



Beiträge zur  
Gesundheitsberichterstattung  
des Bundes

**Erster telefonischer Gesundheitssurvey  
des Robert Koch-Instituts – Methodische Beiträge**



Beiträge zur  
Gesundheitsberichterstattung  
des Bundes

**Erster telefonischer Gesundheitssurvey  
des Robert Koch-Instituts – Methodische Beiträge**

*Martin Kohler, Angelika Rieck, Susan Borch, Thomas Ziese*

# Inhaltsverzeichnis

<b>1</b>	<b>Der telefonische Gesundheitssurvey 2003 – Baustein zum Gesundheitsmonitoringsystem für Deutschland</b> .....	<b>5</b>
	<i>Thomas Ziese</i>	
1.1	Einleitung .....	5
1.2	Themen .....	6
1.3	Eckdaten .....	6
1.4	Vorteile, Grenzen und Ziele .....	7
1.5	Literatur .....	8
<b>2</b>	<b>Methodische Aspekte der Fragebogenentwicklung</b> .....	<b>9</b>
	<i>Angelika Rieck, Susan Borch, Martin Kohler</i>	
2.1	Einleitung .....	9
2.2	Notwendigkeit der Berücksichtigung der Erhebungsmethode .....	9
2.3	Maßnahmen zur Entwicklung des Fragebogens .....	10
2.3.1	Fragebogendramaturgie .....	10
2.3.2	Filterführung .....	13
2.3.3	Optimierung der Operationalisierungen .....	13
2.3.3.1	Mehrdimensionale bzw. komplexe Fragen .....	14
2.3.3.2	Umfangreiche bzw. komplexe Antwortkategorien .....	15
2.3.3.3	Schwer memorierbare Antworten .....	16
2.3.3.4	Reihenfolge von Antwortvorgaben .....	16
2.3.3.5	Hinweistexte und Administrationsanweisungen .....	17
2.3.4	Weitere Techniken/Maßnahmen .....	18
2.4	Fazit .....	18
2.5	Literatur .....	19
<b>3</b>	<b>Feldarbeit</b> .....	<b>21</b>
	<i>Susan Borch, Angelika Rieck</i>	
3.1	Einleitung .....	21
3.2	Personal .....	21
3.3	Datenerhebung und Datenhaltung .....	21
3.4	Mitarbeiterschulung .....	23
3.4.1	Interviewer .....	23
3.4.2	Supervisoren .....	25
3.5	Qualitätssicherung .....	25
3.5.1	Interviewführung und Schulung .....	27
3.5.2	Supervision .....	28
3.6	Fazit .....	29
3.7	Literatur .....	29

<b>4</b>	<b>Stichprobenkonzeption und Stichprobenrealisierung</b> .....	<b>31</b>
	<i>Martin Kohler</i>	
4.1	Einleitung .....	31
4.2	Technische Rahmenbedingungen .....	31
4.3	Stichprobendesign .....	32
4.4	Anrufmanagement und Stichprobenbearbeitung .....	33
4.4.1	Das Anruf- und Rückrufmanagement .....	35
4.4.2	Prozesssteuerung im Call Back Management .....	37
4.5	Datenhaltung .....	44
4.6	Sampleausschöpfung und Response .....	44
4.6.1	Leitende Kriterien für die Klassifikation der disposition codes .....	45
4.7	Datenkontrolle und -aufbereitung .....	48
4.8	Gewichtung der Daten .....	49
4.8.1	Designgewichtung .....	49
4.8.2	Anpassungsgewichtung .....	50
4.8.3	Güte der Gewichtung .....	50
4.9	Ergebnisse zu Response und Nonresponse .....	51
4.9.1	Bekannte Gründe für Nonresponse .....	51
4.9.2	Analyse soziostruktureller und gesundheitsbezogener Merkmale von Teilnehmern und Nicht-Teilnehmern .....	52
4.9.3	Intensität der Samplennutzung .....	53
4.10	Zusammenfassung .....	55
4.11	Literatur .....	56

# 1 Der telefonische Gesundheitssurvey 2003 – Baustein zum Gesundheitsmonitoringsystem für Deutschland

Thomas Ziese

## 1.1 Einleitung

Die Gesundheitsberichterstattung des Bundes (GBE) informiert laufend über den Gesundheitsstatus, das Gesundheitsverhalten und die Gesundheitsversorgung der Bevölkerung in Deutschland. Durch datengestützte Beschreibungen und Analysen können vordringliche Handlungsbedarfe ermittelt und an diesen orientierte Interventionen geplant und umgesetzt werden. Daten aus amtlichen Statistiken, Prozessdaten der Sozialversicherungsträger, Registerdaten zu einzelnen Krankheiten, Meldedaten gemäß dem Infektionsschutzgesetz und themenbezogene Einzelstudien leisten dabei einen wertvollen Beitrag, sind jedoch meist nicht verknüpfbar, decken nur einen Teil der GBE-relevanten Themen ab und erlauben in der Regel keine Differenzierung nach Merkmalen wie beispielsweise dem sozialen Status. Darüber hinaus sind die meisten statistischen Erfassungssysteme und Erhebungsprogramme an ganz bestimmte Aufgaben und Zwecke gebunden und können nur schwer an aktuelle gesundheitspolitische Fragestellungen angepasst werden.

Diese Defizite können die in Deutschland seit Mitte der 1980er Jahre durchgeführten repräsentativen Gesundheitssurveys in weiten Teilen kompensieren. Im Gegensatz zu den meisten anderen Datenquellen lassen sich zudem subjektive Einstellungen, Wahrnehmungen und Verhaltensorientierungen berücksichtigen, die sowohl unter dem Gesichtspunkt der Prävention als auch im Hinblick auf die Heterogenität der Versorgungsbedarfe und der Selbsthilfe im Gesundheitswesen zunehmend an Bedeutung gewinnen.

Die bisherigen Health Examination and Interview Surveys liefern eine breite und valide Datengrundlage, sind aber aufgrund des hohen personellen, logistischen und finanziellen Aufwands nur in mehrjährigen Abständen realisierbar. Ein Gesundheitsmonitoring setzt jedoch eine kontinuierliche Beobachtung in kürzeren Zeit-

intervallen voraus, d. h. es muss die regelmäßige und zeitgerechte Verfügbarkeit gewährleistet sein. Nur so lassen sich aktuelle Trends aufzeigen und im Zusammenhang mit gesundheitspolitischen Interventionen bewerten (Kurth et al. 2005). Aus diesem Grund führt das Robert Koch-Institut seit kurzem zusätzlich repräsentative telefonische Befragungen durch, mit denen die Zeitabstände zwischen den Untersuchungssurveys überbrückt werden sollen.

Telefonische Befragungen haben nicht nur den Vorteil der deutlich geringeren Kosten, sie sind außerdem inhaltlich flexibel und können schneller realisiert werden. In den USA nutzt das Centers for Disease Control (CDC 1998) schon seit über 20 Jahren jährliche repräsentative telefonische Gesundheitssurveys, um die Daten aus den in etwa acht- bis zehnjährigen Abständen stattfindenden nationalen Health Examination Surveys zu ergänzen.

Die Zielsetzung der telefonischen Gesundheitssurveys entspricht denen der nationalen Untersuchungssurveys des RKI, also zum einen die Bereitstellung aktueller Daten für Politik und Wissenschaft zur Verbreitung ausgewählter Erkrankungen vorwiegend mit chronischem Verlauf, zum anderen die Bereitstellung von Informationen zur Verbreitung von wichtigen Risikofaktoren und gesundheitsbezogenem Verhalten in der bundesdeutschen Bevölkerung. Daher wurde bereits bei der Konzeption der telefonischen Surveys auf eine Fortschreibung zentraler Themen der vorangegangenen Untersuchungssurveys, insbesondere des Bundes-Gesundheitssurveys 1998 (BGS98) geachtet. Um Entwicklungen und Trends im Krankheitsgeschehen identifizieren zu können, wurde daher auf viele Instrumente und Fragen zurückgegriffen, die sich im BGS98 bewährt haben. Weitere thematische Bezüge ergaben sich aus dem Themenspektrum der GBE des Bundes, den Schwerpunktbereichen von gesundheitsziele.de sowie aus aktuellen gesundheitspolitischen Bezügen.

## 1.2 Themen

Das thematische Spektrum des telefonischen Gesundheitssurveys 2003 (GSTelo3) umfasst folgende Bereiche:

### **Subjektive Gesundheit**

- ▶ Allgemeine subjektive Gesundheit (European Health Module, EHM)
- ▶ Gesundheitsbezogene Lebensqualität (SF8)

### **Krankheiten**

- ▶ Diabetes, -behandlung, -komplikationen
- ▶ Krankheiten der Atmungsorgane (chronische Bronchitis, Asthma bronchiale)
- ▶ Gelenkerkrankungen (Arthrose, Polyarthritits)
- ▶ Osteoporose und -mitteleinnahme
- ▶ Menopause
- ▶ Rückenschmerzen und -behandlung
- ▶ Schwindel
- ▶ Krankheiten der Augen (Nutzung von Sehhilfen, Kurz- und Weitsichtigkeit, Glaukom, Katarakt, Makula-Degeneration)
- ▶ Krankheiten der Ohren (Nutzung eines Hörgeräts, Tinnitus)
- ▶ Psoriasis
- ▶ Hepatitis
- ▶ Depressionen (Selbsteinschätzung)
- ▶ Herz-Kreislauf-Erkrankungen (Hypertonie, erhöhte Blutfette und Cholesterinwerte, Angina pectoris, Herzinfarkt, Herzinsuffizienz, Herzrhythmusstörungen)
- ▶ Schlaganfall
- ▶ Krebserkrankungen

### **Gesundheitsverhalten, Vorsorge, allgemeine Inanspruchnahme**

- ▶ Sport
- ▶ Alkoholkonsum
- ▶ Rauchen, Hilfsmittel zur Rauchentwöhnung
- ▶ Teilnahme an Selbsthilfegruppen
- ▶ Gripeschutzimpfung, Gesundheits-Check-Up
- ▶ Inanspruchnahme der gesundheitlichen Versorgung

### **Behinderung und körperliche Einschränkungen**

- ▶ Schwerbehinderung, Grad der Behinderung
- ▶ Minderung der Erwerbsfähigkeit, Grad der Minderung der Erwerbsfähigkeit
- ▶ Activities of daily life (ADL)

### **Soziodemografie**

- ▶ Sozialschicht (Bildung, Einkommen, berufliche Stellung)
- ▶ Migrationshintergrund
- ▶ Arbeitslosigkeit
- ▶ Krankenversicherung
- ▶ Sozialhilfeempfang

Rückenschmerzen und Diabetes bilden thematische Schwerpunkte. In beiden Krankheitsbereichen werden Inanspruchnahme von Ressourcen des Gesundheitssystems sowie vom Arzt durchgeführte Untersuchungen und die Einhaltung therapeutischer Leitlinien oder Standards erhoben. Ein weiterer Schwerpunkt liegt im Bereich der vorbeugenden bzw. kurativen Behandlung von Osteoporose bei Frauen im Alter ab 45 Jahren.

## 1.3 Eckdaten

Der GSTelo3 wurde, mit Finanzierung durch das Bundesministerium für Gesundheit und soziale Sicherung, im Zeitraum von September 2002 bis März 2003 realisiert. Es wurden erstmals 8.313 Personen repräsentativ für die deutschsprachige Wohnbevölkerung ab 18 Jahren u. a. zu Krankheiten, zu ihrem Gesundheitsverhalten und zur Inanspruchnahme von Leistungen des Gesundheitswesens befragt. Die Stichprobenbasis bildete ein nach dem Gabler-Häder-Design gezogenes Sample. Unmittelbar nach Abschluss der Feldphase und der Datenaufbereitung wurde mit der Auswertung der erhobenen Informationen begonnen; inzwischen liegen ein deskriptiver Ergebnisbericht (Kohler, Ziese 2005) und erste themenspezifische Publikationen (Lampert, Burger 2004) vor.

## 1.4 Vorteile, Grenzen und Ziele

Telefonische Befragungen haben nicht nur die bereits erwähnten Vorzüge, zeitnäher und mit geringerem Kostenaufwand auch aktuelle Daten zu Veränderungen in gesundheitsrelevanten Bereichen bevölkerungsrepräsentativ gewinnen zu können. Weitere Vorteile telefonischer Befragungen liegen

- ▶ in der Verfügbarkeit eines nationalen Auswahlrahmens durch das Gabler-Häder-Design, das Klumpungseffekte nahezu ausschließt. Damit sind prinzipiell die Möglichkeiten für regionalisierte Analysen gegeben.
- ▶ Telefonische Befragungen sind darüber hinaus auch effizienter in der Feldbearbeitung. Schwer erreichbare Bevölkerungsgruppen sind mit geringerem Aufwand besser zu erreichen, was allerdings entsprechend häufige Kontaktversuche voraussetzt.
- ▶ Die zentrale Unterbringung der Interviewer in einem Telefonstudio ermöglicht eine effizientere Kontrolle der Interviewdurchführung und verbessert die Datenqualität.

Natürlich stehen den Vorteilen telefonischer Befragungen auch Nachteile gegenüber. So bestimmt der ausschließlich akustische Kontakt zwischen Interviewern und Befragten sowohl die Auswahl der Themen mit als auch deren Operationalisierung im Erhebungsinstrument. Dies betrifft u. a. auch die schlechtere Erreichbarkeit spezifischer Teilpopulationen der Bevölkerung mittels der telefonischen Befragungsmethode (z. B. Migranten, soziale Randgruppen, junge Erwachsene, die nur noch über Handy zu erreichen sind). Darüber hinaus steigt der Vorbereitungsaufwand telefonischer Befragungen durch eine methodenadäquate Anpassung gängiger, zumeist für schriftliche oder face-to-face-Erhebungen konzipierte Erhebungsinstrumente. Die damit einhergehende Notwendigkeit der Zerlegung komplexer Fragen in ihre Teildimensionen führt in der Folge zu einer aufwändigeren Aufbereitung der Daten für die Analyse.

Im Kontext des derzeit im Aufbau begriffenen Monitoringsystems (Kurth et al. 2005) spielen die telefonischen Gesundheitssurveys eine tragende Rolle als Bindeglied zwischen den weiterhin

konzeptionell vorgesehenen Examination Surveys. Damit kommt dem GSTelo3 die Funktion einer Piloterhebung zu. Das ist auch der Grund, weshalb im vorliegenden Beitrag zur GBE methodische Aspekte bei Telefonbefragungen vertiefend in einem gesonderten Band behandelt werden, für die in der Darstellung der Ergebnisse des GSTelo3 sonst nicht genügend Raum gegeben werden könnte.

In den hier vorgestellten Einzelbeiträgen werden zentrale methodische Aspekte computergestützter Telefonsurveys behandelt, die nicht nur für den GSTelo3, sondern auch für künftige Gesundheitssurveys von Bedeutung sind.

Im ersten Beitrag geht es um die Notwendigkeit eines erhebungsmethodisch angemessenen Fragebogens und die dazu im GSTelo3 umgesetzten Maßnahmen. Zu diesen zählen Fragebogendramaturgie, Filterführung und Optimierung der Operationalisierungen mit den Schwerpunkten Zerlegung von komplexen Fragen bzw. Antwortkategorien, Festlegung der Reihenfolge für die Antwortvorgaben und Vereinheitlichung retrospektiver Zeitfenster.

Anschließend werden wichtige Aspekte der Feldarbeit des GSTelo3 vorgestellt. Dazu gehören eingesetztes Personal, Datenerhebung und -haltung, Interviewer- und Supervisorschulung sowie die Qualitätssicherungsmaßnahmen. Aufgrund ihrer Bedeutung für Datenqualität, Ausschöpfung und Item-Nonresponse stehen Schulung und feldbegleitende Qualitätssicherung im Zentrum der Darstellung.

Der letzte Beitrag geht ausführlich auf die technischen Rahmenbedingungen, das Stichprobendesign, den Feldverlauf bzw. die Ablaufsteuerung, die Datenaufbereitung und daraus abgeleitet auf die Gewichtung des GSTelo3 ein. Abschließend werden die Ergebnisse einer Non-Responderanalyse vorgestellt und die Auswirkungen der Intensität der Sampleausschöpfung diskutiert.

Die Beiträge zeigen, dass der GSTelo3 hohen Qualitätsansprüchen gerecht wird. Das schließt jedoch in einzelnen Fällen nicht aus, dass methodenspezifische Effekte auch Einfluss auf die erzielten Resultate haben können. Das genauer zu bestimmen, bleibt zukünftigen Methodenstudien vorbehalten.

## 1.5 Literatur

- CDC (1998) Behavioral Health Risk Factor Surveillance Systems's User Guide. Centers for Disease Control, Atlanta
- Kohler M, Ziese T (2004) Telefonischer Gesundheitssurvey des Robert Koch-Instituts zu chronischen Krankheiten und ihren Bedingungen. Beiträge zur Gesundheitsberichterstattung des Bundes. Robert Koch-Institut, Berlin
- Kurth B-M, Ziese T, Tiemann F (2005) Gesundheitsmonitoring auf Bundesebene: Ansätze und Perspektiven. Bundesgesundheitsblatt 3: 261–272
- Lampert T, Burger M (2004) Rauchgewohnheiten in Deutschland – Ergebnisse des telefonischen Bundes-Gesundheitssurveys 2003. Das Gesundheitswesen 66: 511–517



## 2 Methodische Aspekte der Fragebogenentwicklung

Angelika Rieck, Susan Borch, Martin Kohler

### 2.1 Einleitung

Es ist leicht, schlechte Fragen zu stellen (Fowler 2001). Entgegen landläufigen Meinungen gehört die gute Formulierung nicht zu den Bildungsgütern, die quasi nebenbei an der Brust der Alma Mater eingesogen werden. Sie verlangt vielmehr ein spezielles Studium. Vor allem bei CATI<sup>1</sup> scheint der Kunst der Fragebogenentwicklung oft nicht die angemessene Bedeutung zugebilligt zu werden: »Niemand würde auf die Idee kommen, einen Fragebogen, der für mündlich-persönliche Interviews konzipiert wurde, für eine schriftliche Befragung einzusetzen. Leider wird nicht immer beachtet, dass auch beim Telefoninterview auf die Gegebenheiten des Instrumentes Rücksicht genommen werden muss.« (Anders 1990)

Im Folgenden wird zuerst kurz die Notwendigkeit eines erhebungsmethodisch angemessenen Fragebogens aufgezeigt, die sich insbesondere aus den Unterschieden in der kommunikativen Situation beim Telefoninterview im Vergleich zum face-to-face Interview und zur schriftlichen Befragung ergibt (Friedrichs 1990; Dillman 1978; Frey et al. 1990). Diese methodischen Besonderheiten bedürfen aufgrund ihrer Bedeutung für Datenqualität, Item-Nonresponse und Ausschöpfung bei der Entwicklung des Fragebogens besonderer Beachtung. Im weiteren Verlauf des Artikels werden die wichtigsten Maßnahmen vorgestellt, die dazu für den telefonischen Gesundheitssurvey 2003 (GSTel03) umgesetzt wurden.

### 2.2 Notwendigkeit der Berücksichtigung der Erhebungsmethode

Bei Telefoninterviews erfolgt die Kommunikation im Gegensatz zu face-to-face-Interview oder schriftlicher Befragung ausschließlich über verbale bzw. vokale Kanäle (Friedrichs 1990; Dillman 1978; Frey et al. 1990). Diese (fern-) mündliche Darbietung aller Informationen hat Konsequenzen für das Verhalten von Befragten und Interview-

wern bei der Datenerhebung, die bei der Fragebogenentwicklung angemessen berücksichtigt werden müssen (Friedrichs 1990; Dillman 1978).

Telefoninterviews werden laut Friedrichs (1990) von den Befragten im Vergleich zu face-to-face-Interviews als unangenehmer empfunden. Bedingt durch die geringere Zahl kommunikativer Kanäle sei ein angenehmes emotionales Klima zwischen den Interviewpartnern (sog. Rapport) schwerer zu erreichen. Durch das Bemühen des Befragten um Substitution der fehlenden nonverbalen Merkmale des Interviewers sowie die Suche nach Vertrauen und Bestärkung würde beim Interviewer eine falsche Reaktion zur Verbesserung des emotionalen Klimas provoziert. In Folge dessen seien Abweichungen der Interviewführung von der Standardisierung, die sich beispielsweise äußern in Veränderungen des Wortlautes der Fragen zur Anpassung an das vermutete Niveau des Befragten, spontanen Formulierungen von Überleitungen sowie verkürzten Pausen oder im Überbrücken von Pausen durch Füllphrasen (z. B. »ja, ja«, »hmm«), bei Telefonsurveys häufiger anzutreffen als bei face-to-face-Interviews.

Deshalb sollte der Fragebogen für Telefonsurveys noch stärker an einen natürlichen Kommunikationsfluss angepasst werden (Friedrichs 1990), sodass beispielsweise Überleitungen im Erhebungsinstrument vorgesehen sind und nicht vom Interviewer improvisiert werden müssen. Außerdem ist es für Telefoninterviews wichtig, den Fragebogen an die Alltagssituation der Probanden anzupassen (Frey et al. 1990), um einen möglichst unmittelbaren, einfachen Zugang zu den gewünschten Informationen zu erreichen. Dabei hat die Fragebogendramaturgie, d. h. die optimale Aneinanderreihung einzelner Fragen und Fragengruppen sowie die Integration von Überleitungstexten und Filterfragen, eine besondere Bedeutung für die Beantwortung der Fragen und deren Qualität (Fuchs, Lamnek 1990; Frey et al. 1990; Dill-

<sup>1</sup> CATI steht für computer-assisted telephone interviewing, also computergestütztes Telefoninterview.

man 1978). Hinsichtlich des Abbruchverhaltens von Probanden ist die Gestaltung der Kontakt- und Eingangsphase für Telefoninterviews ohne schriftliche Vorankündigung (so genannte Kaltkontaktierung) besonders wichtig, weil die ersten Minuten für das Zustandekommen eines vollständigen Interviews maßgeblich sind (Dillman 1978; Frey et al. 1990).

Die Formulierung von Fragen und Antwortvorgaben hat im Vergleich zu anderen Erhebungsverfahren beim Telefoninterview eine noch höhere Priorität. Nur so kann eine Frage, die am Telefon vorgelesen wird, mindestens genauso gut werden wie eine vorgelegte bei einer postalischen Befragung oder eine visuell unterstützte beim face-to-face-Interview (Wüst 1998). Eine besondere Herausforderung für die Operationalisierung entsteht aus der beim Telefoninterview begrenzten Komplexität, weil die Fragen und Antwortvorgaben vollständig von den Probanden memoriert werden müssen. Das erfordert bei Telefoninterviews den konsequenten Einsatz von kürzeren und einfacheren Fragentexten mit weniger und kürzeren Antwortkategorien (Dillman 1978). Insgesamt sind allgemeine Lehrbuchregeln zur Fragenformulierung also bei einer CATI-Erhebung noch strikter einzuhalten als bei face-to-face-Interviews oder schriftlichen Befragungen (Frey et al. 1990).

Beim face-to-face-Interview und vor allem bei schriftlichen Befragungen sind Phasen längeren Nachdenkens eher möglich als beim schnelleren Telefoninterview (Friedrichs 1990), wo die Interviewer mit ihrer Vorlesegeschwindigkeit das Tempo für die Befragten vorgeben (Schwarz et al. 1991). Dies führen Schwarz et al. (1989, 1991) als entscheidenden Beitrag zum schlechteren recall beim Telefoninterview an. Deshalb sind für Telefonsurveys besonders sorgfältig Vorkehrungen zu treffen, um den Interviewten mehr Zeit zum Verstehen der Fragen und zum Nachdenken einzuräumen.

Da die Befragten im Telefoninterview unklare Fragen nicht so einfach unbeantwortet überspringen können, wie bei einem schriftlichen Fragebogen, bzw. nicht lange selbständig über die Bedeutung einer Frage nachdenken (Schwarz et al. 1991), bitten sie die Interviewer um Erläuterung oder Hilfestellung. Wenn die Interviewenden keine Anweisungen dafür haben, wie sie in dieser Situation reagieren sollen, weichen sie von der Standardi-

sierung ab, indem sie hilfsbereit ihre eigene Meinung über die Intention der Frage wiedergeben, um den Rapport aufrecht zu erhalten (Friedrichs 1990). Wird dieser Aspekt jedoch bei der Entwicklung des Fragebogens (Fowler 2001) und der Interviewerschulung (Borch, Rieck 2005) angemessen berücksichtigt, können Telefonsurveys hinsichtlich der korrekten und vollständigen Beantwortung einen Vorteil gegenüber schriftlichen Fragebögen geltend machen (Dillman 1978).

### 2.3 Maßnahmen zur Entwicklung des Fragebogens

Probanden können ermüden, unwillig werden, Fragen falsch verstehen, antworten ohne nachzudenken oder auch das Interview schlicht und einfach kurzentschlossen abbrechen. Folgende Maßnahmen sollten das Gegenteil erreichen. Zusätzlich sind sie konzipiert, um ein weitgehend einheitliches Interviewerverhalten zu unterstützen. Ihre Erprobung und Weiterentwicklung war Aufgabe von zwei Pretests.

#### 2.3.1 Fragebogendramaturgie

Nach Dillman (1978) dient die Dramaturgie des Fragebogens bei Telefoninterviews hauptsächlich der Erhöhung der Motivation sowie der Verringerung von Verständnisproblemen der Probanden und damit zur Reduzierung der Anforderungen an die Befragten. Fühlen sich Interviewte aufgrund einer schlechten Dramaturgie überfordert, weil die Fragen nicht nach Themen bzw. Gegenständen gruppiert und innerhalb der Themen nicht nach Antwortformaten sortiert präsentiert werden, besteht die Gefahr, dass diese nicht um wiederholtes Vorlesen der Fragen bitten, sondern eine willkürliche Antwort geben (ebd.).

Deshalb wurden zur Erreichung einer guten Fragebogendramaturgie für den GSTelo3 thematisch zusammengehörende Fragen gemäß einer alltagsnahen Perspektive zu Blöcken zusammengefasst. So entspricht beispielsweise die gemeinsame Thematisierung von Prävalenz, Diagnostik und Therapie von Rückenschmerzen eher einem alltagsnahen Gespräch als die separate Abhandlung in getrennten Blöcken zur Therapie, Diagnostik usw., d. h. als eine analytische Blockbil-

dung. Außerdem erfährt die Erinnerungsfähigkeit durch diese Fokussierung auf einen Gegenstand Unterstützung. Innerhalb der Blöcke wurde mit möglichst einfachen und einführenden Fragen begonnen, die den Gegenstand allgemein thematisieren, während spezifischere Items erst später folgten. Diese Aneinanderreihung vom Allgemeinen zum Speziellen sollte von den Probanden als nachvollziehbar und logisch empfunden werden sowie ihre Erinnerungsleistung verbessern. Ähnliche Themen wurden zu Bereichen zusammengelegt, wie z. B. der Bereich (ärztlich diagnostizierter) Krankheiten, der Bereich Gesundheitsverhalten oder der Bereich krankheitsunspezifisch abgefragter Therapie und Inanspruchnahme.

Blöcke bzw. Bereiche wurden durch Überleitungstexte voneinander getrennt, um die Befragten zu entlasten, die Motivation aufrecht zu erhalten, gedanklich auf den neuen Themenbereich hinzuführen und Überstrahlungseffekte zu verringern. Bei langen Blöcken bzw. Themenbereichen signalisierten anschließende Überleitungen den Probanden die Bewältigung eines (weiteren) Teils des Interviews. Eine entsprechende Programmierung bewirkte, dass diese Texte von den Interviewern genauso wichtig genommen wurden wie Fragentexte. Das stellte ein korrektes Vorlesen sämtlicher Überleitungen sicher. Die Erfahrungen während der Feldphase haben gezeigt, dass Überleitungstexte bei künftigen Telefonsurveys sogar noch häufiger eingesetzt werden sollten.

Für die Kontakt- und Eingangsphase vor Beginn des eigentlichen Interviews wurden Vorstellungstexte und Einleitungstexte entwickelt, die gemeinsam mit einem bestimmten (in der Schulung eingeübten) Vorgehen der Interviewer (Borch, Rieck 2005) dazu dienen, innerhalb von wenigen Minuten bei den Probanden Vertrauen zu wecken, zu informieren und zur Teilnahme zu motivieren. Am inhaltlichen Beginn des Interviews (warming-up-Phase) befinden sich einfache, auf den Gegenstand hinführende und Rapport erzeugende Fragen zur allgemeinen subjektiven Gesundheit (Ziese 2005). Die anspruchsvollsten Blöcke, Diabetes mellitus und insbesondere Rückenschmerzen (ebd.), wurden nach der warming-up-Phase in den Hauptteil des Fragebogens eingefügt, um die Wirkung der im Laufe des

durchschnittlich 21-minütigen Interviews auftretenden Ermüdung abzuschwächen. Der letzte Teil des Interviews beinhaltet die soziodemografischen Fragen (ebd.).

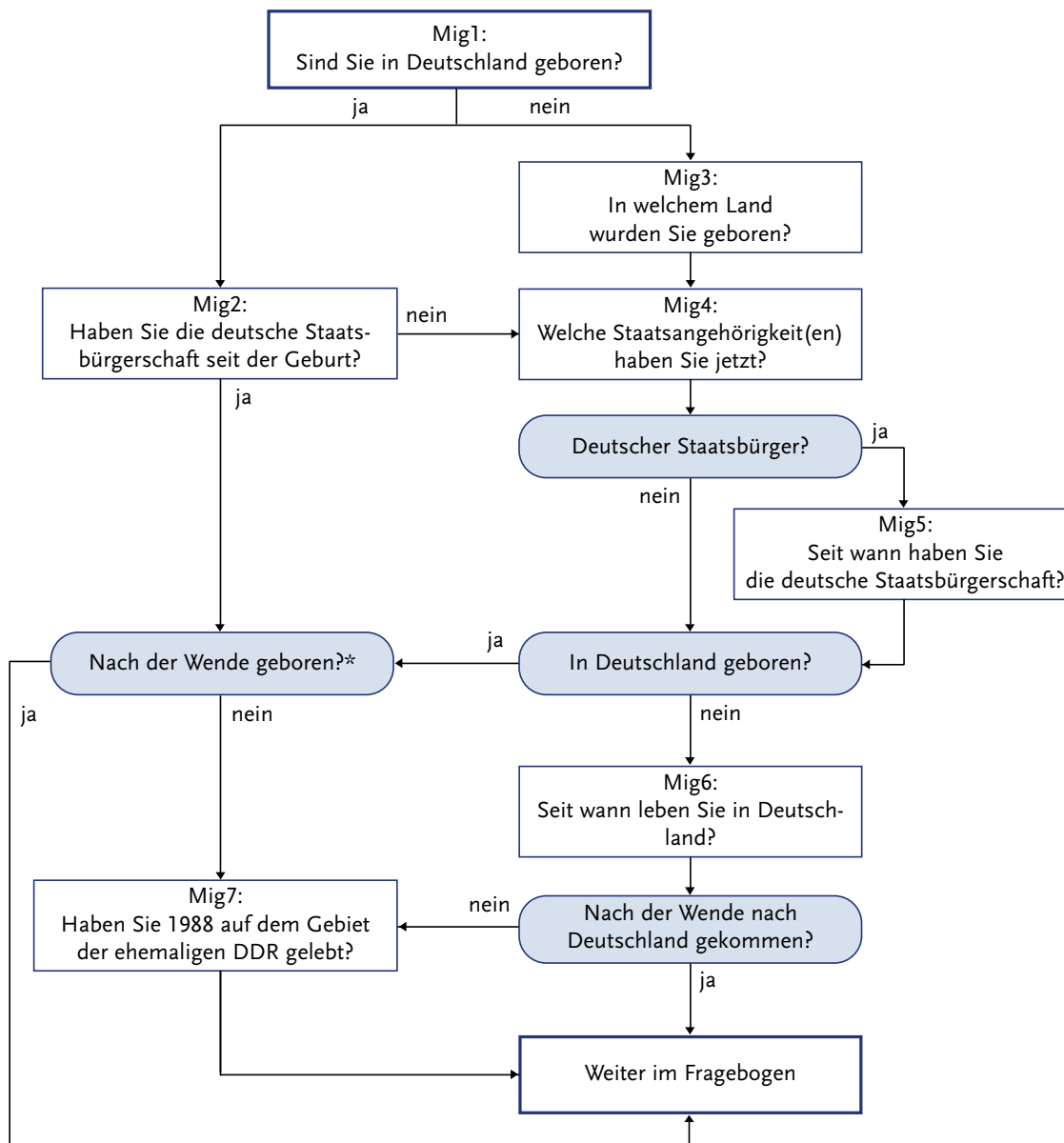
Im Hauptteil des Fragebogens wurden zuerst Krankheitsfragen (fast ausschließlich als ärztliche Diagnosen), dann gesundheitsverhaltensbezogene Themen und schließlich allgemeine, nicht krankheitsspezifische Aspekte der Krankheitsversorgung bzw. Inanspruchnahme des Gesundheitsversorgungssystems behandelt (ebd.). Diese Aneinanderreihung von Bereichen dürfte eher der Erwartung der Probanden entsprechen und deshalb weniger Widerstände bzgl. der Beantwortung der Fragen provozieren als die umgekehrte Reihenfolge. Die vermutete höchste Bedeutungszuweisung für die Krankheitsfragen durch die Befragten stimmt zudem mit der zentralen Stellung dieser Variablen für die Erreichung der Studienziele (ebd.) überein. Die exponierte Positionierung dieser Fragen sollte eine wache Präsenz und eine hohe Motivation der Interviewten gewährleisten, um für diese Daten die bestmögliche Güte (im Sinne von Validität und Reliabilität) zu erreichen.

Der Fragebogen sollte einerseits hinsichtlich Themen und Fragearten bzw. Antwortformaten klar strukturiert und nicht zu heterogen aufgebaut sein, andererseits sollte auch keine Monotonie entstehen. Da der gesamte Fragebogen des GSTel03 fast ausschließlich aus Tatsachenfragen bestand, war es um so wichtiger, das Interview abwechslungsreich zu gestalten, indem sich Bereiche bzw. Blöcke mit anspruchsvolleren Items mit leichteren abwechselten. Beispielsweise befanden sich zwischen den beiden anspruchsvollsten Blöcken zu Diabetes und Rückenschmerzen vier leichter zu beantwortende Fragen und ein kurzer Block zu Osteoporose. Die bereits thematisierten Überleitungstexte wurden ebenfalls gezielt gegen das Aufkommen von Monotonie eingesetzt. Zum Beispiel gibt es in dem langen Fragenblock zu Rückenschmerzen mehrere Überleitungen. Vor allem innerhalb umfangreicher Blöcke, wie Diabetes mellitus, wurde darauf geachtet, lange Batterien von Fragen, die mit ja oder nein beantwortbar waren, durch andere Antwortformate aufzulockern. So kamen im Diabetesblock zusätzlich standardisierte Fragen mit bis zu fünf vorgelesenen Antwortkategorien und offene Fragen (z. B. nach der Anzahl der ärztlichen Untersuchungen der Füße

in den letzten 12 Monaten) vor. Diese Maßnahmen zur Auflockerung der Tatsachen- und Faktfragen waren allerdings nicht so wirksam, wie die ein Jahr später für den Telefonsurvey 2004 vorgenommene Integration anderer Fragearten, d. h. Einstellungs- und Wissensfragen. Deshalb wurden die

Interviewenden im Survey 2003 stärker bzgl. der Aufrechterhaltung der Motivation der Befragten gefordert als beim Folgesurvey. Auf diese Aufgaben wurden sie daher in der Interviewerschulung entsprechend vorbereitet (Borch, Rieck 2005).

**Abbildung 1**  
Filter für Migrationshintergrund



Hinweis: Fragetexte nicht vollständig abgebildet

\* Diese Filterbedingung wird erst ab 2006 relevant, da Minderjährige nicht zur Grundgesamtheit gehören.

□ Fragen ○ Filterbedingung

### 2.3.2 Filterführung

Filter sind ein zentrales Instrument von CATI. In diesem Punkt unterscheidet sich CATI nicht von CAPI<sup>2</sup>, jedoch von konventionellen face-to-face-Interviews und erst recht von schriftlichen Fragebögen. Die Filtertechnik diente zur Optimierung der Dramaturgie als auch der Operationalisierungen. Die Notwendigkeit einer stringenten Filterführung ergab sich aus dem komplexen und umfangreichen Erhebungsinstrument des GSTelo3, das aus 248 Fragen bzw. Items bestand. Mit ihrer Hilfe konnten für bestimmte Probandengruppen redundante bzw. widersinnige Fragen vermieden und das Interview insgesamt so kurz wie möglich gehalten werden, wie z. B. von Fowler (2001) gefordert. Der Verzicht auf redundante Variablen sollte zur Motivierung der Befragten beitragen, gegen eine vorzeitige Ermüdung wirken und damit zur Erhöhung der Güte der wichtigen zu erfassenden Items führen sowie Abbrüchen vorbeugen. Eine situationsabhängige Steuerung oder Variation von Fragen (bzw. Teilen davon) und Überleitungen mittels Filterung führt darüber hinaus zu einem natürlicheren Kommunikationsfluss zwischen den Interviewpartnern.

Der Beitrag der Filtertechnik zur Optimierung der Operationalisierungen ist Gegenstand des nächsten Abschnittes. Dramaturgisch wirken Filter beispielsweise bei der Einspielung unterschiedlicher Überleitungstexte in Abhängigkeit von bereits beantworteten Fragen.<sup>3</sup>

Die Einsparung von redundanten Fragen als dramaturgische Maßnahme soll am Beispiel von Diabetes angedeutet werden: Der Diabetesblock bestand insgesamt aus 29 Fragen, wobei Frauen teilweise andere Fragen bekamen als Männer. Probanden ohne Angabe von Diabetes verließen diesen Block bereits nach der ersten Frage. Trat die Krankheit bei Frauen nur während einer Schwangerschaft auf, wurden lediglich vier Fragen gestellt. Befragte mit Insulintherapie und Blutzuckerselbstkontrollen bekamen drei Fragen mehr. Die Nichtkenntnis von Diabetiker-Schulungen führte zur Einsparung von drei Fragen usw.

Geschickte Filterführung verhinderte z. B. bei der Erfassung des Migrationshintergrundes die Konfrontation der Probanden mit widersinnigen Fragen, wie der schematischen Darstellung in Abbildung 1 entnehmbar ist.

Für alle anderen gegenstandsorientierten Blöcke des Fragebogens gab es ebenfalls eine gezielte Anordnung von Fragen und eine stringente Filterführung. Auch wenn aufgrund von einzelnen Filterbedingungen teilweise nur wenig eingespart wurde, führte das insgesamt zu einer erheblichen Reduktion der Anzahl gestellter Fragen für viele Probanden und damit zu einer spürbaren Entlastung.

Der konsequente Einsatz der Filtertechnik stellte jederzeit und für jede Probandengruppe die korrekte Auswahl und Aneinanderreihung der zu beantwortenden Fragen sicher (Dillman 1978), ohne dass Interviewer oder Probanden damit beansprucht wurden, wie dies beispielsweise bei einem nicht computergestützten face-to-face-Interview oder insbesondere bei einer schriftlichen Befragung der Fall wäre. Durch diese intensive Nutzung der Filtertechnik entsteht allerdings ein erhöhter Aufwand bei der Programmierung des Fragebogens und der Aufbereitung der Daten für die Auswertung.

### 2.3.3 Optimierung der Operationalisierungen

Warum einfach fragen, wenn es auch kompliziert geht? Hier geht es vor allem um die Zerlegung von Fragen mit zu komplexen Formulierungen oder zu komplexen Antwortkategorien, die Festlegung der Reihenfolge für die Antwortvorgaben und die Vereinheitlichung retrospektiver Zeitfenster. Hinweistexte sollten Unklarheiten minimieren, um ein einheitliches und richtiges Verständnis der Fragen zu gewährleisten.

2 computerassistiertes persönliches Interview bzw. computergestütztes face-to-face-Interview

3 Nach den einführenden Fragen zur allgemeinen subjektiven Gesundheit erschienen in Abhängigkeit von deren Beantwortung zwei unterschiedliche Überleitungen. Bei eher besserem allgemeinen Gesundheitszustand folgte: »Auch wenn das alles möglicherweise gar nicht auf Sie zutrifft, möchte ich Sie im Folgenden kurz nach einigen Krankheiten fragen.« Bei schlechterem Zustand kam hingegen: »Als nächstes möchten wir Sie nach einigen Erkrankungen fragen, die häufiger auftreten.«

### 2.3.3.1 Mehrdimensionale bzw. komplexe Fragen

Komplexe Frageformulierungen sind durch sprachliche Überfrachtung bzw. Mehrdimensionalität gekennzeichnet, weshalb sie von Befragten unterschiedlich verstanden werden (Friedrichs 1980). Dies betraf insbesondere Fragen, die für schriftliche Fragebögen oder persönliche Interviews entwickelt wurden bzw. für den Einsatz per CATI noch nicht ausreichend operationalisiert waren. Da ein Ziel des GSTelo<sub>3</sub> in der Fortschreibung wichtiger Items aus vergangenen Surveys bestand, mussten diese Fragen inhaltlich äquivalent in die CATI-taugliche Form transformiert werden. Für komplexe Frageformulierungen erfolgte eine Zerlegung in einfachere Teilfragen (Dillman 1978), die so angeordnet wurden, dass die Anzahl der zu stellenden Fragen mittels Filterführung minimiert wird.

Einfaches Beispiel: Statt der Frage »Wie viele Stunden in der Woche haben Sie durchschnittlich in den letzten drei Monaten Sport getrieben?« verwenden wir zwei Fragen, nämlich »Denken Sie mal an die letzten drei Monate. Haben Sie da Sport gemacht?« plus »Und wie viele Stunden in der Woche waren es in etwa?« Die zweite Frage bleibt den Nichtsportlern aufgrund von Filterung erspart, während sie den Sportlern Zeit zum Nachdenken gewährt.

Ähnlich war das Vorgehen für andere Fragen zur Erfassung von Zeit-, Häufigkeits- oder Mengenangaben, weil diese besonders hohe Anforderungen an das Erinnerungsvermögen der Befragten stellen. Dabei wurde zunächst unter Berücksichtigung der Güte der Daten und der später intendierten Auswertungen geprüft, wie differenziert die Informationen erhoben werden müssen. Wenn möglich wurden diese Angaben dann nicht offen, sondern in Kategorien erfasst, wie von Dillman (1978) vorgeschlagen. Für das o. g. Sportbeispiel gibt es vier Antwortvorgaben.

In einigen Fällen konnten anspruchsvolle Fragen zu Zeit- oder Häufigkeitsangaben sogar durch eine einfachere Frage substituiert werden. Beispielsweise interessierte für Probanden mit Diabetes, ob der Augenhintergrund – wie empfohlen – einmal jährlich augenärztlich untersucht wurde. Eine Frage nach dem Zeitpunkt der letzten Untersuchung, wie im Bundes-Gesundheitssurvey 1998 gestellt, liefert diese Information. Bei einer derartigen Operationalisierung wären jedoch

im Rahmen des Telefoninterviews die Erinnerungsfähigkeit und Motivation der Probanden stark strapaziert worden. Zudem waren erhebliche Zweifel an der Verlässlichkeit (Güte) der Angaben zu befürchten. Schließlich wäre die interessierende Dimension auf diese Weise hinsichtlich des Erkenntnisinteresses differenzierter operationalisiert als erforderlich. Deshalb wurde eine einfachere Lösung gewählt, nämlich ob diese Untersuchung in den letzten 12 Monaten erfolgte. Das ist eine vergleichsweise einfache, mit ja oder nein zu beantwortende Frage.

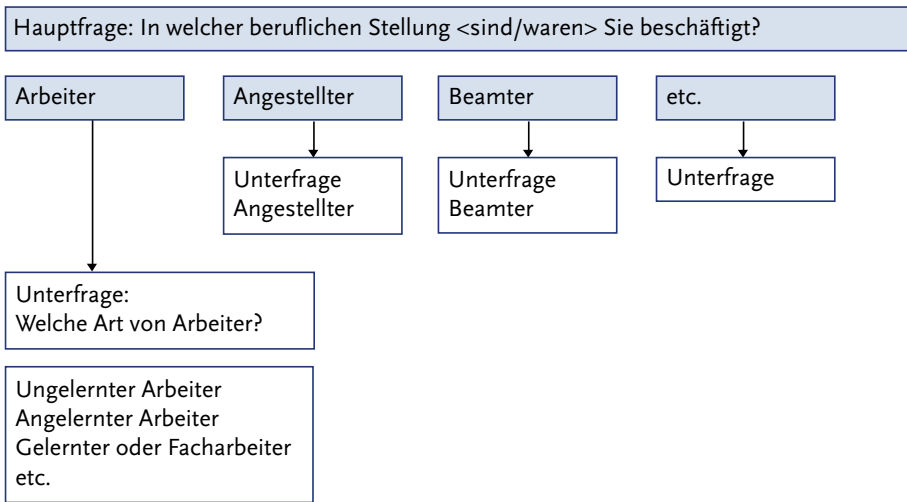
Für viele Gegenstände des GSTelo<sub>3</sub> interessierten konkrete Zeiträume. Die Handhabung der Zeitdimension verdeutlicht, wie eng die Vereinfachung von Fragen mit der Dramaturgie verknüpft ist. Um die Befragten nicht durch die Verwendung einer Vielzahl retrospektiver Zeitfenster zu verwirren, erfolgte – soweit dies mit dem Erkenntnisinteresse der Studie vereinbar war – eine Vereinheitlichung der Zeitdimension. Diese Vereinheitlichung war vor allem für nahe beieinander liegende sowie thematisch zusammenhängende Fragen von Bedeutung, weil sich im Pretest zeigte, dass die Probanden sonst besonders häufig zu Verwechslungen der Zeitdimensionen neigten. Hauptsächlich wurden die bisherige Lebenszeit, die letzten 12 Monate und die letzten 4 Wochen thematisiert. Innerhalb einiger Themen bzw. Blöcke ließen sich unterschiedliche Zeitfenster nicht vermeiden. Besondere Vorkehrungen sollten in diesen Fällen Verwechslungen vorbeugen: Einerseits wurde die Zeitdimension im Rahmen der Formulierung der Fragen besonders betont<sup>4</sup> und auch wiederholt benannt<sup>5</sup>. Andererseits stellten differierende Zeitfenster ein eigenes dramaturgisches Element dar: Beispielsweise wurde bei Rückenschmerzen zuerst nach der bisherigen Lebenszeit gefragt, dann nach den letzten zwölf Monaten und schließlich nach dem Vortag der Befragung. Das heißt, die Fragen wurden nach der Zeitdimension sortiert, um eine optimale Filterführung zu ermöglichen.

Neben einer Vereinheitlichung der Zeitfenster geht es auch um deren optimale Festlegung in Abhängigkeit von der Erinnerungsfähigkeit der

4 zum Beispiel: »Denken Sie mal an die letzten 3 Monate. Haben Sie da Sport gemacht?«

5 beispielsweise innerhalb eines Überleitungstextes, der der Frage voranging

**Abbildung 2**  
Hauptfrage und Unterfrage für Arbeiter



Befragten, was am o.g. Sportbeispiel verdeutlicht werden kann: Die wöchentliche Stundenzahl sportlicher Aktivität lässt sich für die letzten drei Monate wesentlich besser und schneller memorieren und damit valider erfassen, als für die letzten 12 Monate.<sup>6</sup>

### 2.3.3.2 Umfangreiche bzw. komplexe Antwortkategorien

In der Literatur werden für Telefoninterviews maximal vier bis fünf Antwortkategorien empfohlen, wobei die genaue Festlegung dieser Anzahl von deren Länge abhängt (Frey, Oishi 1995). Deshalb erfolgte für Fragen mit umfangreicheren Antwortkategorien möglichst eine Zerlegung in mehrere Teilfragen (Dillman 1978). So wurde zum Beispiel die berufliche Stellung nicht in einer Frage mit sehr vielen und langen Antwortkategorien erhoben, wie bei der schriftlichen Befragung des Bundes-Gesundheitssurveys 1998, sondern zerlegt in eine Frage nach der allgemeinen beruflichen Stellung mit den Kategorien Arbeiter, Angestellter, Beamter usw., der gefiltert nach der gewählten Kategorie eine weitere Frage zum differenzierten Beschäftigungsstatus folgt. Es gibt also eine Unterfrage für die Arbeiter, eine andere für die Angestellten, eine für die Beamten etc. Abbildung 2 demonstriert diese zweistufige Abfrage für die Arbeiter.<sup>7</sup>

Bei einigen Fragen bzw. Dimensionen mit zu vielen Antwortvorgaben führte die Zerlegung zu mehr als zwei Abfragestufen. Das verlangte allerdings eine entsprechende Gruppierung der Antwortkategorien zu Teilfragen mit hierarchischer Aneinanderreihung und differenzierter Filterführung (Wüst 1998). Solche Abfragen auf mehreren Stufen fanden u. a. innerhalb der Soziodemografie Anwendung. Dabei entstand eine Mischung aus hierarchisch und gleichberechtigt angeordneten Fragen.

Ein gutes Beispiel dafür ist auch die Erfassung des Anrufergebnisses mit Hilfe der so genannten disposition codes (Kohler 2005; Borch, Rieck 2005). 48 Einzelkategorien (z. B. Freizeichen, Anrufbeantworter, Geschäftsanschluss) erforderten eine übersichtliche und strukturierte Präsentation auf dem Bildschirm, um die Interviewer nicht zu überfordern und Qualitätsprobleme bei den für die Berechnung der Responserate zentralen Variablen zu vermeiden (Kohler 2005). Sie wurden so auf neun Bildschirmseiten organisiert, aneinandergereiht und gefiltert, dass die Interviewer in einer konkreten Kontaktsituation nur eine bis maximal vier hierarchisch angeordnete Seiten durchgehen mussten.

<sup>6</sup> Zu Erinnerungsfehlern bei Retrospektivfragen vgl. z. B. Bernard et al. (1984) und Schwarz, Sudman (1994).

<sup>7</sup> Für die Hauptfrage wurden ‚sind‘ oder ‚waren‘ situationsabhängig, d. h. filtergesteuert in Abhängigkeit vom zuvor erhobenen aktuellen Erwerbsstatus, eingeblendet.

Nicht alle Fragen mit relativ umfangreichen bzw. vergleichsweise anspruchsvollen Antwortkategorien ließen sich mit Hilfe der bisher beschriebenen Schritte in Teilfragen zerlegen. Dafür gab es unterschiedliche Gründe. Beispielsweise passten die allgemeineren, abstrakteren Gruppenbezeichnungen für die gesetzliche Krankenversicherung, wie Ersatzkassen, Innungskrankenkassen etc., nicht zur Lebensrealität der Probanden. Denn die Befragten kennen zum Teil nur den konkreten Namen der Krankenkasse oder irgendeine Kurzform davon und nicht diese Gruppenbezeichnung. Deshalb mussten diese Kassentypen durch Beispiele ergänzt und mit Hilfe einer umfangreicheren Liste erhoben werden. Bei Zuordnungsproblemen gab es die Möglichkeit einer freitextlichen Eingabe. Eine Zerlegung in mehrere hierarchisch angeordnete und gefilterte Fragen wäre hier keine geeignete Lösung gewesen. Die notwendige Unabhängigkeit der Beantwortung der allgemeineren Einstiegsfrage vom Wissen hinsichtlich der Antwortvorgaben der nachgeschalteten, detaillierteren Unterfragen diskutiert auch Dillman (1978). Sie ist eine notwendige Voraussetzung für die Vermeidung von Güteproblemen (bzgl. Validität und Reliabilität) hinsichtlich sämtlicher Teilfragen. Außerdem müssen die zu zerlegenden Antwortkategorien überhaupt sinnvoll gruppierbar sein (Wüst 1998).

### 2.3.3.3 *Schwer memorierbare Antworten*

Wie übermittelt man nicht weiter zerlegbare Fragen mit relativ schwer zu memorierenden Antwortalternativen? Eine geeignete Maßnahme ist beispielsweise die doppelte Nennung der Kategorien sowohl als Bestandteil des Fragentextes als auch des Antwortteils der Frage, d. h. das mindestens zweifache Vorlesen, wie von Dillman (1978) empfohlen. Oder es erfolgte eine Anweisung an die Interviewer, bestimmte Fragen oder Antwortkategorien besonders langsam und mehrfach vorzulesen (siehe unten, Administrationsanweisungen). Außerdem war die Art und Weise des Vorlesens von Fragen und Antworten, wie z. B. ihr Tempo, ihre Intonation und einzulegende Pausen, intensiver Gegenstand der Interviewerschulung und Qualitätssicherung (Borch, Rieck 2005).

Da schwer memorierbare Antwortvorgaben – ebenso wie anspruchsvollere Frageformulierungen – eine Verlangsamung des Interviewtempos erfordern, sind Maßnahmen zur Regulierung des Tempos, wie eben genannt, ein wichtiger Bestandteil gelungener CATI-Fragebögen, die Interviewerschulung und Qualitätssicherung entlasten und zur Erhöhung der Güte der Daten beitragen.

### 2.3.3.4 *Reihenfolge von Antwortvorgaben*

Bei der Optimierung der Operationalisierungen hinsichtlich der Antwortkategorien waren neben Anzahl und Umfang auch deren Reihenfolge zu berücksichtigen. In der Literatur (z. B. Schwarz et al. 1991) gibt es vor allem für Meinungsfragen die Diskussion, dass in Abhängigkeit vom eingesetzten Erhebungsverfahren eher mit einer bevorzugten Auswahl der erst- oder letztgenannten Kategorien zu rechnen ist (response order effects). Für die im GSTelo3 verwendeten Fragearten stand allerdings die Festlegung einer optimalen Reihenfolge der Antwortvorgaben unter dem Aspekt ihrer Nutzbarkeit für die Interviewer und Befragten in der interaktiven Situation am Telefon im Vordergrund, um eine hohe Güte bei der Antwortauswahl hinsichtlich der intendierten Zieldimension zu erreichen (Fowler 2001).

Beispielsweise sollte die Facharztbezeichnung des den Diabetes mellitus behandelnden Arztes mittels fünf vorgelesener Antwortkategorien ermittelt werden. Dabei sollte eine Arztgruppe ausgewählt werden und zwar jene, die die Interviewten hauptsächlich therapiert. Die Antwortvorgaben wurden dafür vom Speziellen (Diabetikerzentrum) zum Allgemeinen (praktischer Arzt/Allgemeinarzt) sortiert. Diese Reihenfolge war wichtig, weil ansonsten die Gefahr von Demotivierung steigen würde. Denn es provoziert sowohl bei Befragten als auch Interviewenden Widerstände, eine für Telefoninterviews recht große Anzahl von Antwortkategorien weiter durchzuarbeiten, nachdem bereits eine (scheinbar) zutreffende Antwort gefunden wurde. Da die meisten Diabetespatienten hauptsächlich vom Hausarzt oder vom nicht auf Diabetes spezialisierten Internisten behandelt werden, war mit diesem Problem häufiger zu rechnen. Und für die vergleichsweise wenigen Kranken, die sich zur Therapie hauptsächlich ins Diabetiker-



zentrum begeben, würde das Risiko einer Fehlklassifikation steigen, wenn diese zusätzlich hausärztlich versorgt werden. Denn sie würden bei einer Anordnung der Kategorien vom Allgemeinen zum Speziellen eventuell ungeduldig und vorschnell das Gebiet Hausarzt auswählen. Eine zweckmäßige Aneinanderreihung unterstützt beide Interviewpartner hinsichtlich der vollständigen Einbeziehung der Antwortvorgaben und führt damit zu einer besseren Datenqualität.

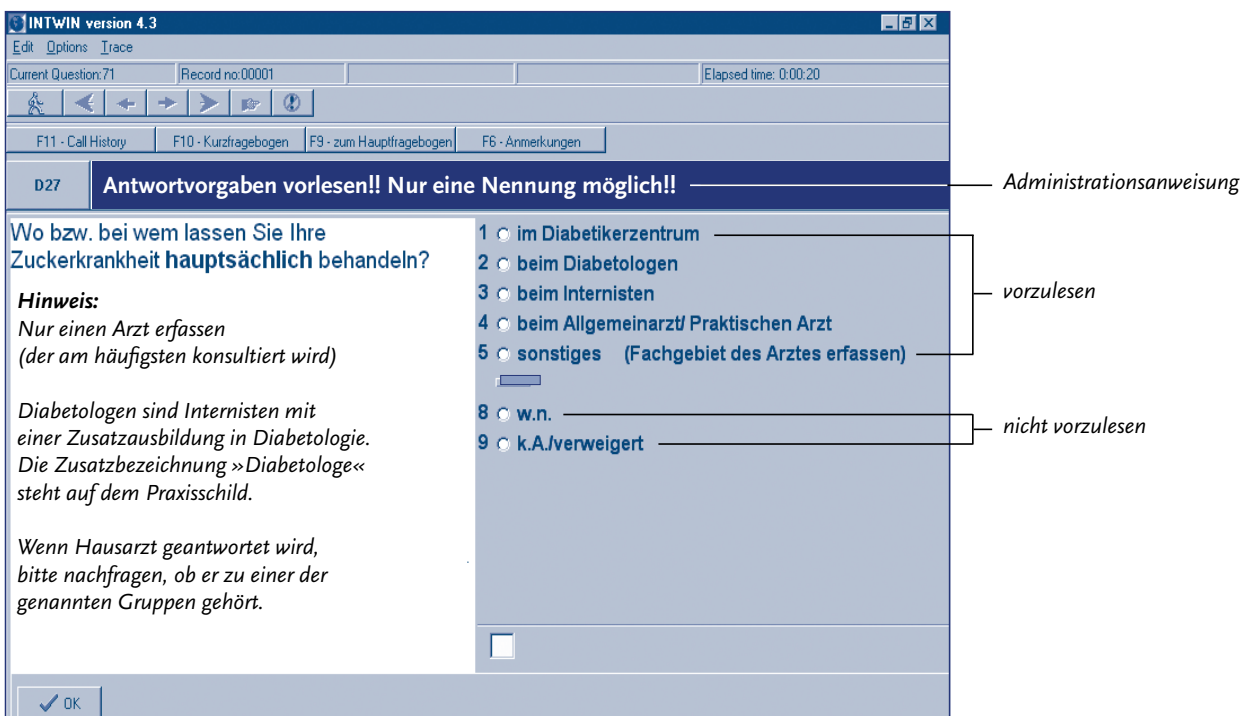
Gegenüber diesem Beispiel bestand die Herausforderung bei der Erfassung der Krankenversicherung darin, die Befragten und Interviewer zu motivieren, aus einer längeren Liste von Krankenkassentypen, die durch Beispiele für die wichtigsten Kassennamen ergänzt waren, sorgfältig eine Auswahl zu treffen. Die Anordnung der Antwortvorgaben erfolgte hier nach ihrer Auswahlwahrscheinlichkeit, d. h. die vermutlich häufigeren Kategorien standen vor den seltenen. Damit mussten nach erfolgreicher Zuordnung zu den ersten vorgelesenen Antwortkategorien einige recht spezielle Krankenversicherungen nicht mehr vorgelesen werden.

### 2.3.3.5 Hinweistexte und Administrationsanweisungen

Wie eingangs erwähnt, reagieren Interviewte auf Unklarheiten mit Rückfragen. Daher ist ein wichtiger Bestandteil der Operationalisierung, wie die Interviewer mit Nachfragen umgehen sollen, d. h. ob bzw. welche Erläuterungen sie geben sollen. Dafür wurden in den Fragebogen Hinweistexte integriert (vgl. Abbildung 3). Durch sie standen den Interviewern beispielsweise für Krankheitsbegriffe (z. B. Asthma bronchiale) mögliche Synonyme (z. B. Bronchialasthma, Lungenasthma, allergisches Asthma) unmittelbar zur Verfügung.

Ein weiterer Aspekt der Operationalisierung besteht in der Festlegung, welche Antwortkategorien vorgelesen werden sollen (Dillman 1978). Programmierte Administrationsanweisungen und eine markante graphische Gestaltung gewährleisteten diesbezüglich eine einheitliche Handhabung (Abbildung 3). Andere Administrationsanweisungen drosseln, wie erwähnt, das Interviewtempo. Auch dadurch wurden die Anlässe reduziert, in denen die Interviewer improvisieren mussten (wie von Dillman (1978) gefordert). Der Umgang mit Erläuterungen sowie die Administration der Fragen und Antworten waren aufgrund ihrer

Abbildung 3  
Bildschirmansicht des programmierten Fragebogens



Bedeutung für die Standardisierung der Erhebungssituation zudem Kernbestandteil der Interviewerschulung (Borch, Rieck 2005).

### 2.3.4 Weitere Techniken/Maßnahmen

Neben den bisher genannten Schritten zur Optimierung des Erhebungsinstrumentes kamen für den GSTelo3 eine Reihe weiterer zur Anwendung. Sie wurden aufgrund der Schwerpunktsetzung in diesem Aufsatz (insbesondere Reduktion von Komplexität) bisher nicht thematisiert. Die wichtigsten sollen abschließend kurz erwähnt werden:

- ▶ Tabellen wurden integriert, wenn die selbe Frage für eine Reihe von Items gestellt werden sollte und der Gegenstand sich im Rahmen von Telefoninterviews dafür eignete. Die Inhalte tabellarisch organisierter Fragen wurden sequentiell präsentiert (wie bei Dillman (1978) dargestellt) und nicht simultan wie bei schriftlichen Fragebögen (Hippler et al. 1992).
- ▶ Komfortabel programmierte Listen, z. B. für die Erfassung des Geburtslandes, dienten als Ersatz für eine freitextliche Erfassung.
- ▶ Für die Festlegung der Anordnung von Fragen und Antwortkategorien wurden weitere Aspekte berücksichtigt, bzgl. der Fragen vor allem question order effects (Schuman, Presser 1981; Schwarz et al. 1989, 1991).
- ▶ Eine numerische Skala mit verbalisierten Endpunkten diente zur Erfassung der Stärke der Rückenschmerzen.<sup>8</sup>
- ▶ Die visuelle Gestaltung des Fragebogens unterstützt die Interviewer (Dillman 1978): eine einheitliche Platzierung, Formatierung, Hervorhebung und Formulierung führte dazu, dass Administrationsanweisungen, Hinweise, Fragentexte und Antwortkategorien auf den ersten Blick gefunden, verstanden und richtig vorgelesen wurden.<sup>9</sup>
- ▶ Anmerkungen zu den programmierten Plausibilitätskontrollen finden sich bei Borch, Rieck (2005).

In der Literatur werden zusätzlich andere Maßnahmen diskutiert, die beim GSTelo3 aufgrund der behandelten Themen, eingesetzten Fragearten etc. von nachgeordneter Bedeutung waren. Dazu

gehört beispielsweise bei Meinungsfragen der Einsatz von zufallsgesteuerten Rotationen (Permutation) für Antwortkategorien bzw. Items zur Vermeidung von response order effects (Dillman 1978; Schwarz et al. 1991, 1992).

## 2.4 Fazit

Es ist wichtig, die Besonderheiten der Erhebungssituation, beispielsweise von Telefoninterviews, bei der Entwicklung von Erhebungsinstrumenten zu berücksichtigen. Das bedeutet jedoch nicht, dass jede Erhebungsmethode grundsätzlich andere Arten von Operationalisierungen verlangt. Vielmehr geht es vor allem darum, die allgemeinen Regeln zur Fragenformulierung und Fragebogenentwicklung, die bereits vor Jahrzehnten für face-to-face-Interviews und schriftliche Befragungen entwickelt wurden (König 1957, 1967; Scheuch 1973; Friedrichs 1980), in der Praxis wirklich einzuhalten. Was also in diesem Aufsatz aufgrund der Betonung der Besonderheiten bzw. Akzente verschiedener Erhebungsmethoden in den Vordergrund der Darstellung rückte, ist bei genauer Betrachtung eben häufig keine Abweichung von diesen allgemeinen Kriterien, sondern lediglich ihre konsequente Anwendung. Beispielsweise sind sprachlich überfrachtete oder unklare Frageformulierungen auch bei schriftlichen Befragungen problematisch, auch wenn die Probanden dort mehr Zeit zum Nachdenken haben. Denn es bleibt offen, ob dieses längere Nachdenken zum – im Sinne des Erkenntnisinteresses – richtigen Ergebnis führt. Und Interviewer leisten hier keine Unterstützung. Lange Antwortlisten wirken bei schriftlichen Befragungen ebenfalls ermüdend, was die Gefahr des vorzeitigen Abbruchs und der nachlassenden Sorgfalt bei der Beantwortung dieser Frage in sich birgt. Allerdings ist eine überschaubare Anzahl von Antwortvorgaben für Telefoninterviews eben noch wichtiger als für schriftliche Fragebögen. Hinzu kommt, dass es für Telefonsurveys bessere Möglichkeiten gibt als für schriftliche Befragungen, die

<sup>8</sup> Zum Ersatz von visuellen Analogskalen durch numerische Skalen für Telefonsurveys vgl. z. B. Wüst (1998) und Paice, Cohen (1997).

<sup>9</sup> Gute graphische Gestaltungsmöglichkeiten, die Einbindung langer Listen u.ä. waren zentrale Kriterien für die Auswahl der CATI-Software.

Filtertechnik bei der Zerlegung langer Antwortlisten einzusetzen. Das bedeutet: einzelne Regeln zur Fragebogenentwicklung gelangen in Abhängigkeit vom Erhebungsmodus zu unterschiedlicher Bedeutung.

Für diejenigen Arten von Operationalisierungen, bei denen in Abhängigkeit von der Erhebungsmethode Modifikationen unvermeidlich sind, stellt sich die Frage, in wie fern es sich dabei nicht nur um inhaltlich äquivalente Transformationen handelt, die letztendlich zu dem gleichen Ergebnis führen. Dies zu klären, bleibt Aufgabe zukünftiger Forschung. Wüst (1998) unterstellt eine Abhängigkeit der Stärke des Effektes vom Ausmaß der Modifikation, was plausibel ist. Jedoch kann man auch im Sinne von Dillman (1978) argumentieren, dass erhebungsmethodisch adäquate Transformationen zu äquivalenteren bzw. vergleichbareren Ergebnissen führen, als inadäquate Operationalisierungen. Und davon gehen die Autoren und Autorinnen dieses Bandes aus. Solange eine endgültige Klärung dieser Fragestellung nicht erfolgt ist und eine Gleichheit der Operationalisierungen nicht festgestellt wurde, sollten Studienergebnisse jedoch auch vor dem Hintergrund möglicher Unterschiede zwischen den Erhebungsmethoden interpretiert werden.

## 2.5 Literatur

- Anders M (1990) Praxis der Telefonbefragung. In: Forschungsgruppe Telekommunikation (Hrsg) *Telefon und Gesellschaft*. Band 2. Spiess, Berlin, S 426–436
- Bernard HR, Killworth P, Kronenfeld D, Sailer L (1984) The problem of informant accuracy: The validity of retrospective data. *Annual Review of Anthropology* 13: 495–517
- Borch S, Rieck A (2005) Feldarbeit (in diesem Band)
- Dillman D A (1978) *Mail and Telephone Surveys. The Total Design Method*. Wiley, New York
- Fowler FJ (2001) Why it is easy to write bad questions, *ZUMA-Nachrichten* 48: 49–66
- Frey J H, Kunz G, Lüschen G (1990) *Telefonumfragen in der Sozialforschung*. Westdeutscher Verlag, Opladen
- Frey JH, Oishi SM (1995) *How to conduct interviews by telephone and in person*. Sage, Thousand Oaks
- Friedrichs J (1980) *Methoden der empirischen Sozialforschung*. Westdeutscher Verlag, Opladen
- Friedrichs J (1990) Gesprächsführung im telefonischen Interview. In: Forschungsgruppe Telekommunikation (Hrsg) *Telefon und Gesellschaft*. Band 2. Spiess, Berlin, S 413–425
- Fuchs M, Lamnek S (1990) Fragebogendramaturgie und Abbruchverhalten. *Planung und Analyse* 3: 101–107
- Hippler HJ, Schwarz N (1992) The impact of administration modes on response effects in surveys. *ZUMA*, Mannheim
- Kohler M (2005) Stichprobenkonzeption und Stichprobenrealisierung (in diesem Band)
- König R (Hrsg) (1957) *Das Interview. Formen, Technik, Auswertung*. *Praktische Sozialforschung*, Bd. 1., 2. Auflage. Verlag für Politik und Wirtschaft, Köln
- König R (Hrsg) (1967 ff.) *Handbuch der empirischen Sozialforschung*. 2 Bde. 2. Auflage. Ferdinand Enke Verlag, Stuttgart
- Paice JA, Cohen FL (1997) Validity of a verbally administered numeric rating scale to measure cancer pain intensity. *Cancer Nursing* 20: 88.93
- Scheuch EK (1973) Das Interview in der empirischen Sozialforschung. In: König R (Hrsg) *Handbuch der empirischen Sozialforschung*, Bd. 2, 3. Auflage. Ferdinand Enke Verlag, Stuttgart, S 66–190
- Schuman H, Presser S (1981) *Questions and answers in attitude surveys*. Academic Press, New York
- Schwarz N, Bishop G, Hippler HJ, Strack F (1989) Psychological sources of response effects in self-administered and telephone surveys. *ZUMA-Arbeitsbericht* 89/01, ZUMA, Mannheim
- Schwarz N, Strack F, Hippler HJ, Bishop G (1991) The Impact of Administration Mode on Response Effects in Survey Measurement. *Applied Cognitive Psychology* 5: 193–212
- Schwarz N, Sudman S (Eds.) (1994) *Autobiographical memory and the validity of retrospective reports*. Springer, New York, Heidelberg
- Wüst AM (1998) Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften als Telefonumfrage. *ZUMA-Arbeitsbericht* 98/04, ZUMA, Mannheim
- Ziese T (2005) Der telefonische Gesundheitssurvey 2003 – Baustein zum Gesundheitsmonitoringsystem für Deutschland (in diesem Band)



## 3 Feldarbeit

*Susan Borch, Angelika Rieck*

### 3.1 Einleitung

Gesundheitssurveys haben am Robert Koch-Institut eine Tradition, die bis in die 1980er Jahre zurückreicht. Während die früheren Studien aus Examinationsmodulen, selbst auszufüllenden Probanden-Fragebögen und computergestützten ärztlichen Interviews bestanden, handelt es sich beim telefonischen Gesundheitssurvey 2003 (GSTelo3) um die erste Studie, die mittels computergestützter Telefoninterviews (CATI, computer-assisted telephone interviewing) durchgeführt wurde.

Gegenüber anderen Erhebungsformen weist CATI in Bezug auf die Feldarbeit wichtige Vorteile (Weitkunat 1998) auf: Durch die unmittelbare elektronische Erfassung der Daten werden eine komfortable Filterführung der Fragen (Rieck et al. 2005) und Plausibilitätsprüfungen während der Dateneingabe ermöglicht. Außerdem entfällt durch die Arbeit im Telefonstudio das aufwändige Aufsuchen bzw. Einladen der Befragten, was auch die Qualitätssicherung der Interviewführung sehr erleichtert.

Bedingt durch das neue Design (Ziese 2005) musste das Vorgehen bei der Feldarbeit in vielen Punkten neu entwickelt werden, welches im vorliegenden Artikel mit einem speziellen Fokus auf Schulung und Qualitätssicherung dargestellt werden soll. Denn Datenqualität, Ausschöpfung<sup>1</sup> und Item-Nonresponse werden durch das Interviewerverhalten beeinflusst.

Außerdem werden das eingesetzte Personal sowie die Datenerhebung und -haltung beschrieben.

### 3.2 Personal

Das Feldteam bestand aus folgenden Personengruppen: Drei wissenschaftliche Mitarbeiter gehörten zur Feldleitung, die insbesondere verantwortlich war für die Optimierung des Erhebungsinstrumentes unter methodischen Aspekten (Rieck et al. 2005), die Stichprobenkonzeption (Kohler 2005), die Schulung der Interviewer und

Supervisoren, Qualitätssicherung, Erstellung des Operationshandbuchs, fachliche Aufsicht sowie die Organisation und Koordination der Feldarbeit.

Mit der Supervision der Interviewer wurden fünf studentische Hilfskräfte betraut. Ihre Aufgaben bestanden vor allem in der Anleitung der Interviewer während der Datenerhebung, Qualitätssicherung der Interviewführung, Berichterstattung an die Feldleitung, Unterstützung der Interviewerschulung, Rekrutierung sowie der Organisation des Interviewereinsatzes. Dabei wurden Studentinnen und Studenten sozialwissenschaftlicher Fächer mit Interviewerfahrungen verpflichtet, die auch über ein hohes Maß an sozialer Kompetenz, Flexibilität und Einsatzbereitschaft verfügen.

Als Interviewerinnen und Interviewer wurden Honorarkräfte eingesetzt, welche sich hauptsächlich aus Studierenden sozialwissenschaftlicher Fächer rekrutierten. Die Kontaktaufnahme erfolgte telefonisch, um die Eignung z. B. anhand Stimmmodulation und Sprechtempo beurteilen zu können (Friedrichs 1990). Es kam zwar vor, dass Bewerber aufgrund eines starken Dialektes abgelehnt wurden; insgesamt spielte es jedoch eine größere Rolle, welchen Eindruck sie im Bewerbungsverfahren bezüglich ihrer Zuverlässigkeit und kommunikativen Kompetenz machten.

Alle im Projekt beschäftigten Mitarbeiterinnen und Mitarbeiter wurden hinsichtlich des Datenschutzes informiert und verpflichtet.

### 3.3 Datenerhebung und Datenhaltung

Die Feldzeit des GSTelo3 begann am 2. September 2002 und endete am 14. März 2003. In diesem Zeitraum wurden von insgesamt 81 Interviewern

<sup>1</sup> Die Berechnung von Ausschöpfung bzw. Response ist ein mehrstufiger Prozess, bei dem alle neutralen und nichtneutralen Ausfallgründe berücksichtigt werden. Diese Berechnung ist bei CATI-Studien im Kontext zufallsgenerierter Telefonnummern nach dem Gabler-Häder-Design (Gabler, Häder 1998) komplizierter als bei registerbasierten Auswahlverfahren (Kohler 2005).

8.362 Frauen und Männer befragt, davon waren 8.318 Interviews verwertbar.<sup>2</sup> Die Geschlechtsverteilung im Interviewerstamm war ausgewogen, zum größten Teil handelte es sich um Studierende zwischen 20 und 30 Jahren.

Es wurden zwei Pretest durchgeführt, um zu prüfen, ob die Fragen handhabbar waren, die Interviewerhinweise ausreichten, die disposition codes<sup>3</sup> alle Fälle abdeckten und ob das Schulungskonzept praktikabel war. Dabei wurde u. a. festgestellt, dass die Lerninhalte durch die Aufteilung in einen »Theorietag« und einen »Praxistag« zu abstrakt bleiben, was zu einem erhöhten Aufwand für Qualitätssicherung und Nachschulung führte (vgl. Borch et al. 2004). Daher wurde die Schulung für die Hauptphase neu konzipiert; außerdem wurde das Erhebungsinstrument überarbeitet, um die Handhabbarkeit für die Interviewer zu verbessern (Rieck et al. 2005).

Die Datenerhebung erfolgte im neu geschaffenen CATI-Labor des Robert Koch-Institutes, welches 19 Interviewerarbeitsplätze und einen Supervisorarbeitsplatz umfasst. Die Anrufe wurden vorwiegend wochentags zwischen 16 und 20 Uhr durchgeführt, zusätzlich fanden einige Vormittags- und Samstagseinsätze statt, um die Responserate für die in den oben genannten Zeiten schlecht erreichbaren Personen zu erhöhen. Während der Datenerhebung waren zwei Supervisoren als Ansprechpartner für die Interviewer sowie für die Qualitätssicherung der Interviewführung im CATI-Labor anwesend. Außerdem stand für Rückfragen der Supervisoren ein Mitglied der Feldleitung zur Verfügung, welches auch weitere qualitätssichernde Maßnahmen durchführte, wie z. B. statistische Zwischenauswertungen der Daten sowie die Qualitätssicherung der Supervision.

Für die Datenerhebung wurde eine spezielle CATI-Software eingesetzt (Kohler 2005). Ein Modul steuerte z. B. die An- und Rückrufe, wobei die Datensätze zufällig unter den Interviewern verteilt wurden und der Anwahlvorgang automatisch erfolgte (auto-dialing). Ein weiteres Modul diente zur Beobachtung der Interviewerproduktivität und anderer Kenngrößen, d. h. es zeigte interviewerspezifisch die Zahl der vollständigen Interviews, vollständigen Non-Responderinterviews und Abbrüche sowie die durchschnittliche Gesprächsdauer an.

Im Fragebogen-Modul waren die disposition codes, die Einleitungstexte sowie die Items mit

Antwortkategorien und Interviewerhinweisen programmiert. Die mit diesem Modul umgesetzte komfortable Filterführung steuerte den Ablauf der Variablen und begrenzte dadurch auch die Dauer der Befragung für den einzelnen Probanden auf ein Minimum (Rieck et al. 2005). Außerdem erfolgte mit Hilfe entsprechend programmierter Regeln eine Prüfung auf unplausible Daten, die hauptsächlich durch Tipp- oder Verständnisfehler der Interviewer bzw. durch (erinnerungsbedingt) widersprüchliche Angaben der Probanden entstehen: Die Interviewer wurden sofort auf Inkonsistenzen hingewiesen, sodass Korrekturen unmittelbar nach der Fehleingabe möglich waren. Dadurch konnte die Qualität der erfassten Daten erheblich besser gewährleistet werden, als bei den zeitversetzt erfolgenden und daher stets auf reduzierten Informationen basierenden Korrekturen anderer Erhebungsverfahren.

Die durchschnittliche Länge der Interviews im GSTelo3 betrug zirka 21 Minuten. Diese variierte jedoch erheblich für bestimmte Befragten-Gruppen: Jüngere Probanden absolvierten die Befragung in der Regel schneller als ältere, Frauen langsamer als Männer.<sup>4</sup>

2 44 Interviews wurden im Zuge der Datenaufbereitung aus dem Enddatensatz entfernt. Dabei handelt es sich um Datensätze mit dokumentierten Fehlern (z. B. nachträglich festgestellte Nichtzugehörigkeit zur Grundgesamtheit, falsche Erfassung von Alter bzw. Geschlecht) und/oder dokumentierten gravierenden Verständnisproblemen der Befragten bezüglich der gestellten Fragen (z. B. aufgrund von Alter oder Bildungsbarrieren).

3 Die disposition codes sind ein Klassifikationsschema zur Erfassung aller vorkommenden Anwahlergebnisse (z. B. Freizeichen, Terminvereinbarung, Interview vollständig, Ansagen wie »kein Anschluss unter dieser Nummer«). In ihnen wird der Status eines Datensatzes zur Steuerung der Rückrufe festgehalten. Die verwendeten disposition codes wurden speziell für das Projekt entwickelt und sind vergleichsweise differenziert (vgl. Kohler 2005).

4 Die durchschnittliche Interviewdauer variiert nach Altersgruppen von ca. 18 Minuten (18- bis 20-Jährige) bis ca. 30 Minuten (über 80-Jährige) und nach Geschlecht zwischen ca. 20 Minuten (Männer) und ca. 22 Minuten (Frauen).

Bei Vorhandensein bestimmter Krankheiten stieg die durchschnittliche Interviewdauer, da in diesem Fall filterbedingt mehr Fragen gestellt wurden (Rieck et al. 2005).<sup>5</sup>

Zur datenschutzgerechten Umsetzung der Studie gehörte einerseits die entsprechende Handhabung der Daten: Die Interviewsoftware war lokal auf den CATI-Arbeitsplätzen installiert, die Datenspeicherung erfolgte zentral und passwortgeschützt auf einem auch physisch gegen unbefugte Zugriffe gesicherten Server. Der Zugang zum CATI-System wurde beschränkt und dokumentiert. Die Verwaltung der Daten erfolgte ausschließlich anonym über die fortlaufende Datensatznummer. Außerdem wurden die Befragten vor Beginn des Interviews sowie vor der Abfrage soziodemografischer Daten darüber informiert, dass ihre Angaben freiwillig sind, nur anonymisiert ausgewertet werden und sie auch die Möglichkeit haben, einzelne Fragen nicht zu beantworten.<sup>6</sup>

### 3.4 Mitarbeiterschulung

Die Schulung von Interviewern und Supervisoren hat eine zentrale Stellung innerhalb der Feldarbeit: Die Interviewer sollen bei der Kontaktaufnahme und bei der Befragung flexibel und aufmerksam auf den Gesprächspartner eingehen sowie souverän und seriös wirken, um die Motivation zu erhalten und Abbrüche zu vermeiden. Gleichzeitig sollen sie sich an klar definierte Regeln halten, wie z. B. die vorgegebenen Frage korrekt vorzulesen und ausschließlich erlaubte Hinweise zu geben. Die Interviewer haben also in der Erhebungssituation zwei gegensätzliche Rollen auszufüllen (Friedrichs 1990), auf die sie in der Schulung möglichst umfassend und realitätsnah vorbereitet werden müssen.

Die Supervisoren haben aufgrund ihrer Handlungsanweisungen an die Interviewer eine wichtige Funktion bei der Umsetzung der standardisierten Erhebungssituation. Zu ihren Aufgaben gehört die ständige Betreuung und Nachschulung der Interviewer im Rahmen der Qualitätssicherung. Außerdem berichten sie der Feldleitung über den Prozess. Damit die Supervisoren die methodische Relevanz von Ereignissen angemessen einschätzen können, sind kontinuierliche Schulung und Betreuung erforderlich.

#### 3.4.1 Interviewer

Zu Beginn der Feldphase wurden mehrere Wochen lang verschiedene Teilnehmergruppen geschult, bis ein ausreichender Interviewerstamm aufgebaut war. Bei Bedarf fanden jedoch auch später einzelne Lehrgänge statt. Sie wurden von zwei Mitgliedern der Feldleitung durchgeführt und durch Supervisoren unterstützt.

Jeder Bewerber hatte eine intensive 15-stündige Eingangsschulung zu absolvieren, denn diese ist für die Qualität der Daten (Billiet, Loosveldt 1988) sowie für die Optimierung der Ausschöpfung (Kohler 2005) von großer Bedeutung. Zuerst besuchten sie zwei sechstündige Schulungstage, bei denen sich Theorie- und Praxisblöcke abwechselten. Danach arbeiteten die neuen Interviewer zwei bis drei Abende unter enger Betreuung der Supervisoren und der Feldleitung. Schließlich wurden in einer dreistündigen Sitzung noch präzisierete Vorgehensregeln vermittelt, die erst mit der Praxiserfahrung nachvollziehbar waren. Hier wurden u. a. die bisherigen Erfahrungen der Interviewer in der Gruppe diskutiert und eventuell aufgetretene Unklarheiten beseitigt. Nach jedem inhaltlichen Block wurden Schulungsmaterialien ausgeteilt, die den behandelten Stoff vertiefen und die Teil des Operationshandbuchs (Borch et al. 2004) sind.

5 Die beiden umfangreichsten Bereiche des Fragebogens bei den Krankheiten betreffen Diabetes und Rückenschmerzen: Bei Probanden, die weder Diabetes noch Rückenschmerzen haben, betrug die durchschnittliche Interviewdauer ca. 20 Minuten. Bei ausschließlichem Vorliegen von Rückenschmerzen stieg dieser Wert auf ca. 23 Minuten, bei ausschließlichem Vorliegen von Diabetes auf ca. 26 Minuten. Bei Vorhandensein beider Erkrankungen lag die Dauer bei ca. 31 Minuten. Bei diesen Krankheiten ist zusätzlich zu berücksichtigen, dass auch innerhalb der Fragenbereiche – in Abhängigkeit von den greifenden Filtern – erhebliche Unterschiede hinsichtlich der Dauer auftraten.

6 In wenigen Fällen wurde die Erlaubnis zur Datenspeicherung von Probanden nachträglich widerrufen. Ihre Angaben wurden gelöscht.

Ziel der Schulung war die Vermittlung von Qualitätsstandards; die Lerninhalte lassen sich in folgenden Punkten zusammenfassen:

- ▶ Hintergrund der Studie, Auftraggeber und Zielsetzung,
- ▶ Aufbau, Inhalt und Besonderheiten des Erhebungsinstrumentes,
- ▶ Sicherer technischer Umgang mit der CATI-Software,
- ▶ Vorgehen in der Kontaktphase des Interviews incl. angemessenes Auftreten, Argumentationsstrategien, korrekter Durchführung der Zielpersonenermittlung und zweckmäßigem Umgang mit den disposition codes,
- ▶ Vollständige, aussagekräftige und datenschutzgerechte Dokumentation der Zielpersonenermittlung,<sup>7</sup>
- ▶ Angemessenes Auftreten gegenüber den Befragten im weiteren Verlauf des Interviews,
- ▶ Angemessener Umgang mit Antworten wie »weiß nicht« und »keine Angabe«,<sup>8</sup>
- ▶ Vollständige, aussagekräftige und datenschutzgerechte Dokumentation von Informationen über mangelnde Qualität der Antworten bezüglich einzelner Fragebogenitems, kompletter Interviews oder nachträglich festgestellter Fehler, wie die Nicht-Zugehörigkeit zur Grundgesamtheit,
- ▶ Konstanthalten der Erhebungssituation.<sup>9</sup>

Für die Qualität der erhobenen Daten ist der letzte Punkt von essentieller Bedeutung; dieser wurde daher besonders vertiefend behandelt. Zum Konstanthalten der Erhebungssituation zählen so grundlegende Dinge, wie die Interviews ausschließlich in deutscher Sprache zu führen, die persönlichen Einstellungen gegenüber den Befragten zu verbergen sowie auf eigenmächtige Interpretationen und Erläuterungen von Fragen zu verzichten. Im programmierten Fragebogen waren daher präzise Interviewerhinweise dazu enthalten, welche Erklärungen bei Nachfragen der Probanden gegeben werden dürfen (Rieck et al. 2005).

Außerdem wurden allgemeine Regeln vermittelt, wie die Interviewer mit unerwarteten Rückfragen, d. h. nicht in den Interviewerhinweisen berücksichtigten Aspekten, umgehen sollten. So galt es die Interviewer dafür zu sensibilisieren, dass jede Frage wörtlich zu verstehen ist und nicht

thematisierte Aspekte daher auch nicht in die Beantwortung einfließen sollen. Beispiel: Die Frage zu Rückenschmerzen in den letzten 12 Monaten enthält keine Eingrenzung auf bestimmte Ursachen, folglich gehören ausschließlich während einer Schwangerschaft aufgetretene Rückenbeschwerden ebenfalls dazu.

Gegenstand war zudem das Vorlesetempo und das Gewähren von ausreichend Zeit zum Überlegen für die Befragten. Letzteres war bei der Handhabung schwieriger, zeitbeanspruchender Fragen, z. B. nach Häufigkeiten oder Mengen (Rieck et al. 2005) relevant. Die Interviewer sollten in diesem Zusammenhang unter anderem eine entsprechende Sensibilität bezüglich der Unsicherheitsäußerungen von Probanden entwickeln und lernen situationsadäquat zu reagieren (Dillman 1978).

Daneben wurde der Verzicht auf Interpretation der Antworten durch die Interviewer trainiert. Dies bedeutete z. B., dass sie keine Zahlen- oder Zeitangaben selbst errechnen bzw. runden oder Antworten eigenmächtig den Kategorien zuordnen durften. Zum Beispiel bestand die Aufgabe der Interviewer bei der Frage nach dem allgemeinen Gesundheitszustand nicht darin, eine von den Kategorien abweichende Antwort, etwa »für mein Alter ganz gut«, zu interpretieren und zuzuordnen, sondern darin, die ältere Probandin um die Auswahl einer der fünf Vorgaben zu bitten und diese dafür ggf. nochmals vorzulesen.

Zur Optimierung der Ausschöpfung (Kohler 2005) wurden der korrekte Umgang mit den umfangreichen disposition codes und das Vorgehen in der Kontaktphase mit der Zielpersonenermittlung ausführlich besprochen und trainiert. Gegenstand war ebenfalls die Vermittlung eines im Pretest erprobten Vorgehens in der Kontakt-

7 Sie ist notwendig, wenn das Interview zu einem anderen Zeitpunkt geführt werden soll.

8 Die Antwortvorgaben »weiß nicht« und »keine Angabe« wurden von den Interviewern nicht vorgelesen, eine von den Probanden entsprechend geäußerte Antwort jedoch akzeptiert. Waren sich die Befragten unschlüssig darüber, ob sie eine Frage beantworten wollen, wiesen die Interviewer einerseits auf die Freiwilligkeit der Angaben und andererseits auf die anonymisierte Auswertung der Daten hin.

9 Die zum Beispiel bei Prüfer und Stiegler (2002) dargestellten allgemeinen Regeln zur Standardisierung der Erhebungssituation wurden für den GSTelo3 adaptiert und präzisiert.



phase, welches vor allem auf die eindeutige Abgrenzung von anderen telefonischen, insbesondere kommerziellen Befragungen abhebt. Dieses bestand hauptsächlich in der namentlichen Vorstellung des Interviewers, der Nennung des Robert Koch-Institutes als verantwortliche und des Bundesministerium für Gesundheit und Soziale Sicherung als auftraggebende Institution sowie der Betonung des wissenschaftlichen Ziels der Studie. Wichtig waren in diesem Zusammenhang auch Argumentationsstrategien zur zusätzlichen Information unsicherer Befragter und der Hinweis auf weitere Informationsmöglichkeiten hinsichtlich der Studie über die Internetseiten des Robert Koch-Institutes sowie mittels einer Rückrufnummer, unter der ein Supervisor Fragen beantwortete und ggf. das Gespräch zur Feldleitung durchstellte. Da es sich mit durchschnittlich 21 Minuten um ein relativ langes Interview handelte, wurde besondere Rücksicht auf das Zeitbudget der Befragten genommen, indem bei Bedarf angeboten wurde, einen günstigeren Termin zu vereinbaren oder ein bereits begonnenes Interview zu unterbrechen und zu einem späteren Zeitpunkt fortzusetzen.

Als Schulungsmethoden kamen Lehrgespräche, Gruppendiskussionen, praktische Arbeit mit dem Interviewprogramm sowie selbständig zu bearbeitende Übungsaufgaben zum Einsatz. Im Rollenspiel wurden unter Einbeziehung aller Teilnehmerinnen und Teilnehmer typische Fragen und Probleme simuliert und analysiert. Bei Probeinterviews befragten sich die angehenden Interviewer unter Echtbedingungen zunächst gegenseitig, um im zweiten Schritt dann Freunde und Bekannte anzurufen, um das Erlernte nochmals anzuwenden.

Teilweise hatten die Bewerber Vorerfahrungen, die bei der Gestaltung der Schulung berücksichtigt werden mussten. Diese reichten von qualitativer Interviewführung bis zu CATI im Rahmen von Markt- und Meinungsforschung. Erfahrene Schulungsteilnehmer verfügten mehrheitlich über sehr gut ausgeprägte kommunikative Fähigkeiten sowie Souveränität bei der Interviewanbahnung, die unerfahrene erst erwerben mussten. Andererseits handhaben sie beispielsweise disposition codes und Verweigerungen teilweise anders als für diese Studie gefordert. Es wurde daher darauf geachtet, die Trainingsbedürfnisse aller Teilnehmergruppen abzudecken.

### 3.4.2 Supervisoren

Die studentischen Hilfskräfte wurden schon früh in die Projektplanung und –durchführung einbezogen, da die Mitarbeit bei den Feldvorbereitungen auch eine implizite Schulungsfunktion erfüllt. Die konkreten Tätigkeiten im Kontext der Supervision wurden während des Pretestes und Hauptfeldes bzw. bei deren Vorbereitung erlernt. Schwerpunkte der Supervisorenschulung waren neben der Qualitätssicherung, Dokumentation, Betreuung und Nachschulung der Interviewer auch inhaltliche Hintergründe von Items und methodische Aspekte der Studie. Zusätzlich galt es, den Umgang mit der relevanten Software, die projektspezifischen Verwaltungsabläufe sowie ein angemessenes Auftreten gegenüber den Interviewern zu vermitteln. Während der gesamten Studiendauer erfolgte eine kontinuierliche Weiterbildung der Supervisoren im Rahmen der Qualitätssicherung.

Als Schulungsmethoden sind vor allem die persönliche Anleitung durch die Feldleitung als training on the job, die Durchführung von Interviews zu Test- und Schulungszwecken, Gruppendiskussionen im Rahmen der wöchentlichen Teamsitzungen sowie die Lektüre des Operationshandbuchs und anderer Literatur zu nennen.

### 3.5 Qualitätssicherung

Die Feldarbeit wurde kontinuierlich anhand definierter Kriterien kontrolliert, um daraus konkrete Maßnahmen für die Qualitätssicherung abzuleiten. Hierbei wurden Strategien zur Ermittlung von Verbesserungsbedarf sowie zur Implementierung von Optimierungen in den Routinebetrieb kombiniert. Nachfolgend werden die eingesetzten qualitätssichernden Strategien bezüglich der Eingangsschulung, Interviewführung und Supervision dargestellt. Eine wichtige Grundlage der Qualitätssicherung war das Operationshandbuch (vgl. Borch et al. 2004). Es stellte die Anleitung zum standardisierten Vorgehen bei der Studierendurchführung dar und diente als Schulungsmaterial und Nachschlagewerk. In zielgruppenspezifischen Teilen wurden die unterschiedlichen an der Datenerhebung beteiligten Gruppen (Interviewer, Supervisoren, Feldleitung) je nach Aufgaben,

**Tabelle 1**  
**Checkliste zur Qualitätssicherung der Interviewführung**

Datum: \_\_\_\_\_ Interviewer/in: \_\_\_\_\_ Supervisor/in: \_\_\_\_\_

Interviewtechniken					
Wählt die Zielperson (ZP) korrekt aus	1	2	3	4	5
Führt das Auswahlverfahren der ZP reibungslos aus	1	2	3	4	5
Hält sich wörtlich an die vorgegebenen Texte	1	2	3	4	5
Liest die Fragen vollständig vor	1	2	3	4	5
Liest die Antwortkategorien vollständig vor	1	2	3	4	5
Führt das Interview in einem guten Tempo	1	2	3	4	5
Interview ist fließend	1	2	3	4	5
Lässt die ZP nicht abschweifen	1	2	3	4	5
Motiviert die ZP weiterzumachen	1	2	3	4	5
Erklärt situationsbedingt	1	2	3	4	5
Hält sich bei Erläuterungen an die vorgegebenen Hinweise	1	2	3	4	5
Formgerechte Verabschiedung	1	2	3	4	5
<i>Freitextliche Erläuterung zur Bewertung der Interviewtechniken:</i>					
_____					
Nachfragen					
Fragt situationsbedingt nach	1	2	3	4	5
Stellt keine Suggestivfragen	1	2	3	4	5
Besteht auf exakte Antwortkategorien	1	2	3	4	5
Antwortet auf Nachfragen inhaltlich richtig	1	2	3	4	5
Drückt sich auf Nachfragen hin verständlich aus	1	2	3	4	5
<i>Freitextliche Erläuterung zur Bewertung des Nachfrageverhaltens:</i>					
_____					
Verhalten					
Ist höflich und zuvorkommend	1	2	3	4	5
Klingt sicher	1	2	3	4	5
Klingt nicht gelangweilt	1	2	3	4	5
Spricht Fremdwörter korrekt aus	1	2	3	4	5
Hat eine deutliche Aussprache	1	2	3	4	5
Treibt den Probanden nicht zur Eile an	1	2	3	4	5
Klingt nicht sarkastisch	1	2	3	4	5
Zeigt Interesse gegenüber ZP	1	2	3	4	5
Lässt sich von ZP stilistisch nicht beeinflussen	1	2	3	4	5
<i>Freitextliche Erläuterung zur Bewertung des Verhaltens:</i>					
_____					
<i>Traten Probleme auf?</i>					
_____					
<i>Wie ist der Gesamteindruck?</i>					
_____					

Würdest Du gerne von diesem Interviewer befragt werden?

Ja

Nein

Vorkenntnissen und ihrer Einbindung in den Gesamtprozess gezielt angesprochen. Für die Qualitätssicherung war auch die kontinuierliche Dokumentation, z. B. in einem Logbuch, durch die Supervisoren wichtig. Außerdem verlangt die Optimierung der Feldarbeit die kontinuierliche und systematische Reflexion des Vorgehens. In diesem Kontext hat sich z. B. der Austausch zwischen Supervisoren und Feldleitung in der wöchentlichen Teamsitzung bewährt.

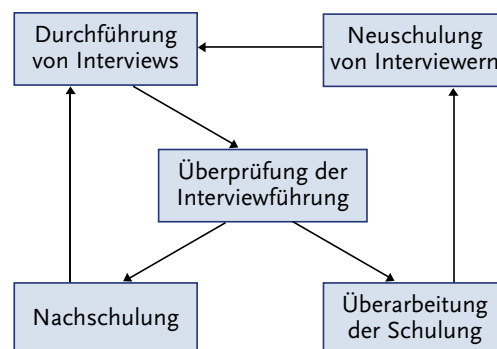
### 3.5.1 Interviewführung und Schulung

Qualität bezüglich der Interviewführung wurde als kontinuierliche Einhaltung der in der Schulung vermittelten Qualitätsstandards und als Befolgung der Datenschutzbestimmungen definiert.

Die Strategien zur Ermittlung von Verbesserungsbedarf bei der Interviewdurchführung fußen auf verschiedenen Säulen. So wurde Wert auf eine konstruktive und kooperative Atmosphäre gelegt, in der die Optimierung der Datenerhebung im Vordergrund stand,<sup>10</sup> denn die Betreuung durch die Supervisoren sollte als Hilfestellung bei Problemen und nicht als Überwachung empfunden werden.<sup>11</sup> In diesem Arbeitsklima machten die Interviewer häufig von der Möglichkeit Gebrauch, bei Unklarheiten im Umgang mit komplexen Situationen nachzufragen.

Weiterhin werteten die Supervisoren mehrmals täglich die von einem Modul der CATI-Software generierten interviewerspezifischen Kenngrößen aus. Dabei wurden vor allem überdurchschnittliche Werte (z. B. eine extrem kurze Interviewdauer oder eine auffällig hohe Anzahl von Abbrüchen) als Indikator für Nachschulungsbedarf berücksichtigt. Bei derartigen Anzeichen, aber auch als Routine in regelmäßigen Intervallen, erfolgte eine Bewertung der Interviewführung durch die Supervisoren unter Einsatz einer speziellen Checkliste. Diese war angelehnt an die des Behavioral Risk Factor Surveillance System (CDC 1998), wurde jedoch für die Bedingungen des GSTelo3 weiterentwickelt. In dieser Checkliste (vgl. Tabelle 1) haben die Supervisoren Noten von eins bis fünf für einzelne Bewertungsdimensionen vergeben, wobei eins den besten und fünf den schlechtesten Wert darstellt. Um die Entwicklung eines Interviewers nachvollziehen zu

Abbildung 1  
Qualitätszyklus der Interviewführung und Schulung



können, war es hier auch wichtig, freitextliche Erläuterungen zur Bewertung festzuhalten. Schließlich wurden von der Feldleitung regelmäßig Zwischenauswertungen durchgeführt, um Auffälligkeiten und Interviewereffekte zu lokalisieren. Hierfür wurden die Daten dahingehend ausgewertet, ob sich die Antworten, die einzelne Interviewer eingeben, signifikant von denen ihrer Kolleginnen und Kollegen unterscheiden.

Neben diesen im Ablauf festgelegten Maßnahmen zur Qualitätssicherung (vgl. Abbildung 1) haben die Supervisoren natürlich auch während anderer Tätigkeiten, wie der Dokumentation, ständig mit »halben Ohr« die Interviews verfolgt, was quantitativ die größte Rolle für die Aufdeckung von Nachschulungsbedarf spielte. Diese Nachschulungen bestanden aus einem kurzem lobenden oder kritisierenden, aber stets motivierenden Feedback zwischen zwei Interviews. Diese wurden

<sup>10</sup> Daher erfolgte die Vergütung der Interviewer im GSTelo3 auch durch eine Mischkalkulation von festem Stundenlohn für das Durchführen von Anwahlversuchen und einem Leistungsanteil für vollständige Interviews und Nonresponder-Interviews. Durch den Leistungsanteil sollte einerseits ein Anreiz zum effizienten Arbeiten gesetzt werden, andererseits sorgte das Fixum dafür, dass die Interviewer qualitäts- und nicht quantitativ orientiert arbeiteten. Die Honorierung der Anwahlversuche ist notwendig, um eine gute Ausschöpfung der zufallsgenerierten Telefonnummern zu erreichen. Dafür mussten die Interviewer motiviert werden, auch wenig erfolgversprechende Telefonnummern mehrfach (entsprechend den Regeln für das Anruf- und Rückrufmanagement (Kohler 2005)) anzuwählen.

<sup>11</sup> Die Förderung einer entspannten, kooperativen Arbeitsatmosphäre im CATI-Studio als Ausdruck der Wertschätzung gegenüber den Interviewern wird auch von Niehoff (1998) als wichtig befunden.

von den Supervisoren dokumentiert. Zeigte sich in der Dokumentation eine Häufung von ähnlichen Unklarheiten und Auffälligkeiten bei mehreren Interviewern, wurden Maßnahmen ergriffen, um diese künftig zu reduzieren beziehungsweise zu vermeiden. Dazu gehörte einerseits die Nachschulung der sich im Einsatz befindlichen Interviewer als Gruppe vor Arbeitsbeginn, die überwiegend in den ersten Monaten der Feldzeit vorkam. Andererseits sind die im Rahmen der Qualitätssicherung gemachten Erfahrungen als didaktische Weiterentwicklung und verbesserte inhaltliche Schwerpunktsetzung in Konzept und Unterlagen für künftige Eingangsschulungen eingeflossen. Damit sind die Qualitätssicherung der Interviewführung und der Schulung in einem gemeinsamen Prozess eng miteinander verknüpft (vgl. Abbildung 1).

Angestrebt wurde, dass die zunehmende Verbesserung der Eingangsschulung von Interviewern einen positiven Einfluss auf die Qualität der von ihnen durchgeführten Interviews hat. Jedoch kann in diesem Rahmen nicht auf alle Details eingegangen werden, die für die Interviewer erst mit der Praxiserfahrung nachvollziehbar werden. Daher war die kontinuierliche Nachschulung von Interviewern während der gesamten Feldzeit nicht durch einen optimierten Grundkurs ersetzbar. Alle neu aufgetretenen inhaltliche Probleme wurden dokumentiert, ihre Lösung verbindlich geregelt und das konkretisierte Vorgehen durch die Formulierung von Praxisbeispielen in einer Liste der am häufigsten gestellten Fragen veranschaulicht.<sup>12</sup> Es kam vor, dass einzelne Interviewer nicht mehr eingesetzt wurden, z. B. wenn sie trotz mehrmaliger Nachschulung nicht in der Lage waren, die Qualitätsstandards umzusetzen, sich uneinsichtig zeigten oder unzuverlässig waren. Dies waren jedoch Ausnahmen; teilweise stellten sich die Mitarbeiterinnen und Mitarbeiter mit den größten »Anlaufschwierigkeiten« und anfangs überdurchschnittlichem Betreuungsaufwand mit zunehmender Übung als sehr gute Interviewer heraus.

### 3.5.2 Supervision

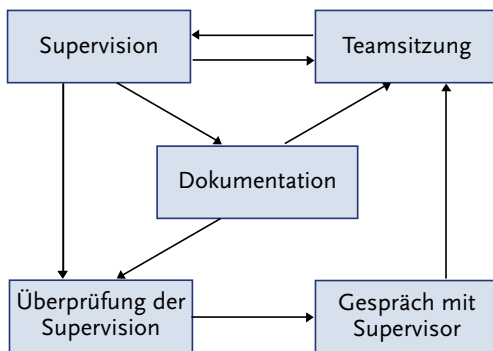
Qualität in der Supervision wurde wie folgt definiert:

- ▶ Regelmäßige und sachgemäße Qualitätssicherung bezüglich der Interviewführung,
- ▶ Erteilung einheitlicher und sachgemäßer Anweisungen, vor allem während der Nachschulung,
- ▶ Motivierender Kommunikationsstil gegenüber den Interviewern, insbesondere im Rahmen der Nachschulung,
- ▶ Vollständige Durchführung der täglich vorzunehmenden technischen Vorbereitungen und Tests,
- ▶ Vollständige und aussagekräftige Dokumentation technischer Probleme, inhaltlicher Unklarheiten und des Nachschulungsbedarfs von Interviewern,
- ▶ Sensibilität gegenüber den (insbesondere methodischen) Konsequenzen von Ereignissen und Vorgehensweisen,
- ▶ Einhaltung des Datenschutzes bezüglich der Befragten- sowie der Interviewerdaten.

Die Qualitätssicherung erfolgte auf zwei Wegen: erstens mittels Überprüfung der Supervision durch Mitglieder der Feldleitung. Dafür waren diese täglich einige Zeit im CATI-Labor anwesend, beobachteten die Supervisoren bei ihrer Arbeit, sichteten die Dokumentationsmittel und verschafften sich so einen Überblick über die Einhaltung der oben genannten Kriterien. Zweitens berichteten die Supervisoren selbständig über den allgemeinen Schichtverlauf, aber auch über Klärungs- und Verbesserungsbedarf. Durch die Diskussion der alltäglichen Vorkommnisse in der wöchentlichen Teamsitzung erfolgte eine kontinuierliche Weiterbildung der Supervisoren (Abbildung 2).

<sup>12</sup> Beispiel: Ist eine Herzmuskelsuffizienz auch eine Herzinsuffizienz? Antwort: Ja.

**Abbildung 2**  
Qualitätszyklus Supervision



### 3.6 Fazit

Die Methode des computergestützten Telefoninterviews bietet ausgezeichnete Ansatzpunkte für eine systematische, beständige und effektive Steuerung der Datenerhebung und damit zur Aufrechterhaltung einer konstanten Erhebungssituation. Voraussetzungen für erfolgreiche Telefoninterviews sind sowohl die gründliche Vorbereitung der Interviewer und Supervisoren als auch ihre aufmerksame Begleitung in der Feldphase.

Nur so ist z. B. der vorgegebene Umgang mit den disposition codes und den Hinweisen zur Interviewführung zu gewährleisten. Nur so kann sensibel auf Qualitätsschwankungen eingegangen und rechtzeitig auf Schwierigkeiten reagiert werden.

CATI ist insofern keine einfach anwendbare, quasi automatisierbare Erhebungsroutine, sondern ein sensibles Instrument, dessen Vorteile nur durch erhebliche Fachkompetenz, Qualitätsbewusstsein und viel Engagement der Beteiligten ausgeschöpft werden können.

### 3.7 Literatur

- Billiet J, Loosveldt G (1988) Improvement of the quality of responses to factual survey questions by interviewer training. *Public Opinion Quarterly* 52: 190–211
- Borch S, Rieck A, Blümel M (2004) *Operationshandbuch erster Telefonischer Bundes-Gesundheitssurvey 2003*, Robert Koch-Institut, Berlin
- CDC (1998) *Behavioral Risk Factor Surveillance System Users Guide*. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Atlanta
- Dillman D A (1978) *Mail and Telephone Surveys. The Total Design Method*. Wiley, New York
- Friedrichs J (1990) Gesprächsführung im telefonischen Interview. In: Forschungsgruppe Telekommunikation (Hrsg) *Telefon und Gesellschaft. Band 2*. Spiess, Berlin, S 413–425
- Häder S, Gabler S (1998) Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. In: Gabler S, Häder S, Hoffmeyer-Zlotnik JHP (Hrsg) *Telefonstichproben in Deutschland*. Westdeutscher Verlag, Opladen, S 101–119
- Kohler M (2005) Stichprobenkonzeption und Stichprobenrealisierung (in diesem Band)
- Niehoff R (1998) Intervieweranforderung und -auswahl, In: Statistisches Bundesamt (Hrsg) *Interviewereinsatz und -qualifikation*, Metzler-Poeschel, Stuttgart, S 51–65
- Prüfer P, Stiegler A (2002) *Die Durchführung standardisierter Interviews: Ein Leitfaden*. ZUMA How-to-Reihe Nr. 11. ZUMA, Mannheim
- Rieck A, Borch S, Kohler M (2005) Methodische Aspekte der Fragebogenentwicklung (in diesem Band)
- Weitkunat R (1998) *Computergestützte Telefoninterviews als Instrument der sozial- und verhaltenspidemiologischen Gesundheitsforschung*, Logos-Verlag, Berlin
- Ziese T (2005) *Der telefonische Gesundheitssurvey 2003 – Baustein zum Gesundheitsmonitoringsystem für Deutschland* (in diesem Band)



## 4 Stichprobenkonzeption und Stichprobenrealisierung

Martin Kohler

### 4.1 Einleitung

Der telefonische Gesundheitssurvey 2003 (GSTelo3) wurde von September 2002 bis März 2003 vom Robert Koch-Institut im Auftrag des Bundesministeriums für Gesundheit und Soziale Sicherung durchgeführt. Die Vorgehensweise bei telefonischen Interviews unterscheidet sich in vielen Aspekten von schriftlichen oder face-to-face Befragungen, die bisher primär im Gesundheitsbereich eingesetzt wurden. Daher wird im Folgenden ausführlicher auf die technischen Rahmenbedingungen, auf das Stichprobendesign, auf den Feldverlauf bzw. die Ablaufsteuerung sowie auf die Datenaufbereitung und daraus abgeleitet auf die Gewichtung eingegangen. Abschließend werden die Ergebnisse einer Nonresponderanalyse des GSTelo3 vorgestellt und die Auswirkungen der Intensität der Sampleausschöpfung untersucht.

### 4.2 Technische Rahmenbedingungen

Die Datenerhebung fand unter Einsatz von gesondert geschulten Interviewern im eigens am Robert Koch-Institut eingerichteten CATI-Labor direkt an den 19 CATI-Arbeitsplätzen statt<sup>1</sup>. Grundlage hierzu bildete ein programmierter Fragebogen, der vergleichbar mit computer-gestützten Interviews (CAPI)<sup>2</sup> sowohl eine Datenkontrolle während der Befragung als auch die Ablaufsteuerung mittels programmierter Filter übernimmt<sup>3</sup>.

Die technische Umsetzung der telefonischen Befragung basiert auf der CATI-Software Interviewer Suite 4.3 der Firma Voxco (Voxco 2001). Die CATI-Software ist ein komplexes System mit mehreren Modulen, basierend auf einem Datenbanksystem (Voxco 2001). Alle Daten werden auf einem zentralen Fileserver gehalten, zu dem besondere Zugangsrechte erforderlich sind. Zusätzlich werden Hard- und Software für den Telefonie-Server benötigt.

Das Interviewer-Suite Programm übernimmt auch die Anrufverwaltung und die Steuerung der »call-backs«, d. h. der Anrufe, bei denen noch kein abschließender Kontakt mit einer Zielperson im Haushalt stattfinden konnte. Die Anwahl wird dabei von einem »auto-dialing«-System übernommen. d. h. der Anwahlvorgang der eingespielten Rufnummern wird automatisch durch das Programm durchgeführt. Fehlanrufe bei Wahlwiederholung sind nahezu ausgeschlossen und eine eindeutige, nachvollziehbare Stichprobenrealisierung ist gewährleistet.

1 ausführlicher hierzu Borch, Rieck (2005) in diesem Band, CATI = computer-assisted telephone interviewing

2 CAPI = computer-assisted personal interviewing; zu den Abkürzungen vgl. auch Couper et al. (1998)

3 zur Fragebogenentwicklung siehe Rieck et al. 2005, in diesem Band

### 4.3 Stichprobendesign

Das Stichprobendesign des GSTelo3 bildet die volljährige, deutschsprachige Wohnbevölkerung in Privathaushalten der Bundesrepublik ab, sofern sie über Festnetzanschlüsse erreichbar ist. Stichprobenbasis bildet ein nach dem Gabler-Häder-Design (Gabler, Häder 1999) erstelltes Telefonnummernsample, das auch die Basis für das ADM Mastersample liefert und damit die Grundlage fast aller bundesweit repräsentativen Telefonbefragungen in Deutschland bildet (ADM/BVM 1998; Heckel 2001).

Der Ausgangspunkt dieses Verfahrens besteht darin, aus dem Pool öffentlich zugänglicher Nummernverzeichnisse eine Bestandsliste sämtlicher in der BRD vorhandener Festnetzanschlüsse zu erstellen. Jedoch ist zu berücksichtigen, dass es seit der Liberalisierung des Telekommunikationsmarktes zu Beginn der 1990er Jahre den Teilnehmern freigestellt ist, sich in öffentlichen Telefonverzeichnissen eintragen zu lassen. Seither wächst die Anzahl der nicht mehr in öffentlich zugänglichen Verzeichnissen registrierten Teilnehmer beständig (Heckel 2001). Um diese »Dunkelziffer« der Nicht-Eingetragenen (so genannte Non-Pubs) mit gleicher Wahrscheinlichkeit erreichen zu können, haben Gabler und Häder den in den USA bei telefonischen Befragungen verwendeten Waksberg-Ansatz (Waksberg 1978) auf die bundesrepublikanischen Verhältnisse zu übertragen versucht. Die Behebung der »Dunkelziffer« der Non-Pubs erfolgt im Rahmen des Gabler-Häder-Designs dadurch, dass aus dem Gesamtbestand der öffentlich zugänglichen Rufnummernverzeichnisse zunächst alle »Service-Nummern« (Vorwahl 0180 u.ä.), Telefax-Nummern, Doppeleinträge sowie Sammelnummern, die nur aus einer rein gewerblichen Bindestrich-Nummer bestehen, automatisiert gelöscht werden. In dem bereinigten Nummernpool werden dann die letzten beiden Ziffern abgeschnitten und so genannte Nummernblöcke gebildet. Diese werden abschließend wieder durch die Ziffernfolgen »00« bis »99« ergänzt und bilden so den Auswahlrahmen für bevölkerungsrepräsentative Telefonstichproben in der Bundesrepublik Deutschland. (Gabler, Häder 1999)

Aus dieser hypothetischen Grundgesamtheit<sup>4</sup> wird dann uneingeschränkt zufällig die Anzahl der erforderlichen Telefonnummern gezogen, die sich

aus der Sollzahl der vorgegebenen Interviews multipliziert mit einem Unsicherheitsfaktor ergibt. Der Unsicherheitsfaktor ist notwendig, weil sich die Existenz eines Anschlusses hinter den generierten Nummer erst im Laufe der Feldarbeit ergibt<sup>5</sup>.

Das Gabler-Häder-Design entspricht damit im Kern einem Random-Route Ansatz, und gewährleistet auch das Erreichen von Haushalten, die sich nicht in öffentlich zugänglichen Telefonverzeichnissen eintragen lassen. Dies ermöglicht eine bundesweit repräsentative Zufallsauswahl aller Privathaushalte, die über einen Festnetzanschluss verfügen. Die Repräsentativität auf Personenebene wird durch eine zweite Auswahlstufe erreicht. Dabei wird in Mehrpersonenhaushalten nur diejenige volljährige Person als Zielperson ermittelt und befragt, die beim ersten Kontakt des entsprechenden Haushalts als nächste Geburtstag hat (next-birthday-Methode)<sup>6</sup>.

Ein erheblicher Vorteil des beschriebenen Stichprobendesigns liegt zudem darin, dass mit dem Auswahlrahmen nach dem Gabler-Häder-Design auch Klumpungseffekte nahezu ausgeschlossen sind und bei genügend großer Gesamtzahl auch regionalisierbare Stichproben realisiert werden können. Registerbasierte Auswahlverfahren können dies nicht ohne zusätzliche Einschränkungen leisten.

Die Tabelle 1 zeigt das im Rahmen des GSTelo3 verwendete und von ZUMA-Mannheim<sup>7</sup> zur Verfügung gestellte Sample der zufallsgenerierten Rufnummern, die im Rahmen des GSTelo3 benutzt wurden (n=45.819) sowie die Eintragsdichte<sup>8</sup>.

4 Für das Jahr 2001 besteht der Gesamtrahmen aus 71,7 Millionen Nummern, von denen 30,7 Millionen tatsächlich eingetragen sind. (Heckel 2002)

5 Dieses so genannte ‚Randomized Last Digits‘-Verfahren ist gegenüber einer reinen Zufallsziffernanzahl, d. h. der Generierung der gesamten Telefonnummer (‚Random Digit Dialing‘), für Deutschland aufgrund der heterogenen Rufnummernstruktur effizienter, da zwischen den Blöcken häufig große Lücken existieren. Es wird hierbei unterstellt, dass der Anteil von nicht eingetragenen Rufnummern, die einen eigenen Block bilden, vernachlässigbar gering ist. Heckel (2001); Häder, Gabler (2000)

6 Nach Expertenmeinung ist diese Methode für die Befragten leichter zu handhaben als die last-birthday-Methode.

7 Zentrum für Methoden und Analysen, Mannheim (ZUMA) <http://www.gesis.org>

8 Unter dem Begriff »Eintragsdichte« versteht man den Anteil der Personen/Haushalte, der sich in ein öffentliches Telefonbuch hat eintragen lassen.



Tabelle 1

Das Nummernsample des GSTelo3 – Eintrag im Telefonverzeichnis und Personenkontakte

Personen-Kontakt unter der angewählten Nummer	Eintrag in Telefonbuch		Gesamt
	Nein	Ja	
Status unklar	0%	0,1%	15
Nein	77,3%	24,7%	26.291
Ja	22,7%	75,2%	19.513
<b>Gesamt</b>	<b>28.444</b>	<b>17.375</b>	<b>45.819</b>
	<b>62,1%</b>	<b>37,9%</b>	<b>100%</b>

Tabelle 1 zeigt, dass bei über einem Fünftel der nicht eingetragenen Telefonnummern ein Privathaushalt anzutreffen ist. Der zusätzliche Aufwand, dabei auch einen hohen Anteil ungültiger Nummern anzuwählen, ist damit durchaus berechtigt.

#### 4.4 Anrufmanagement und Stichprobenbearbeitung

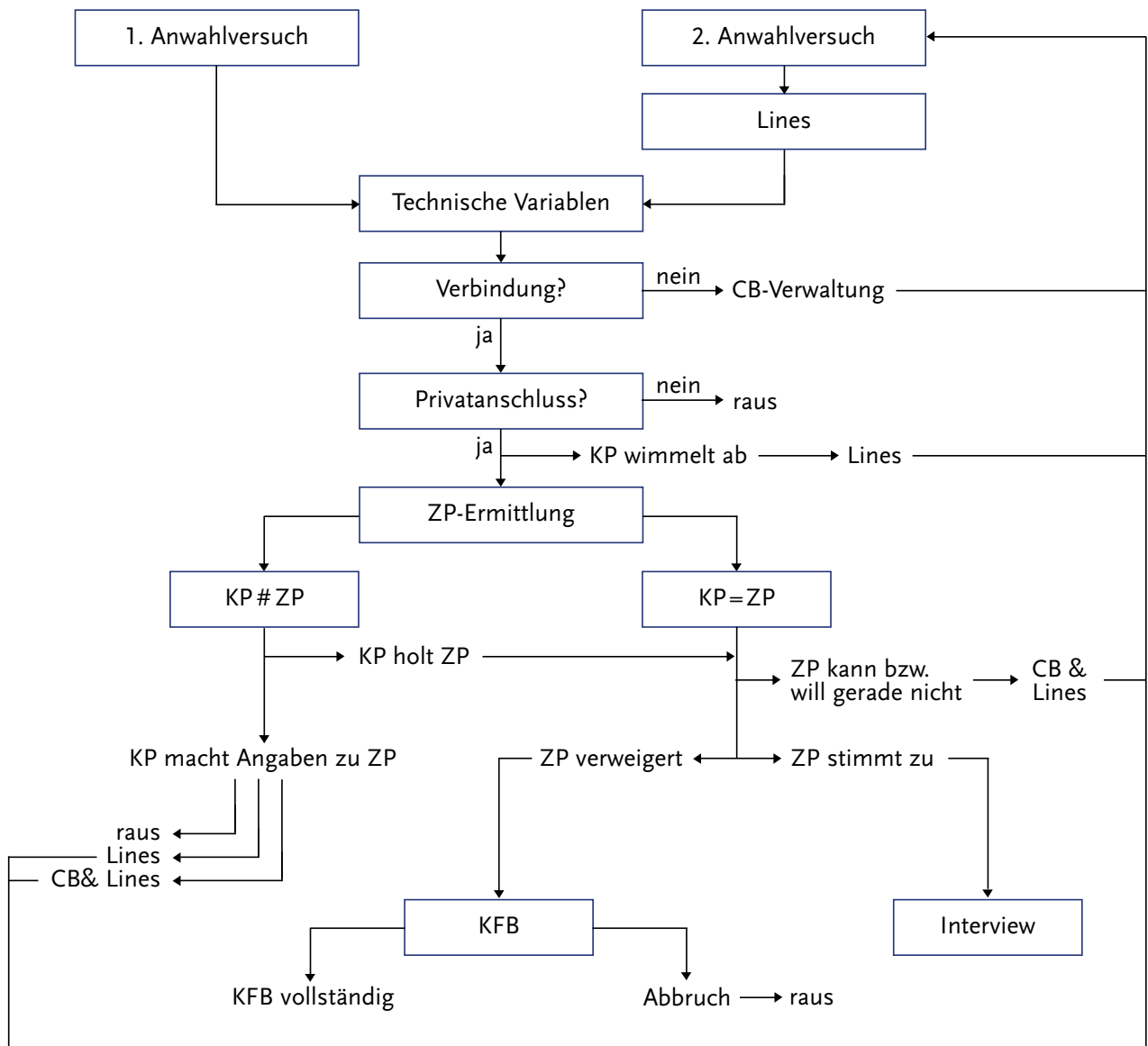
Das für den GSTelo3 gewählte stichprobentheoretische Konzept bestimmt auch maßgeblich das Vorgehen während der Feldarbeit. Im Unterschied zu Ansätzen, bei denen man im Voraus bekannte Personen versucht, telefonisch zu kontaktieren – z. B. basierend auf einer Stichprobe von Einwohnermeldeamtsdaten oder durch vorhergehende Screening-Interviews –, liegt im GSTelo3 als Ergebnis der Stichprobenziehung eine Menge von zu kontaktierender Rufnummern vor, bei denen zunächst nicht bekannt ist, welche tatsächlich existieren und welche davon Anschlüsse von Privathaushalten sind<sup>9</sup>.

Nach dieser Klärung erfolgte dann beim ersten Kontakt mit der zufällig angetroffenen Person (Kontaktperson oder kurz: KP) die Ermittlung der Zielperson im Haushalt (kurz: ZP) nach der next-birthday-Methode. Nur mit dieser Person durfte sofort ein Interview durchgeführt werden oder es war ein geeigneter Rückruftermin zu vereinbaren, an dem die ermittelte Zielperson voraussichtlich zu erreichen ist.

Selbstverständlich gelingt die Klärung, ob es sich um einen Privathaushalt handelt, die Zielpersonermittlung oder gar die Durchführung

9 Zum Begriff des Privathaushalts vgl. auch Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. (ADM), Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V. (ASI) (1999), Statistisches Bundesamt (2004); sowie die Anweisungen im OP-Handbuch zum GSTelo3, Borch et al. (2004). Ein Interview durfte nur dann unter einem Firmenanschluss durchgeführt werden, wenn unter derselben Telefonnummer auch ein Privathaushalt zu erreichen ist (z. B. bei Selbstständigen mit wenigen oder keinen Angestellten).

Abbildung 1  
Anrufsteuerung im GSTelo3



CB – call-back  
 KP – Kontaktperson  
 ZP – Zielperson  
 KFB – Kurzfragebogen  
 lines – Kommentare

des Interviews mit der Zielperson nur selten beim ersten Anwahlversuch einer Telefonnummer. Dies kann häufig erst durch mehrmalige Kontaktversuche festgestellt werden. Hier gilt es einen Mittelweg zu finden zwischen Effizienz der Anwahlversuche und den geforderten Qualitätsansprüchen. Wie an anderer Stelle dargelegt wurde, ist davon auszugehen, dass sich die schwer unter ihrer Festnetznummer erreichbaren Personen systematisch von den leichter zu Erreichenden unterscheiden (Häder 1996).

Die Verwaltung der Rufnummern mit dem jeweils von den Interviewern zu protokollierenden Anwahlergebnis, in dem gegebenenfalls auch Kommentare zur Zielpersonermittlungen enthalten sind, übernimmt dabei das Anruf- und Rückrufmanagement, das den jeweiligen Anforderungen entsprechend konzipiert und programmiert werden muss.

#### 4.4.1 Das Anruf- und Rückrufmanagement

Die Konzeption eines Anruf- und Rückrufmanagements bedarf einiger Vorüberlegungen und kann aus unterschiedlichen Perspektiven betrachtet werden. Aus Sicht der Interviewer sollte es möglichst allen Eventualitäten gerecht werden, sodass eine eindeutige Zuordnung des jeweiligen Anwahlergebnisses vorgenommen werden kann. Gleichzeitig darf es aber auch nicht zu unübersichtlich sein, da die Interviewer in der Kontaktpphase besonderen Belastungen ausgesetzt sind und deren Aufmerksamkeit durch den jeweiligen Gesprächspartner in Anspruch genommen wird. Andererseits erfolgt über die Kodierung des Anwahlergebnisses (disposition codes) auch die Steuerung der Wiedervorlage der entsprechenden Rufnummer nach festzulegenden Regeln (callback rules) und berührt damit die Qualität der Stichprobenausschöpfung<sup>10</sup>. Weiterhin wird über die Vergabe der disposition codes auch die Ergebnisdarstellung der Ausschöpfung und damit auch die Berechnung der Ausschöpfungsquote determiniert. Von zentraler Bedeutung hierbei ist die Unterscheidung, ob sich Anrufergebnisse den Kategorien »neutraler« bzw. »nicht neutraler« Ausfälle zuordnen lassen. Dabei sind wiederum eigene Gesichtspunkte zu beachten, auf die weiter unten noch näher eingegangen wird<sup>11</sup>. Die Entwicklung eines Anrufmanagementsystems stellt damit einen Optimierungsprozess teilweise divergierender Zielvorgaben dar und kann nicht allen Bedürfnissen gerecht werden.

10 So wird beispielsweise differenziert, ob eine Verweigerung durch die – zufällig – erreichte Kontaktperson erfolgt, bevor die Zielperson im Haushalt ermittelt werden konnte. Dieses Ergebnis bekommt einen eigenen Code und wird über die Wiederanrufregeln zeitversetzt erneut angerufen. Konnte dagegen im jeweiligen Haushalt bereits die Zielperson ermittelt werden und diese verweigert, muss eine anderer Code vergeben werden und die Telefonnummer wird nicht mehr angewählt.

11 Wenn beispielsweise bereits mehrere Kontaktversuche bei einer Telefonnummer durchgehend das Ergebnis »Freizeichen« erzielt haben, so ist trotzdem nicht eindeutig zu sagen, ob der Anschluss vergeben ist und wenn ja, ob es sich nicht um einen Firmenanschluss handelt. Nur durch den mehrmaligen Wiederanruf zu verschiedenen Uhrzeiten und Wochentagen kann das Risiko minimiert werden, eine Zielperson nicht zu erreichen.

Anders als in den USA liegen im deutschsprachigen Raum bisher keine einheitlichen Standards für die Entwicklung von disposition codes vor<sup>12</sup>. Die rules und disposition codes mussten daher speziell für den GSTelo3 entwickelt und auf Basis der Erfahrungen des Pretests verfeinert werden. Die Komplexität des Prozesses wird dabei aus dem Ablaufschema für die call-back-Verwaltung in der Abbildung 1 ersichtlich.

Prinzipiell sind für das Anruf- und Rückrufmanagement zwei Arten von disposition codes zu unterscheiden:

- ▶ Erstens Codes, die Telefonnummern zugeordnet werden, die nicht mehr kontaktiert werden sollen (z. B. nach vollständig abgeschlossenem Interview oder wenn die Zielperson aus gesundheitlichen Gründen nicht in der Lage ist teilzunehmen), in der Diktion der AAPOR als final codes bezeichnet.
- ▶ Zweitens Ergebnisse, bei denen mindestens ein weiterer Kontaktversuch stattfinden soll, also temporäre Codes, die die so genannten call-backs definieren. Call-backs können mit dem Gesprächspartner vereinbarte und vom Interviewer als Gesprächsergebnis eingegebene Rückruftermine sein, werden aber auch automatisch durch programmierte Algorithmen (call-back rules) generiert, z. B. wenn bei Besetzt- und Freizeichen kein Kontakt stattgefunden hat oder aus technischen Gründen keine Verbindung hergestellt werden konnte.

Die AAPOR unterscheidet darüber hinaus noch eine dritte Kategorie, die so genannten »action codes«, unter denen alle Ereignisse subsummiert werden, die manuelle Eingriffe erfordern und nicht über das call-back Management automatisiert verwaltet werden, z. B. das Zuweisen von Anrufen an bestimmte Interviewer oder manuelles Entfernen eines Datensatzes aus dem call-back-Pool.

Jedes Anrufergebnis wird während der Feldarbeit durch die Interviewer dokumentiert und mit einem entsprechenden Code versehen. Diesen Ergebniscodes werden dann Regeln zugeordnet, nach denen die jeweilige Rufnummer zur Wiedervorlage kommt. Diese »call-back rules« bestimmen u. a., zu welchem Zeitpunkt (nach zehn Minuten, einem Tag, mehreren Tagen etc.) eine mit call-back-Status versehene Telefonnummer wieder in den anzurufenden Pool von Datensätzen eingespielt wird.

Die Erstellung der rules erfolgt in einem eigenen Modul der CATI-Software, indem jedem disposition code zunächst ein Statusmerkmal (call-back »ja« oder »nein«) zugewiesen wird. Falls der Status auf »ja« gesetzt wurde, folgt die Regel, wie in diesem Fall verfahren werden soll. Bei der Erstellung der rules wird auch die Priorität festgelegt, mit der die call-backs abgearbeitet werden. Dies ist insbesondere relevant, wenn zu bestimmten Zeiten mehrere call-backs vorliegen, sodass zwangsläufig einige nur mit Verzögerung bearbeitet werden können<sup>13</sup>.

Außerdem wird im call-backs-Modul die Priorität festgelegt mit der »frische« Datensätze – d. h. Telefonnummern, die bisher noch nicht angerufen wurden – behandelt werden. Dies ist vor allem bei einem bereits stark ausgeschöpften Sample relevant: sind die Nummern im bestehenden call-back-Pool bereits mehrmals angerufen worden, sinkt die Wahrscheinlichkeit eines Kontaktes rapide. Um die Effizienz der Datenerhebung und auch die

<sup>12</sup> Für die USA gibt es Richtlinien der American Association of Political Opinion Research (AAPOR 2004). Die in etwa vergleichbaren Einrichtungen der ADM-Institute in Deutschland oder ZUMA in Mannheim haben keine derartigen Empfehlungen. Die umfangreiche Literatur über Telefonsurveys geht auf diesen Aspekt in der Regel nicht ein.

<sup>13</sup> So versteht sich von selbst, dass mit Zielpersonen vereinbarte Termine mit höherer Priorität behandelt werden müssen, als beispielsweise call-backs, die wegen Besetzt- oder Freizeichen zum gleichen Zeitpunkt zur Wiederanwahl vom System eingespielt werden.

Motivation der Interviewer zu erhalten, ist es in Maßen notwendig, »frische« Nummern zuzulassen. Deren Anteil ist jedoch möglichst gering zu halten, da sie gegen Ende der Feldphase nicht mehr in gleichem Maße systematisch ausgeschöpft werden können, wie früher eingespielte Telefonnummern. Diesem Aspekt wird daher u. a. auch in den Anweisungen zur Durchführung des BRFSS<sup>14</sup> der CDC unter dem Begriff »wind down procedures« ein eigener Abschnitt gewidmet.

Eine weitere Einstellung im call-back-Modul betrifft die Festlegung der maximalen Zahl der Kontaktversuche einer Telefonnummer. Gabler, Häder (1999) empfehlen mindestens zehn Kontaktversuche, Porst u. a. (1998) zitieren Literatur, wonach bereits nach dem dritten Anruf keine deutliche Ausschöpfungssteigerung mehr erzielt werden könne. Im GSTelo3 wurde die Zahl der maximalen Kontaktversuche aus Qualitätsgründen auf 15 festgelegt. Nur wenn ein Anruf (z. B. Freizeichen) trotz wechselnder Uhrzeiten, wechselnder Wochentage und längeren Abständen zwischen den Kontaktversuchen stets das gleiche Ergebnis erzielte, wurden Telefonnummern lediglich neun mal kontaktiert.

Bei Anrufergebnissen, die gegen einen vergebenen Anschluss sprechen (wie z. B. »Leitung tot« oder andere Display-Anzeigen), wird in deutlichem zeitlichem Abstand ein zweiter Kontaktversuch unternommen, um das Erstergebnis zu überprüfen.

#### 4.4.2 Prozesssteuerung im Call-back Management

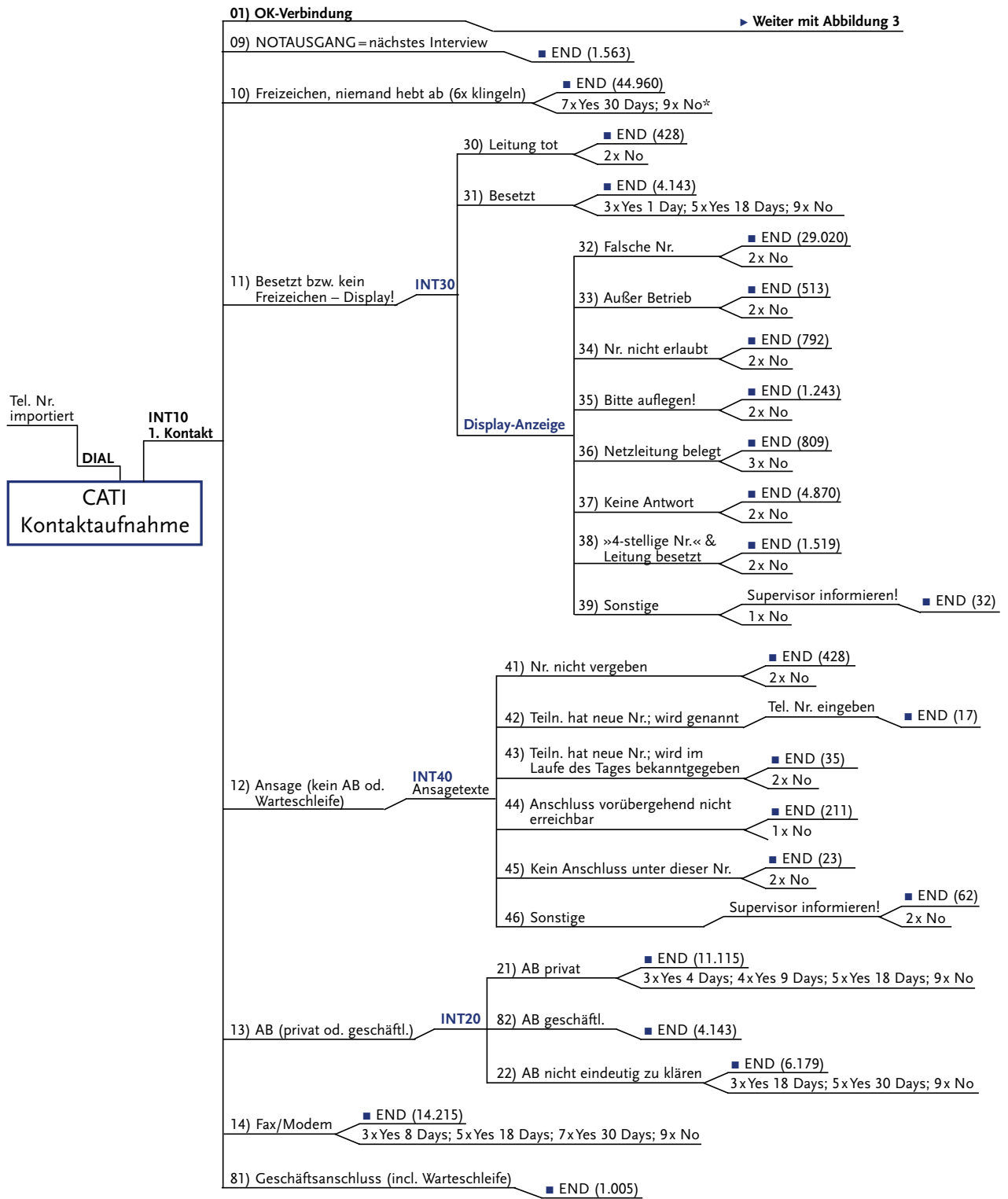
Nachfolgend wird das Anrufmanagement im GSTelo3 ausführlicher behandelt. Dieser Abschnitt richtet sich vorwiegend an technisch interessierte Leser, die selbst ein Anrufmanagement zu betreuen haben oder dies beabsichtigen. Er wird an dieser Stelle behandelt, weil die Dokumentation dieses Aspekts telefonischer Umfragen üblicherweise in Publikationen zu kurz kommt, aber für die Transparenz des Vorgehens im Rahmen von telefonischen Befragungen bedeutsam ist.

Die Abbildungen 2 und 3 zeigen die Komplexität des Anrufmanagements in seiner gesamten Breite. Die Verzweigungen enden jeweils mit dem für einen Anwahlvorgang möglichen Resultat, das die Interviewer zur Dokumentation während der Kontaktabbahnung festhalten müssen – dargestellt durch eine Ziffer mit Klammer, z. B. »10«). Der Text hinter der Klammer – »Freizeichen, niemand hebt ab (6x klingeln)« – beschreibt in Kürze, in welchen Fällen dieser Code zu wählen ist. Nach dem END-Symbol werden in den Baudarstellungen in Klammern die Häufigkeiten wieder gegeben, mit der im Verlauf des GSTelo3 das jeweilige Anwahlergebnis vergeben wurde. Wichtig ist in diesem Zusammenhang auch der untere Zweig der letzten Verzweigung. Hier werden die dem jeweiligen Anwahlergebnis zugeordneten Rückrufregeln ausgewiesen. So zeigt z. B. die Regel für den Code 10 »Freizeichen«, dass nach 7 mal Freizeichen als Anwahlergebnis weitere 2 mal zeitversetzt um 30 Tage angerufen wird. Erst wenn 9 mal in Folge nur das Freizeichen erzielt wurde, wird die jeweilige Telefonnummer aus dem call-back entfernt. Das ist frühestens nach 2 Monaten Feldzeit der Fall.

Die Verzweigungen in den Darstellungen markieren den Wechsel von einer Bildschirmseite zur nächsten. D. h., die Ausdifferenzierung der Anwahlergebnisse zieht sich in der Eingangsphase einer Kontaktaufnahme über mehrere Bildschirmseiten hin, bei der die Interviewer auch die Einleitungstexte beachten müssen. Bereits dadurch wird deutlich, weshalb die adäquate Vermittlung des Umgangs mit dem call-back Management in den Schulungen einen großen Teil der Zeit beansprucht, vgl. auch Borch, Rieck (2005).

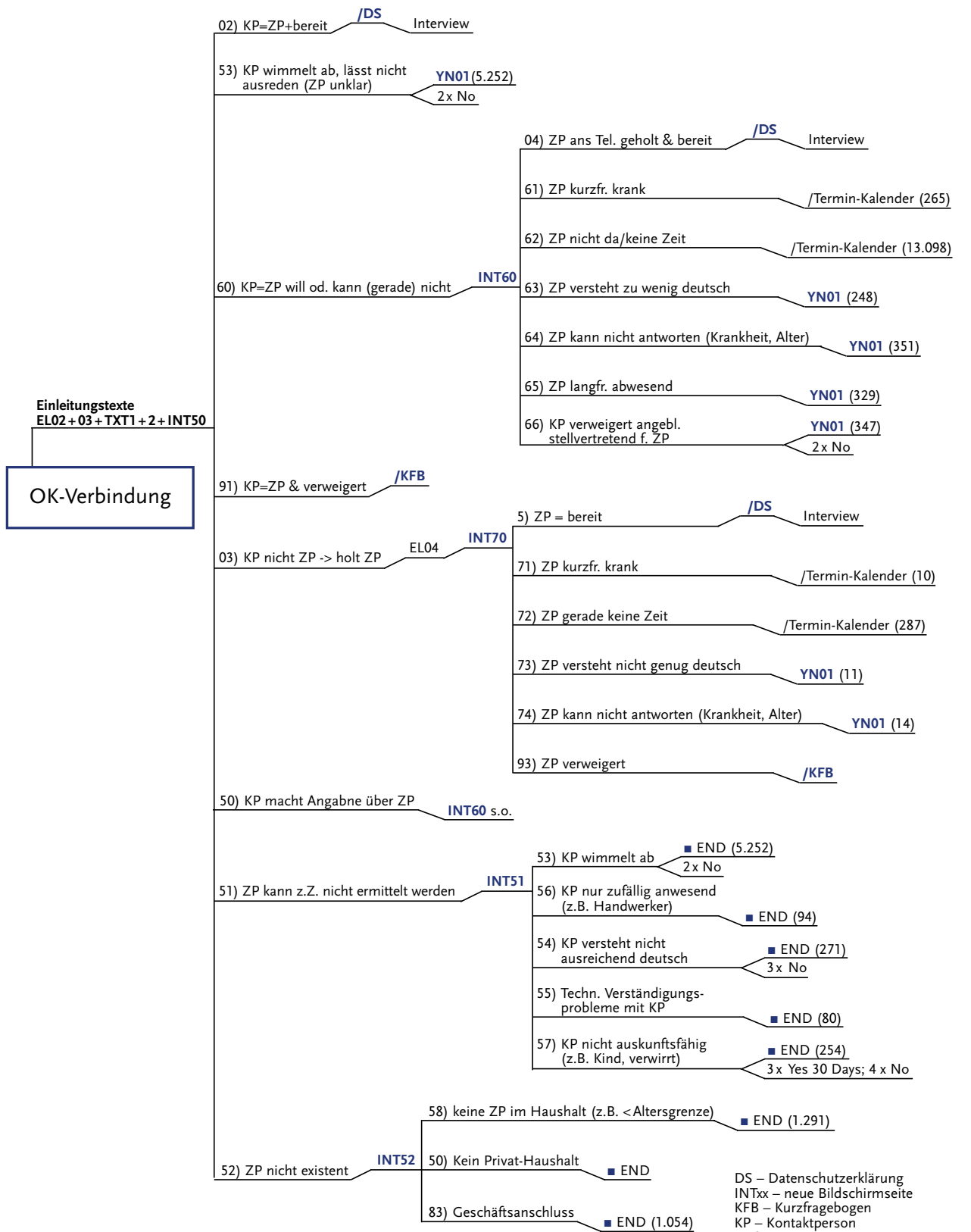
<sup>14</sup> Behavioural Risk Factor Surveillance System – jährliche bundesweite Telefonbefragung zum Gesundheitszustand der Bevölkerung in den Vereinigten Staaten im Auftrag der CDC (1998)

Abbildung 2  
Baumdarstellung Anrufsteuerung bei der Kontaktaufnahme



DS – Datenschutzerklärung  
 INTxx – neue Bildschirmseite  
 KFB – Kurzfragebogen  
 KP – Kontaktperson  
 YN01 – Kommentar ja/nein  
 ZP – Zielperson  
 \* – Rückrufregeln, näheres siehe Text

Abbildung 3  
Baumdarstellung Anrufsteuerung bei der Kontakthanbahnung



DS – Datenschutzerklärung  
 INTxx – neue Bildschirmseite  
 KFB – Kurzfragebogen  
 KP – Kontaktperson  
 YN01 – Kommentar ja/nein  
 ZP – Zielperson  
 \* – Rückrufregeln, näheres siehe Text

Die Abfolge der Bildschirmseiten wird nachfolgend exemplarisch für das Anwahlergebnis 62 »Zielperson hat gerade keine Zeit« und anschließender Terminvereinbarung dargestellt.

Die Abbildung 4 zeigt den Bildschirm nach dem Anwahlvorgang. Es kommt eine Verbindung zustande.

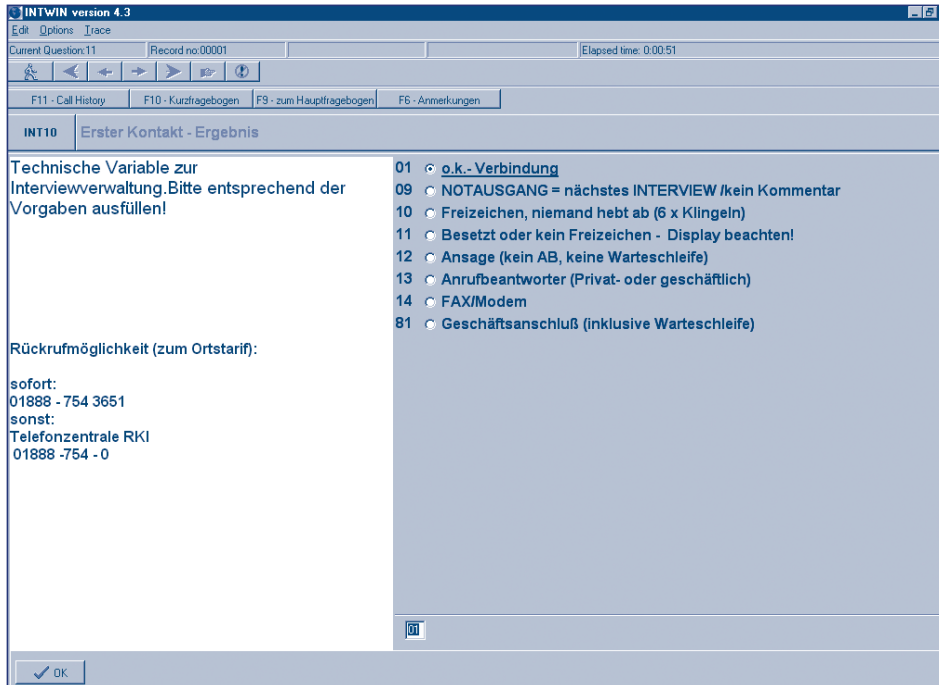


Abbildung 4  
Bildschirm Anwahlvorgang

Nach dem Vorlesen der Einleitungstexte – hier nicht wiedergegeben – stellt sich bei der Ermittlung der Zielperson heraus, dass der gegenwärtige

Gesprächspartner gleichzeitig auch die Zielperson ist, aber gerade keine Zeit für ein Interview hat. Es wird der Code »60« gewählt (Abbildung 5).

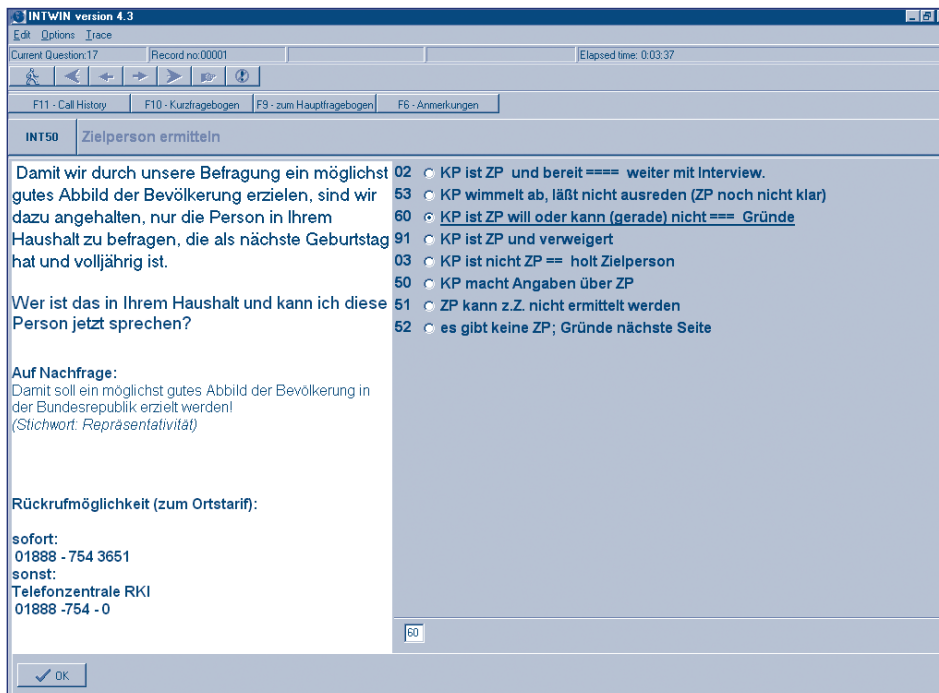


Abbildung 5  
Bildschirm Zielpersonen-  
ermittlung



Darauf hin folgt die Ausdifferenzierung der Gründe der aktuellen Nichtteilnahme(-bereitschaft) auf einer weiteren Bildschirmseite Hier

werden die Gründe der Nichtteilnahme der Zielperson ausdifferenziert (Abbildung 6).

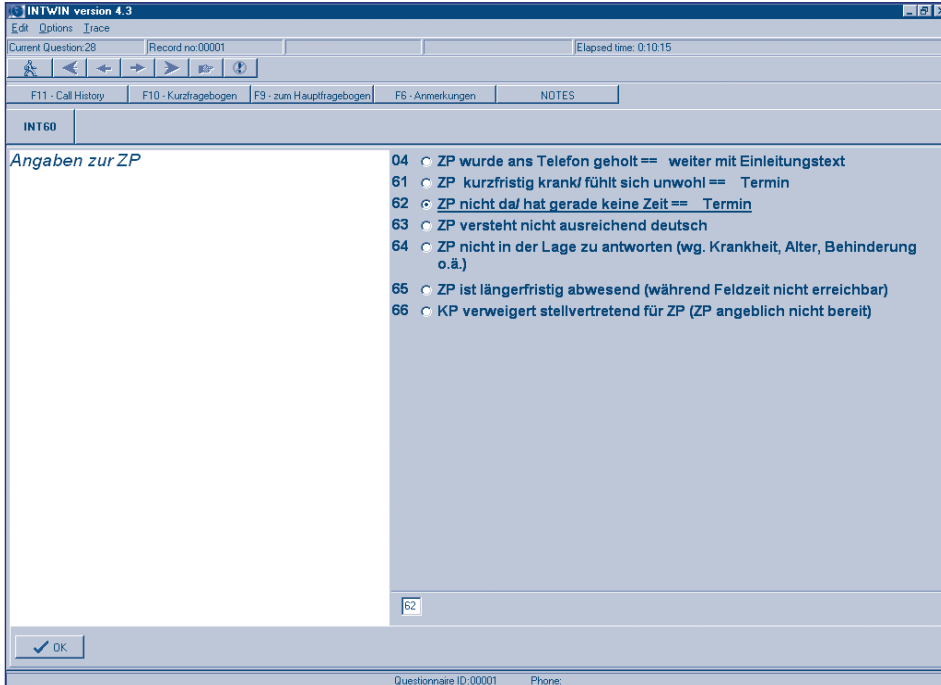


Abbildung 6  
Bildschirm Gründe der Nichtteilnahme

Da die Zielperson gerade keine Zeit hat, aber zu einem späteren Zeitpunkt bereit wäre, ein Inter-

view zu führen, wird ein Termin zu einem späteren Zeitpunkt mit ihr vereinbart (Abbildung 7).

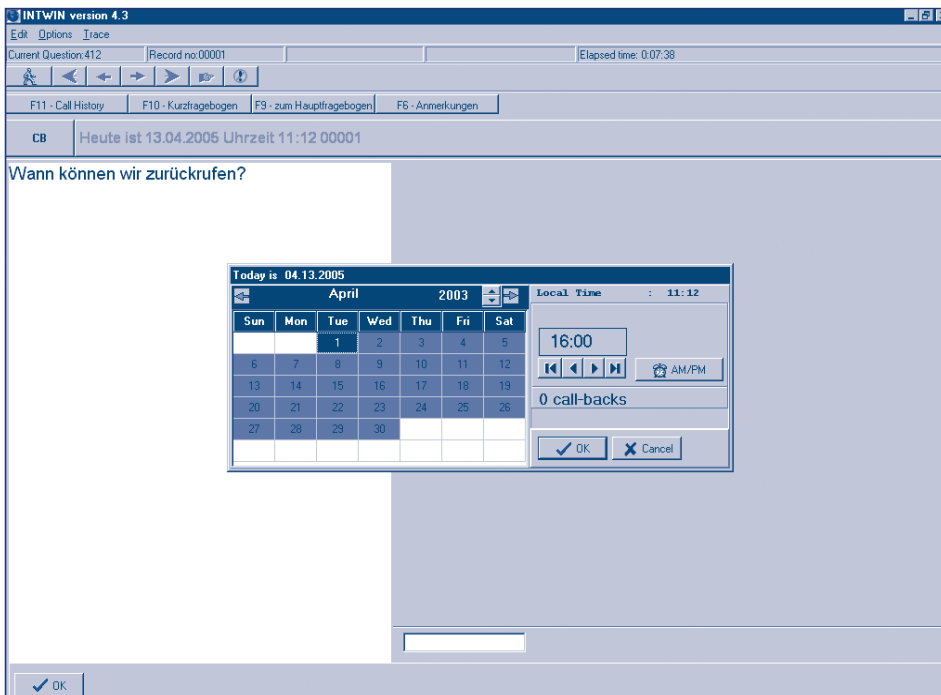


Abbildung 7  
Bildschirm Terminvereinbarung

**Tabelle 2**  
**Häufigkeit der Anwahlergebnisse im Prozessverlauf**

Anwahlergebnis	Häufigkeit	in %
10) Freizeichen, niemand hebt ab (6 x Klingeln)	44.960	27,6
32) Falsche Nummer	29.020	17,8
14) FAX/Modem	14.215	8,7
62) ZP nicht da/hat gerade keine Zeit	13.098	8,0
21) AB privat	11.115	6,8
97) Hauptfragebogen komplett	8.351	5,1
22) AB nicht eindeutig zu klären	6.179	3,8
99) endgültiger Abbruch	5.528	3,4
53) KP wimmelt ab, lässt nicht ausreden (ZP noch nicht bekannt)	5.252	3,2
37) keine Antwort	4.870	3,0
31) Leitung besetzt (später nochmals versuchen)	4.143	2,5
09) NOTAUSGANG = nächstes INTERVIEW/kein Kommentar	1.563	1,0
38) keine Antwort	1.519	0,9
58) keine ZP im Haushalt (z. B. unter Altersgrenze)	1.291	0,8
35) Bitte auflegen!	1.243	0,8
98) Kurzfragebogen beendet	1.108	0,7
83) Geschäftsanschluss	1.054	0,6
81) Geschäftsanschluss (inklusive Warteschleife)	1.005	0,6
08 INTERVIEW beenden ==> Wiederaufnahme möglich	1.001	0,6
36) Netzleitung belegt	809	0,5
34) Nummer nicht erlaubt	792	0,5
33) Außer Betrieb	513	0,3
82) AB geschäftlich	455	0,3
90) INTERVIEW beenden	441	0,3
30) Leitung tot – nichts tut sich	428	0,3
64) ZP nicht in der Lage zu antworten (wg. Krankheit o.ä.)	351	0,2
66) KP verweigert stellvertretend für ZP	347	0,2
65) ZP während Feldzeit nicht erreichbar	329	0,2
72) Zielperson hat gerade keine Zeit -> Termin	287	0,2
54) KP versteht nicht ausreichend deutsch	271	0,2
61) ZP kurzfristig krank/ fühlt sich unwohl	265	0,2
57) KP ist nicht auskunftsfähig (z. B. Kind, verwirrte)	254	0,2
63) ZP versteht nicht ausreichend deutsch	248	0,2
44) Anschluss ist vorübergehend nicht erreichbar	211	0,1
41) Nummer nicht vergeben	140	0,1
59) kein Privathaushalt (Anstalt o.ä.)	120	0,1
56) KP ist nur zufällig anwesend (z. B. Handwerker o.ä.)	94	0,1
55) Technische Verständigungsprobleme mit KP	80	0

Fortsetzung S. 43

Anwahlergebnis (Fortsetzung)	Häufigkeit	in %
49) sonstige Ansage – bitte Supervisor informieren!	62	0
43) TN neue Nummer; Nummer wird bekannt gegeben	35	0
39) sonstige Displayanzeige -> Supervisor informieren!	32	0
45) Kein Anschluss unter diese Nummer	23	0
42) TN neue Nummer; Nummer wird genannt	17	0
74) ZP nicht in der Lage (Alter, Behinderung, Krank)	14	0
73) ZP versteht nicht ausreichend deutsch	11	0
71) ZP kurzfristig krank/fühlt sich unwohl -> Termin	10	0
00) offene Angabe	2	0
<b>Gesamt</b>	<b>163.156</b>	<b>100</b>

Nach Eingabe des Tages und der Uhrzeit des vereinbarten Termins wird der Fragebogen beendet. Die Telefonnummer wird nun im call-back-Managementsystem verwaltet und zum vereinbarten Zeitpunkt einem gerade »freien« Interviewer erneut zugewiesen. Diese Sequenz lässt sich auch in den Abbildungen 2 und 3 über den Ablauf »Kontaktaufnahme«, »OK-Verbindung«, »Einleitungstext« über INT60 bis zum Endcode 62 nachvollziehen. In Abbildung 3 ist auch zu erkennen, dass dieser Ablauf während des Feldverlaufs im GSTelo3 insgesamt 13.098 mal aufgetreten ist.

Abschließend werden die Ergebnisse aller Anwahlvorgänge tabellarisch nach Auftretenshäufigkeit sortiert aufgelistet, die während der gesamten Feldzeit im GSTelo3 durchgeführt wurden (Tabelle 2). Insgesamt wurden während der 112 Tage Feldzeit über 160.000 Anwahlvorgänge durchgeführt.

Bereits die hohe Zahl von Anwahlvorgängen verweist auf zusätzlich zu berücksichtigende infrastrukturelle Erfordernisse bei der Beurteilung und Bewertung der Ergebnisse von telefonischen Surveys. Ohne automatisches Anwahlverfahren kann eine derart hohe Zahl von Anwahlvorgängen in der zur Verfügung stehenden Zeit kaum bewältigt werden, abgesehen von Eingabefehlern, die etwa manuelle Wählverfahren mit sich brächten. Auf die Stunde umgerechnet ergeben sich bei 19 Plätzen und einer vierstündigen Belegung der Arbeitsplätze immer noch durchschnittlich über 19 Wahlvorgänge pro Stunde, die ein Interviewer erbracht hat. Wird zusätzlich noch die Dauer der Interviews abgezogen – im Durchschnitt ca. vier Interviews pro Tag und Interviewer mit durch-

schnittlich 25 Minuten Länge – erhöht sich die Zahl der Anwahlvorgänge auf etwa 40 pro Stunde und Person.

Die Dokumentation der Anwahlvorgänge in der call history erfüllt aber noch andere Zwecke. So ergeben sich aus der Häufigkeit der Anwahlergebnisse im Feldverlauf bereits erste Aufschlüsse über die Qualität der Stichprobenbasis. Darüber hinaus lassen sich auch Fehlsteuerungen im call-back Management identifizieren. Eine bedeutende Rolle spielt die Verlaufsdatei darüber hinaus im Kontext der feldbegleitenden Qualitätssicherung. So lassen sich aus der ungewöhnlichen Häufung bestimmter Anwahlergebnisse Defizite in der Schulung der Interviewer erkennen oder aber auch eventuelle Täuschungsmanöver identifizieren. Die Qualitätsstandards im BRFSS sehen dafür auch spezielle Quotienten der Anwahlergebnisse pro Interviewer vor, die während der Feldzeit laufend kontrolliert werden sollten.

#### 4.5 Datenhaltung

Die Datenerhebung im Rahmen des GSTelo3 erfolgte – wie erwähnt – mittels der CATI-Software Interviewer Suite 4.3 der Firma Voxco, das als Datenbankprogramm sowohl der Datenerfassung während der Befragung als auch der Anrufverwaltung dient. Das Softwareprogramm arbeitet mit so genannten Projekten, die aus mehreren Teildateien bestehen. Die beiden wichtigsten Dateien führen die Anwahlergebnisse (call-backs) und die Datensätze der abgeschlossenen Interviews. Den Kern eines Projektes bildet die Fragebogendatei, in der neben den eigentlichen Fragen zusätzlich in einem Vor- und Abspann technische Variablen definiert werden müssen, die der Anrufverwaltung und -erledigung dienen. Nach Abschluss der Feldphase liegen drei Arten von Datensätzen vor: eine Datei mit den Ergebnissen der ca. 160.000 Anwahlvorgänge (call history), eine Zweite mit dem letzten Ergebnis des jeweiligen Anrufs (call result). Die dritte Datenart besteht aus den vollständig abgeschlossenen Interviews, die Teilmenge der call results sind. Diese Datei bildet – nach Bereinigung um unvollständige Fälle – die Datenbasis für die weitere Auswertung.

#### 4.6 Sampleausschöpfung und Response

Gegenüber registerbasierten Auswahlverfahren ermöglichen zufallsgenerierte Nummernsample keine Bestimmung einer Rücklaufquote im herkömmlichen Sinn, da sich die Stichprobenbasis erst im Vollzug der Feldarbeit realisiert. In der Literatur zu Telefonsurveys wird daher auch von Ausschöpfung anstelle von Rücklauf gesprochen.

Für die Darstellung der Ausschöpfungsquoten von telefonischen Befragungen sind grundsätzlich folgende Punkte von Bedeutung:

- ▶ Das gewählte Stichprobenkonzept (Random vs. Register)
- ▶ Die Festlegung der maximalen Anwahlversuche
- ▶ Die Differenziertheit der disposition codes
- ▶ Damit in Verbindung stehend die Feldsteuerung, d. h. die Regeln, nach denen eine Telefonnummer zur Wiedervorlage zugespielt wird (call-back rules)
- ▶ Das Verfahren, nach dem bei Annäherung an die Sollzahl das Verhältnis »neue« zu »alte« Nummern gesteuert wird (wind-down-procedure)
- ▶ Damit einhergehend die Aufspaltung des Gesamtsamples in Teilsample oder Quoten
- ▶ Die fernmeldetechnische Ausstattung der Telekommunikations-Anlage (insbesondere: Analog- vs. ISDN-Anschluss)
- ▶ Damit verbunden, das verwendetet Wählverfahren (manuell, auto-dialing, predictive-dialing)
- ▶ Funktionalität und Stabilität der Soft- und Hardware sowie die Netzstabilität und
- ▶ nicht zuletzt: die Qualität der Schulung von Interviewerinnen und Supervisoren (Feldteam-Performance)<sup>15</sup>

Wegen der Vielzahl zu berücksichtigender Faktoren hat sich bisher noch kein allgemeiner Standard für die Darstellung der Ausschöpfung bei Telefonstichproben durchgesetzt, zumal sich die angeführten Faktoren auch wechselseitig beeinflussen. Die zentralen Aspekte für die Ergebnisaufbereitung und -darstellung liegen jedoch beim Stichprobenkonzept sowie in der Differenziertheit der disposition codes, sofern mit einem call-back-Managementsystem gearbeitet wird.

15 Vgl. auch Borch, Rieck (2005) in diesem Band

Für die vorliegende Aufbereitung der Anrufergebnisse des GSTelo3 wurden folgende Bezüge zu im deutschsprachigen Raum veröffentlichten Darstellungen hergestellt. Thematisch am nächsten stehen die Untersuchungen des IBE München<sup>16</sup>. Allerdings sind diese Erhebungen registerbasiert und kommen daher mit einer geringeren Anzahl von disposition codes aus. Darüber hinaus unterscheiden sich die Erhebungen noch in der Durchführung. Während im Falle des GSTelo3 auto-dialing, also die programmgesteuerte Anwahl der Rufnummern zum Einsatz kam, wird in den IBE-Erhebungen mit manuellem Anwahlverfahren gearbeitet. Ob die fernmeldetechnische Ausstattung bei den IBE-Erhebungen ISDN oder analog ausgelegt ist, kann den Veröffentlichungen nicht entnommen werden. ISDN-Anlagen gestatten differenziertere disposition codes, weil über Display-Anzeigen eine zusätzliche Kategorisierung der technischen Ausfälle möglich ist. Bei Analoganlagen ertönt stattdessen entweder ein »Besetzt«-Zeichen oder es kommt einfach keine Verbindung zustande. Alle Faktoren zusammen genommen, bestehen deutliche verfahrensbezogene Unterschiede, sodass die Ergebnisdarstellung nur schwer in das IBE-Schema übertragbar wäre.

Dem Vorgehen im GSTelo3 am nächsten kommt das ADM-Verfahren<sup>17</sup>, welches auf einem modifizierten Gabler-Häder Stichprobendesign beruht (Heckel 2002; Heyde 2002). Die wesentlichen Unterschiede bestehen in der Differenziertheit der disposition codes sowie in der Anzahl der maximalen Kontaktversuche pro zufallsgenerierter Telefonnummer<sup>18</sup>.

Die folgende Darstellung der Ausschöpfungsquote orientiert sich primär an der ADM-Schematik, weil deren Erhebungen sowohl hinsichtlich des Stichprobendesigns als auch in den Durchführungsbedingungen dem Vorgehen beim GSTelo3 am nächsten kommen. Aus der IBE-Schematik wird ergänzend die Unterscheidung zwischen Haushalts- und Zielpersonenkontakten übernommen. Der Haupt Gesichtspunkt dabei ist, dass erst durch Personenkontakt sicher gestellt werden kann, ob die zufallsgenerierten Nummern Privathaushalte, also Personen unserer Grundgesamtheit enthalten. Erst dadurch wird eine sichere Bestimmung der stichprobenneutralen Ausfälle möglich.

#### 4.6.1 Leitende Kriterien für die Klassifikation der disposition codes

Bei der Aufbereitung und Verdichtung der Anwahlergebnisse der call history ist zunächst die Unterscheidung zwischen verlaufs- und ergebnisbezogener Betrachtung wichtig, um eine möglichst genaue Zuordnung der generierten Anwahlnummern nach dem Kriterium »Element der Grundgesamtheit«, also der Bestimmung von stichprobenneutralen bzw. nicht neutralen Ausfällen vornehmen zu können. Erst dies ermöglicht die Berechnung einer Responsequote, weil es sich beim Auswahlrahmen des Gabler-Häder Verfahrens um zufallsgenerierte Nummern handelt.

Unter verlaufsbezogener Betrachtung stellt sich zudem die Frage, welche Einordnung Ergebnissen zukommt, deren Status nicht endgültig geklärt werden konnte, bzw. wegen der Regeln und maximaler Anzahl der Anrufe während der begrenzten Feldzeit offen bleiben.

Daraus folgt, dass sowohl die call-back rules als auch die maximale Zahl der Anwahlversuche sowie die Verweildauer eines Datensatzes während der Feldarbeit im Nummern-Pool, Auswirkungen auf die Berechnung der Responsequote haben. Um diesen unterschiedlichen Perspektiven in der Auswertung der call history gerecht werden zu können, wurden die in den disposition codes festgehaltenen Anwahlergebnisse nach folgenden Gesichtspunkten verdichtet. Dabei wurde auf die in den USA durch die AAPOR und CASRO<sup>19</sup> entwickelten Standards soweit möglich Bezug genommen (AAPOR 2004; CASRO 1982).

16 Institut für Biometrie und Epidemiologie München u. a.: (Weitkunat 1998; Fischer et al. 2001; Wiedenmayer 2002)

17 ADM: Arbeitsgemeinschaft deutscher Markt- und Meinungsforschungsinstitute

18 beim ADM-Ansatz: Anzahl der disposition codes in der komprimierten Darstellung = 24, maximale Anzahl Kontakte = 5, gegenüber 46 disposition codes und bis zu 15 Kontaktversuchen beim GSTelo3.

19 Council of American Survey Research Organizations

Ein wesentlicher Unterschied zu den deutschsprachigen Veröffentlichungen beruht dabei auf der Differenzierung zwischen »temporary«, »final«- und »action«-Codes. Dies ist deshalb wichtig, weil der Endzustand der call history nur unzureichende Informationen über den Status der jeweiligen zufallsgenerierten Rufnummer erlaubt. Die ausschließliche Betrachtung des letzten Anwahlergebnisses und dessen Zuordnung zu einer der vier Hauptgruppen

- ▶ Interview – teilweise oder vollständig (eligible, with response)
- ▶ Element der Grundgesamtheit, aber kein Kontakt (eligible, no response)
- ▶ Status der Nummer konnte nicht geklärt werden (unknown eligibility)
- ▶ Nummer nicht Element der Grundgesamtheit (not eligible)

kann je nach der Zuordnung der entsprechenden Codes zu einer der vier Hauptgruppen die Responsquote erheblich beeinflussen. Zeigt beispielsweise eine Anrufsequenz beim ersten Anruf einen Kontakt mit einer Zielperson, die aber zur Zeit des Anrufs keine Zeit hatte, können die Folgeanrufe jeweils mit dem Resultat »Anrufbeantworter« oder »Fax« enden. Diese Codes werden dann aber möglicherweise in Verbindung mit den call-back rules den stichprobenneutralen Ausfällen zugeordnet. Die Empfehlungen der AAPOR sehen daher für die Berechnung der Ausschöpfungsquote vor, den gesamten Feldverlauf in die Analysen einzubeziehen, indem die Endresultate durch das letzte Anrufergebnis mit einer Zielperson im Haushalt ersetzt werden. Konnte kein Zielpersonenkontakt hergestellt werden, soll stattdessen der letzte Kontakt mit einer beliebigen Person des Haushalts als Endergebnis verwendet werden<sup>20</sup>. Das setzt voraus, dass die call Results sequenziell-hierarchisch in einen Endzustand transformiert werden.

Als Ergebnis liegen bereinigte call results vor, die eine exaktere Bestimmung der Ausschöpfungs- und Responsquoten ermöglichen<sup>21</sup>. Der Plural wird hier bewusst gewählt, weil sich in der Literatur kein einheitlicher Standard für die »richtige« Berechnung von Response- und Ausschöpfungsquoten finden lässt und eine Vielzahl von Berechnungsvarianten vorliegen (AAPOR 2004; CASRO 1982;

Latza et al. 2004). Im Folgenden wird daher zwischen Kooperationsrate und Ausschöpfungsquote unterschieden. Die Kooperationsrate ist definiert als Verhältnis der Anzahl derjenigen Befragten, über die mindestens vollständige Angaben des Kurzfragebogens oder aus der Hauptbefragung vorliegen, bezogen auf die Gesamtzahl der ermittelten Zielpersonen. Die Ausschöpfung bezieht dagegen nur die Zahl der auswertbaren vollständigen Interviews auf die Zielpersonenkontakte.

Werden die 9.426 durchgeführten Interviews auf die 15.918 Zielpersonenkontakte bezogen, ergibt sich mit etwas über 59 % eine bei vergleichbaren Telefonbefragungen im deutschsprachigen Raum übliche Kooperationsrate. Die Ausschöpfungsquote liegt mit knapp 53 % ebenfalls im Bereich gängiger Erhebungen. Der Gesundheitsmonitor München berichtet eine Ausschöpfungsquote von 52 % (Wiedenmayer 2002), das Münchener Lebensstilpanel kommt – bei einem aufwändigeren Rekrutierungsverfahren mit vorherigem Anschreiben der Probanden – auf eine Quote von 67,2 % (Weitkunat 1998) und der Iögd NRW berichtet über Ausschöpfungsquoten zwischen 50 bis 60 % (Murza et al. 2003).

20 Human-contact-over-other-outcome rule (AAPOR 2004, S. 9)

21 McCarty (2003) berichtet über Differenzen von durchschnittlich 2 % je nach angewendetem Transformationsalgorithmus.

Tabelle 3  
Ausschöpfung des Nummernsamples im GSTelo3 – Endzustand der Call-Results

Ausschöpfung	Absolut	Brutto- stichprobe	Netto- stichprobe	Personen- kontakt	Zielpersonen- kontakt	Vollständige Interviews
<b>Bruttoansatz</b>	<b>45.819</b>	<b>100,0%</b>				
Telekom-Ansage/Falsche Nummer	18.604	40,6%				
Max-Kontakt erreicht	4.863	10,6%				
FAX/Modem	715	1,6%				
Kein Privathaushalt	2.601	5,7%				
Doppeladresse	–	0,0%				
Falsche Regionalzelle/Quote voll	9	0,0%				
Sprachprobleme	252	0,5%				
Sonstige neutrale Ausfälle	1.268	2,8%				
<b>Neutrale Ausfälle insg. (Summe 1–8)</b>	<b>28.312</b>	<b>61,8%</b>				
<b>Bereinigtes Brutto</b>	<b>17.507</b>	<b>38,2%</b>	<b>100,0%</b>			
Teilnehmer nimmt nicht ab, Status noch offen	242	0,5%	1,4%			
Anschluss besetzt	15	0,0%	0,1%			
Anrufbeantworter	293	0,6%	1,7%			
<b>Personenkontakte insg. (Brutto Privat-Haushalte)</b>	<b>16.957</b>	<b>37,0%</b>	<b>96,9%</b>	<b>100,0%</b>		
Terminvereinbarung nicht wahrnehmbar	73	0,2%	0,4%	0,4%		
Kein Termin möglich während Feldzeit	313	0,7%	1,8%	1,8%		
KP ist beschäftigt, nicht erreichbar	–	0,0%	0,0%	0,0%		
KP gibt keine Auskunft: Thema/Projekt	–	0,0%	0,0%	0,0%		
KP gibt keine Auskunft: anderer/kein Grund	69	0,2%	0,4%	0,4%		
KP verweigert absolut jede Teilnahme	584	1,3%	3,3%	3,4%		
<b>Zielpersonenkontakte insg. (Brutto Zielperson)</b>	<b>15.918</b>	<b>34,7%</b>	<b>90,9%</b>	<b>93,9%</b>	<b>100,0%</b>	
ZP ist beschäftigt, nicht erreichbar	107	0,2%	0,6%	0,6%	0,7%	
ZP gibt keine Auskunft: Thema/Projekt	–	0,0%	0,0%	0,0%	0,0%	
ZP gibt keine Auskunft: anderer/Grund	–	0,0%	0,0%	0,0%	0,0%	
ZP verweigert Teilnahme*	449	1,0%	2,6%	2,6%	2,8%	
ZP ist krank, nicht in der Lage	361	0,8%	2,1%	2,1%	2,3%	
Interview-Verweigerung zu Beginn/Abbruch	5.531	12,1%	31,6%	32,6%	34,7%	
Interview nicht auswertbar	44	0,1%	0,3%	0,3%	0,3%	
<b>System. Ausfälle insg. (Summe 9–11, 12–17, 18–24)</b>	<b>8.081</b>	<b>17,6%</b>				
<b>Durchgeführte Interviews (Netto Zielperson)</b>	<b>9.426</b>	<b>20,6%</b>	<b>53,8%</b>	<b>55,6%</b>	<b>59,2%</b>	<b>100,0%</b>
davon Vollinterviews (Netto Stichprobe)**	8.318	18,2%	47,5%	49,1%	52,3%	88,2%
davon Kurzfragebögen	1.108	2,4%	6,3%	6,5%	7,0%	11,8%

\* nach Statusklärung, ob es sich um einen Privathaushalt handelt

\*\* endgültig, nach Bereinigung

#### 4.7 Datenkontrolle und -aufbereitung

Die Rohdaten des GSTelo<sub>3</sub> wurden nach Abschluss der Feldarbeit einer umfangreichen Plausibilitätsprüfung unterzogen. Neben »out-of-Range«-Ergebnissen und einer »missing values«-Analyse bildeten vor allem Anmerkungen der Interviewer zur Zuverlässigkeit der Angaben der Befragten eine wichtige Grundlage für die Bereinigung der Daten. Hierzu enthält der programmierte Fragebogen ein Kommentarfeld, in das die Interviewer bei Bedarf entsprechende Anmerkungen und Kommentare eintragen konnten. Eine weitere Quelle für die Bereinigung des Rohdatensatzes stellt das während der gesamten Erhebungszeit durch die Supervisoren geführte Log-Buch dar<sup>22</sup>. Darin waren alle besonderen Vorkommnisse während des Feldverlaufs einzutragen u. a. auch zu einzelnen, über die Fragebogennummer identifizierbaren Datensätzen.

Nach Durchsicht der Datensätze und schriftlichen Unterlagen wurden 44 vollständige Interviews aus dem Rohdatensatz entfernt. Die wichtigsten Ausschlussgründe zeigt Tabelle 4.

Am häufigsten trat ein falsch kodiertes Geschlecht als Ausschlussgrund auf. Das ist dem Umstand zu zuschreiben, dass das Geschlecht der Auskunftsperson nicht direkt erfragt sondern durch Einschätzung der Interviewer vergeben wurde. Im Laufe des Interviews, bei geschlechtsspezi-

fischen Fragen ergab sich dann aber die Klärung des »richtigen« Geschlechts. Eine nachträgliche Korrektur der Datensätze musste jedoch wegen der größtenteils geschlechts- und altersabhängigen Filterführung im programmierten Fragebogen unterbleiben und die betroffenen Fälle wurden daher aus dem Datensatz ausgeschlossen. Verständnisprobleme, also Zweifel daran, dass die Befragten den Sinngehalt einzelner Fragen bzw. ganzer Themenkomplexe richtig verstanden haben, bildeten in der Summe die zweithäufigste Ursache für den Ausschluss von Datensätzen. Die Zuordnung zu dieser Kategorie beruht nicht ausschließlich auf den Einschätzungen der Interviewer sondern wurde durch die Feldleitung nach der kritischen Durchsicht derjenigen Fragebögen getroffen, die von den Interviewern mit entsprechenden Anmerkungen versehen wurden. Ein weiterer Problemkomplex, der zu einigen Ausschlüssen führte, betrifft die Zielpersonenermittlung und Haushaltsdefinition. In der Summe wurden acht Fälle aus diesen Gründen aus dem Rohdatensatz ausgeschlossen.

Ein weiterer Fall wurde nach Prüfung der Datensätze wegen des hohen Anteils von »missing values« in den für die Gewichtung benötigten Variablen aus dem Datensatz ausgeschlossen. Nach Bereinigung des Rohdatensatzes um diese Fälle verbleiben 8.318 auswertbare Vollinterviews.

**Tabelle 4**  
Ausschlussgründe von Datensätzen im GSTelo<sub>3</sub>

Ausschlussgrund	Anzahl
Geschlecht falsch kodiert	10
Verständnisprobleme (mental oder körperlich)	10
Nichtzugehörigkeit zum Haushalt, Geschäftsanschluss, Anstaltsbewohner oder Fehler bei Zielpersonenermittlung	8
Sprachliche Probleme	6
Versteckte oder latente Verweigerer (Interviewabbruch, Zweifel an der Ernsthaftigkeit)	6
Falsches Geburtsjahr	2
Betrunken	1
Ausschluss wegen fehlender Angaben in zentralen Merkmalen	1
<b>Summe</b>	<b>44</b>

22 vgl. Borch, Rieck 2005 in diesem Band sowie das OP-Handbuch zum GSTelo<sub>3</sub> (Borch et al. 2004)



## 4.8 Gewichtung der Daten

Analog zum Verfahren beim Bundes-Gesundheitssurvey 1998 wurden die dreidimensionalen Randverteilungen bestehend aus Alter, Geschlecht und Region an die Verteilung der Grundgesamtheit der erwachsenen Bevölkerung in Privathaushalten angeglichen. Dabei wurden die in vergleichbaren Erhebungen üblichen Verfahrenswesen verwendet (Böltken 1976; Gabler et al. 1994).

Vor der Anpassungsgewichtung muss jedoch zunächst eine Kompensation der rein designbedingten Verzerrungen des jeweils verwendeten Stichprobenverfahrens vorgenommen werden (Gabler, Häder 1999; Glemser 2002). Die so genannte Designgewichtung wird nachfolgend beschrieben.

### 4.8.1 Designgewichtung

Die erste Gewichtungsstufe ergibt sich aus der Anlage des Stichprobendesigns des Gabler-Häder Verfahrens, das in der ersten Auswahlstufe zufallsgesteuert einen Haushalt – definiert über eine unter der jeweiligen Anwahlnummer angetroffene Privatperson – auswählt. Innerhalb dieses Haushalts wurde dann wiederum zufällig die jeweilige Befragungsperson nach der next-birthday-Methode ermittelt.

In Verbindung mit der Zunahme der ISDN-Anschlüsse, die es auch Privathaushalten ermöglicht, kostengünstig über mehrere Telefonnummern zu verfügen, haben Haushalte mit ISDN-Anschluss bei mehreren Privatnummern im selben Haushalt eine höhere Wahrscheinlichkeit befragt zu werden. Diese erhöhte Auswahlwahrscheinlichkeit war daher durch den Kehrwert der »Anzahl der privat genutzten Festnetzanschlüsse im Haushalt« zu kompensieren<sup>23</sup>.

Vor dem Hintergrund des Ziels, bevölkerungsrepräsentative Aussagen auf Personenebene für die erwachsene Bevölkerung in der Bundesrepublik zu gewährleisten, ist in einem weiteren Schritt eine Anpassung an die Haushaltsgröße erforderlich, da Telefonstichproben prinzipiell als Haushaltsstichproben angesehen werden können. Personen, die in Haushalten mit mehreren Erwachsenen leben, haben eine um die Anzahl der

Erwachsenen im Haushalt geringere Chance, befragt zu werden. Somit ergäbe sich ein designbedingter Gewichtungsfaktor  $GWd_i^*$  als

$$GWd_i^* = \frac{AEPHH_i}{ATHH_i}, \text{ mit}$$

$AEPHH_i$  = Anzahl der erwachsenen Personen im Privat-Haushalt,  
 $ATHH_i$  = Anzahl der privat genutzten Festnetzanschlüsse im Haushalt.

Sowohl die Erhebung der Haushaltsgröße als auch die der Anzahl der Festnetzanschlüsse eines Haushalts können fehlerbehaftet sein<sup>24</sup>. So ist es eher unwahrscheinlich, dass Einpersonenhaushalte zehn Telefone oder Zehn-Personenhaushalte nur einen Telefonanschluss besitzen. Um die sich aus solchen Extremen ergebenden überproportional hohen Gewichtungsfaktoren zu reduzieren, wurden in Übereinstimmung mit der Literatur (Rösch 1994; Rothe 1994; Gabler, Häder 1997) die Gewichtungsfaktoren nicht mit den Rohwerten der Befragungsergebnisse bestimmt, sondern eine Adjustierung vorgenommen. Nach Testung mehrerer Alternativen<sup>25</sup> wurde schließlich die Wurzeltransformation zur Adjustierung der Rohwerte gewählt. Sie lieferte die besten Ergebnisse gemessen an den Effizienzkriterien minimaler und maximaler Gewichtungsfaktor und Effizienzmaß (s.u.). Die Wurzeltransformation gewährleistet ein Gewicht von 1 für den Einpersonenhaushalt mit einem Telefonanschluss, dämpft aber die sich aus

23 nach neuesten Ergebnissen der Einkommens- und Verbraucherstichprobe verfügen mittlerweile 23,4 % der Privathaushalte über ISDN-Anschlüsse (Statistisches Bundesamt 2003)

24 Im GSTeLo3 wurde die Anzahl der Festnetzanschlüsse, die ausschließlich privat genutzt werden als 2-stellige offene Angabe mit einem Grenzwert von 20 festgehalten. Von einer Kategorisierung z. B. 1, 2, 3 und mehr Anschlüsse wurde wegen der weiten Verbreitung von ISDN-Anschlüssen abgesehen. Es war allerdings überraschend, dass auch Angaben von mehr als 20 Telefonanschlüssen genannt wurden. Darüber hinaus entstehen Ungenauigkeiten auch bei der Abgrenzung von Telefonanschlüssen, die nicht ausschließlich als Fax- oder Modemanschluss betrieben und auch zum Telefonieren genutzt werden. Weiterhin wurde beobachtet, dass in einigen Fällen die Rufnummer, die angewählt wurde, dem oder der Befragten selbst gar nicht bekannt war oder ist. Dieses Phänomen tritt in den laufenden CATI-Projekten mit zunehmender Tendenz auf.

25 u. a. Kappungsgrenzen bei 3, 4 oder 5 und mehr Anschlüssen, log-Transformationen

Extremen ergebenden Faktoren, bei Beibehaltung der Größenrelation. Darüber hinaus ist dieses Verfahren sowohl transparent als auch reversibel, was auf das in vergleichbaren Situationen zumeist verwendete Verfahren der Begrenzung auf einen Maximalwert (z. B. 1, 2, 3 und mehr Telefonanschlüsse etc.) nicht zutrifft.

Zusätzlich ist bei der Designgewichtung noch die Normierung auf den Stichprobenumfang  $n = 8.318$  zu berücksichtigen. Somit ergibt sich das im GSTelo3 verwendete adjustierte und normierte Designgewicht  $GWd_i^*$  nach der Formel:

$$GWd_i^* = \sqrt{\frac{AEPHH_i}{ATHH_i}} \times \frac{n}{\sum \sqrt{\frac{AEPHH_i}{ATHH_i}}}$$

#### 4.8.2 Anpassungsgewichtung

Auf Basis der Designgewichtung wurde in der zweiten Gewichtungsstufe eine Anpassung an die Ergebnisse der amtlichen Statistik mit Stand vom 31.12.2001 vorgenommen. Grundlage hierzu bilden tief gestaffelte Tabellen der Alters- und Geschlechtsverteilung pro Bundesland, die vom Statistischen Bundesamt zur Verfügung gestellt wurden<sup>26</sup>. Dies ermöglichte die Aufbereitung der Referenzdaten in einer den Anforderungen des Datensatzes entsprechenden Weise, sodass das Redressmentgewicht  $GWd_i^*$  als einfache Soll-durch-Ist-Gewichtung gebildet werden konnte. Hierzu wurden aus der dreidimensionalen Verteilung nach den Merkmalen Alter, Geschlecht und Region 72 Gewichtungszellen gebildet. Diese ergaben sich zur Vermeidung von zu geringen Besetzungen in den einzelnen Gewichtungszellen, in dem das Alter in sechs Gruppen<sup>27</sup> eingeteilt und die Bundesländer zu sechs Regionen zusammengefasst wurden<sup>28</sup>. Die Redressmentgewichtung erfolgt nach der Formel:

$$GWr_i = \frac{p_i^{RAS\text{ soll}}}{p_i^{RAS\text{ ist}}}, \text{ mit}$$

$p_i^{RAS\text{ soll}}$  = rel. Zellbesetzung der Kombination aus den Merkmalen Region, Alter, Geschlecht in der Bevölkerung

$p_i^{RAS\text{ ist}}$  = rel. Zellbesetzung der Kombination aus den Merkmalen Region, Alter, Geschlecht in der designgewichteten Stichprobe

Das für die Auswertungen zu verwendende Endgewicht  $w03_i$  ergibt sich schließlich nach der Formel:

$$w03_i = GWd_i^* \times GWr_i$$

#### 4.8.3 Güte der Gewichtung

Als Gütekriterien für Gewichtungen werden in der Literatur (Rösch 1994) der minimale und maximale Gewichtungsfaktor, die Spannweite und Streuung der Faktoren sowie das Effizienzmaß  $E$  genannt. Letzteres ist definiert als die durchschnittliche Summe der Abweichungsquadrate vom »Idealwert« pro Zelle in Prozent:

$$E = \frac{\sum w03_i}{\sum w03_i^2} \times 100$$

Je näher die Effektivität am Wert 100 % liegt, desto geringer sind die gewichtungsbedingten Verzerrungen der Ausgangsdaten. Werte des Effizienzmaßes über 80 % gelten als gute Anpassung.

Das Ergebnis zur Güte der Gewichtung des GSTelo3 zeigt die Tabelle 5 – Gewichtungsvariable  $w03$ . Zum Vergleich werden die Gewichtungsergebnisse des BGS98 mit aufgeführt – Gewichtungsvariable  $w98$  (Stolzenberg 2002).

**Tabelle 5**  
Gütekriterien der Gewichtungen im GSTelo3 und BGS98

Kriterien	w03	w98
Effektivität	87,10%	84,10%
Min. Faktor	0,482	0,275
Max. Faktor	1,669	4,15
Fälle	8.318	7.124

<sup>26</sup> Wir danken dem Statistischen Bundesamtes für die Bereitstellung der Sonderauswertungen

<sup>27</sup> bis 29 Jahre, 30 bis 39 Jahre, 40 bis 49 Jahre, 50 bis 59 Jahre, 60 bis 69 Jahre sowie 70 und mehr Jahre

<sup>28</sup> Nord = Schleswig-Holstein, Hamburg, Bremen, Niedersachsen; Ost = Mecklenburg-Vorpommern, Sachsen Anhalt, Berlin, Brandenburg, Sachsen und Thüringen; Mitte = Hessen, Saarland Rheinland-Pfalz. Die Bundesländer Bayern, Baden-Württemberg und Nordrhein-Westfalen konnten wegen genügend großer Zellbesetzung erhalten bleiben.

Weitere Gütekriterien stellen der maximale und minimale Gewichtungsfaktor dar. Das ist leicht einsehbar, da damit der kleinste bzw. größte Faktor bezeichnet wird, mit der ein Fall in die gewichteten Ergebnisse Eingang findet. Je geringer die Spannweite und je näher der minimale Faktor am Wert »1« liegt, desto geringer ist damit die gewichtungsbedingte Verzerrung. Selbstverständlich kann die Gewichtung keine systematischen Ausfälle bestimmter Personengruppen kompensieren. Daher ist es zur Beurteilung der Güte der erzielten Ergebnisse zusätzlich erforderlich, die verfügbaren Informationen über nicht Befragungswillige nach den Gründen der Nichtteilnahme zu untersuchen und Teilgruppen mit geringerer Teilnahmebereitschaft zu identifizieren. Das wird, soweit es auf Grundlage der verfügbaren Informationen möglich ist, im Folgenden dargestellt.

#### 4.9 Ergebnisse zu Response und Nonresponse

Die Ausschöpfungsquote sagt, unabhängig von deren Höhe, noch nichts über das Ausmaß möglicher Selektivität aus (DFG 1999). Daher stellt sich die Frage: Was wissen wir über die Nonresponder und welche Quellen sind verfügbar? Im Vordergrund stehen dabei vor allem gesundheitliche Gründe, die eine Teilnahme verhinderten sowie soziodemografische Merkmale der Nichtteilnehmer. Neben Geschlecht und Alter interessieren in diesem Zusammenhang auch sprachliche Probleme, die eine Beteiligung der mittlerweile auf nahezu 10 % gestiegenen Anteil der ausländischen Bevölkerung in der Bundesrepublik am GSTelo3 beeinträchtigen können.

Erste Hinweise auf systematische Selektionseffekte ergeben sich aus der Dokumentation der Anwahlergebnisse, wie sie sich in den disposition codes (Tabelle 2) wieder finden. So konnte mit 252 kontaktierten Zielpersonen wegen sprachlicher Gründe kein Interview durchgeführt werden. Weitere 361 ermittelte Zielpersonen sahen sich aus gesundheitlichen Gründen nicht in der Lage, an der Befragung teilzunehmen. Über diese Personen liegen keine weiteren Angaben vor.

Die wichtigste Quelle, die Aufschluss über Gründe der Nicht-Teilnahme liefert, bildet daher der im programmierten Fragebogen enthaltene Kurzfragebogen. Bei jeder Kontaktaufnahme mit einer

Zielperson und bei jedem Abbruch während eines Interviews wurde automatisch vor Beendigung des Gesprächs um dessen Beantwortung gebeten.

Neben den Gründen die zum Abbruch oder zur Nichtteilnahme führten, werden im Kurzfragebogen noch soziodemografische Merkmale sowie einige gesundheitsbezogene Daten erhoben (subjektive Gesundheit, Body-Mass-Index, Raucherstatus, Vorliegen einer chronischen Erkrankung, Hausarzt, AU-Zeiten im letzten Jahr). Da nicht alle Befragten dazu motiviert werden konnten, den Kurzfragebogen vollständig zu beantworten, ergeben sich jedoch unterschiedliche Fallzahlen bei den einzelnen Fragen.

#### 4.9.1 Bekannte Gründe für Nonresponse

Gründe zur Nichtteilnahme werden als Mehrfachnennung mit geschlossenen und einer offenen Antwortmöglichkeit zu Beginn des Kurzfragebogens erhoben. Insgesamt gaben 1.166<sup>29</sup> Probanden Auskunft über ihre Gründe zur Nichtteilnahme. Wie zu erwarten, stehen dabei zeitliche Gründe an erster Stelle. Mit deutlichem Abstand gefolgt von grundsätzlicher Ablehnung von Befragungen. Erfreulicherweise liegen gesundheitliche Gründe an letzter Stelle, sodass daraus der Schluss abgeleitet werden kann, dass ein gesundheitlicher Bias in der realisierten Stichprobe des GSTelo3 zunächst eher ausgeschlossen scheint<sup>30</sup>.

**Tabelle 6**  
Gründe für Nicht-Teilnahme (Mehrfachnennungen)

Begründung für Nicht-Teilnahme	Anzahl	in %
aus zeitlichen Gründen	675	57,9
nehme grundsätzlich nicht teil	259	22,2
aus gesundheitlichen Gründen	83	7,1
sehe keinen Nutzen darin	116	9,9
sonstige Gründe	115	9,9
<b>Antwortende insgesamt</b>	<b>1.166</b>	<b>100</b>

<sup>29</sup> Die Abweichung zu der in Tabelle 2 ausgewiesenen Zahl von 1.108 ergibt sich deshalb, weil nicht alle Teilnehmer den Kurzfragebogen vollständig beantwortet haben.

<sup>30</sup> Im BGS98 lag der Anteil der gesundheitlichen Gründe für die Nicht-Teilnahme bei über 10 %, Stolzenberg (2002)

#### 4.9.2 Analyse soziostruktureller und gesundheitsbezogener Merkmale von Teilnehmern und Nicht-Teilnehmern

Nachfolgend wird anhand der im Kurzfragebogen enthaltenen Angaben der Frage nachgegangen, inwieweit sich die Befragungsteilnehmer von den Nichtteilnehmern hinsichtlich ausgewählter soziodemografischer und gesundheitsbezogener Merkmale unterscheiden. Eine diesbezügliche Analyse ermöglicht erste Hinweise auf Selektionseffekte im Datensatz des GSTelo<sup>3</sup>. Hinsichtlich der demografischen Zusammensetzung unterscheiden sich Nonresponder durch höheres Alter und einen leicht

erhöhten Anteil an Frauen bei den Befragungsteilnehmern. Erwartungsgemäß liegt auch der Ausländeranteil bei den Nonrespondern etwas höher, wofür v.a. sprachliche Probleme ausschlaggebend sein dürften. Allerdings ist dieser Zusammenhang nicht signifikant.

Demgegenüber waren weniger Gesunde und durch mindestens eine chronische Erkrankung betroffene Personen eher bereit, an der Befragung teilzunehmen. Befragungsteilnehmer weisen darüber hinaus einen höheren Anteil von Rauchern aus.

Vor allem bei hochbetagten Frauen und Personen mit niedriger Bildung dürften mögliche Ver-

**Tabelle 7**  
Soziodemografische Merkmale und Befragungsbeteiligung

Teilnahme	Geschlecht		Anzahl	Alter			Anzahl
	Frauen	Männer		<40	40–69	70+	
Nonresponder	56,2%	43,8%	1.541	30,2%	52,2%	17,6%	1.313
Responder	53,4%	46,6%	8.362	35,1%	56,0%	8,9%	8.362
Gesamt*	53,9%	46,1%	9.903	34,4%	55,5%	10,1%	9.675
Signifikanz (chi <sup>2</sup> )	P = 0,046			P = 0,000			
Teilnahme	Bildung		Anzahl	Ausländer		Anzahl	
	Niedrig	Hoch		Nein	Ja		
Nonresponder	69,1%	30,9%	1.117	95,3%	4,7%	1.132	
Responder	58,6%	41,4%	8.196	96,4%	3,6%	8.360	
Gesamt*	59,8%	40,2%	9.313	96,3%	3,7%	9.492	
Signifikanz (chi <sup>2</sup> )	P = 0,000			P > 0,05			

**Tabelle 8**  
Gesundheitsbezogene Merkmale und Befragungsbeteiligung

Teilnahme	Subjektive Gesundheit			Anzahl	Chron. Erkrankung		Anzahl
	Schlecht	Mittel	Gut		Nein	Ja	
Nonresponder	8,1%	18,5%	73,4%	1.123	67,0%	33,6%	1.324
Responder	21,7%	50,3%	27,9%	8.348	61,8%	38,2%	8.316
Gesamt*	20,1%	46,6%	33,3%	9.471	62,6%	37,4%	9.640
Signifikanz (chi <sup>2</sup> )	P = 0,000				P = 0,046		
Teilnahme	Raucher		Anzahl	Gewicht (BMI)			Anzahl
	Nein	Ja		Unter	Norm	Über	
Nonresponder	71,8%	28,2%	1.180	2,7%	73,4%	23,9%	1.101
Responder	66,0%	34,0%	8.360	3,8%	72,3%	23,9%	8.185
Gesamt*	66,8%	33,2%	9.540	3,6%	72,4%	23,9%	9.286
Signifikanz (chi <sup>2</sup> )	P = 0,000			P > 0,05			

\* Wechselnde Fallzahlen ergeben sich daraus, dass nicht alle Befragten den Kurzfragebogen vollständig beantwortet haben

zerrungen, aufgrund deren geringerer Beteiligung an der Befragung am stärksten sein.

Einschränkend ist an dieser Stelle noch anzumerken, dass die hier vorliegenden Ergebnisse nur erste Hinweise auf gruppenspezifische Nonresponse geben können. Zum einen stellen die Nonresponder eine Selektion insofern dar, als sie zumindest noch bereit waren, ein Minimum an Fragen zu beantworten. Zum anderen liegen über den größeren Teil der nicht Teilnahmewilligen keine zusätzlichen Informationen vor. Hier können erst weitere Analysen im Vergleich zu externen Daten, z. B. dem Mikrozensus, tieferen Aufschluss über gruppenspezifischen Nonresponse liefern.

### 4.9.3 Intensität der Samplennutzung

Selektivität bei – telefonischen – Befragungen kann auch aus der systematisch besseren Erreichbarkeit bestimmter Bevölkerungsgruppen resultieren. In diesem Zusammenhang spielt die Intensität der Sampleausschöpfung und die Kontakthäufigkeit einer Telefonnummer eine bedeutende Rolle. Die maximale Kontaktzahl im GSTelo3 wurde aus Qualitätsgründen auf 15 festgesetzt und geht damit – wie oben bereits erwähnt – deutlich über das übliche Maß hinaus.

Nachfolgend wird anhand ausgewählter Merkmale untersucht, welche Konsequenzen sich aus der intensiveren Sampleausnutzung auf die Ergebnisse des GSTelo3 ergeben. Hierzu wurde die Anzahl der Kontakte die zur Durchführung des Interview benötigt wurden in drei Gruppen – bis zu 5, zwischen 5 bis 9 und mehr als 9 Kontaktversuche – zusammengefasst und in den Zeilen der Tabellen 8 bis 13 dargestellt. Die letzte Zeile in den abgebildeten Tabellen gibt immer die Prozentsatzdifferenz der ersten Zeile, bei bis zu 5 Kontaktversuchen und der aktuellen Verteilung – mit maximaler Anzahl der Kontakte – des jeweils dargestellten Merkmals an.

Die intensivere Sampleausnutzung im GSTelo3 verändert die soziodemografische Zusammensetzung in Richtung jüngerer, männlicher Teilnehmer und führt zu einer leichten Zunahme des Anteils der Migranten. Die mittleren und höheren Bildungsgrade steigen ebenfalls mit zunehmender Anzahl der Kontaktversuche an und führen in der Folge zu einer Zunahme des Anteils der höheren Sozialschicht. Dies schlägt sich auch bei den Krankenversicherungsarten in einer Zunahme des Anteils der privat Versicher-

**Tabelle 9**  
Soziodemografische Merkmale und Kontakthäufigkeit

Anrufe	Geschlecht		Alter in Dekaden						Migrant	
	Männer	Frauen	-29	-39	-49	-59	-69	70+	Ja	Nein
Bis 5	45,0%	55,0%	13,3%	19,5%	25,0%	16,6%	15,6%	10,0%	8,7%	91,3%
5–9	53,3%	46,7%	20,5%	24,4%	24,7%	15,7%	10,0%	4,8%	10,7%	89,3%
Mehr als 9	54,8%	45,2%	17,4%	30,4%	24,4%	15,3%	10,0%	2,5%	8,9%	91,1%
Gesamt	46,6%	53,4%	14,4%	20,7%	24,9%	16,4%	14,6%	8,9%	9,0%	91,0%
	1,6%	-1,6%	1,1%	1,2%	-0,1%	-0,2%	-1,0%	-1,1%	0,3%	-0,30%
Signifikanz (chi <sup>2</sup> )	P = 0,000		P = 0,000						P = 0,000	

**Tabelle 10**  
Sozialschichtbezogene Merkmale und Kontakthäufigkeit

Anrufe	CASMIN-Bildungsgruppen			Soz. Schicht n. Winkler-Index (adj.)				Krankenversicherung verdichtet					
	Einfache Bildung	Mittlere Bildung	Höhere Bildung	Fehlend	US (3...8)	MS (9...14)	OS (15...21)	AOK	EKK	BKK	GKV Rest	PKV/ Beihilfe	Rest
Bis 5	31,9%	47,5%	20,6%	1,8%	15,4%	53,4%	29,4%	23,5%	37,3%	16,8%	6,5%	13,8%	2,0%
5-9	25,0%	51,5%	23,5%	1,3%	14,3%	50,7%	33,7%	20,9%	35,5%	19,1%	6,6%	16,3%	1,7%
Mehr als 9	23,4%	48,9%	27,7%	1,9%	11,0%	45,9%	41,2%	19,1%	35,9%	16,6%	6,2%	21,2%	1,1%
Gesamt	30,6%	48,1%	21,3%	1,7%	15,0%	52,7%	30,6%	22,9%	37,0%	17,1%	6,5%	14,6%	1,9%
	-1,3%	0,6%	0,8%	-0,1%	-0,4%	-0,8%	1,2%	-0,6%	-0,3%	0,3%	0,0%	0,7%	-0,1%
Signifikanz (chi <sup>2</sup> )	P = 0,000			P = 0,000				P = 0,000					

**Tabelle 11**  
Erwerbsbeteiligung und Kontakthäufigkeit

Anrufe	Erwerbstätigkeit – Arbeitszeitkonzept									
	Rest	Ausbildung	Vorruhestand	Rente	Arbeitslos	Hausfrau	Teilzeit <15 h	Teilzeit >15 h	Vollzeit	
Bis 5	0,1%	2,8%	3,5%	19,1%	4,0%	8,8%	7,2%	11,5%	42,9%	
5-9	0,1%	3,7%	1,6%	10,8%	3,4%	4,9%	7,5%	11,0%	57,0%	
Mehr als 9	0,2%	2,8%	1,5%	8,5%	2,5%	7,0%	5,9%	9,1%	62,4%	
Gesamt	0,1%	2,9%	3,1%	17,5%	3,9%	8,2%	7,2%	11,3%	45,8%	
	0,0%	0,1%	-0,3%	-1,6%	-0,2%	-0,6%	0,0%	-0,2%	2,8%	
Signifikanz (chi <sup>2</sup> )	P = 0,000									

ten nieder. Hinsichtlich der Erwerbstätigkeit ist ein Ansteigen der Vollzeiterwerbstätigen mit zunehmender Kontakthäufigkeit festzustellen.

Offensichtlich handelt es sich bei den schwerer Erreichbaren um Jüngere, höher Gebildete und überwiegend Vollzeit erwerbstätige Personen, denen eine durchschnittlich bessere Gesundheit unterstellt werden kann. Dies bestätigt sich bei der Betrachtung der gesundheitsbezogenen Merkmale, die einen höheren Anteil normalgewichtiger und gesünder Probanden, mit keinen gesundheitlichen Einschränkungen oder chronischen Erkrankungen aufweisen.

Die Ergebnisse der Nonresponderanalyse lassen im GSTelo3 tendenziell eine Unterschätzung des Anteils der gesünderen Bevölkerung erkennen. Die intensive Feldarbeit, mit einer über das übliche Maß hinausgehenden Sampleausnutzung kompensiert diesen Effekt offensichtlich teilweise, insbesondere in Bezug auf die im gegebenen Kontext relevanten gesundheitsbezogenen Merkmale. Diese Ergebnisse decken sich mit Resultaten, die

Voigt et al. (2003) in epidemiologischen Telefon-surveys feststellen konnten. Darin heißt es, dass intensivere Anstrengungen in der Kontaktabnahme sich in deutlich besseren Schätzern der bevölkerungsbezogenen Merkmale auswirken, die in epidemiologischen Studien von Interesse sind.

Andererseits ergibt sich hinsichtlich der demografischen Zusammensetzung ein überlagernder Effekt aus der geringeren Teilnahmebereitschaft von Personen mit niedriger Schulbildung und deren Kompensation durch höher Gebildete der Mittel- und Oberschicht im Zuge der intensiveren Sampleausschöpfung. Dies trägt auch zur Erklärung der bereits anderenorts festgestellten höheren Beteiligungsbereitschaft der mittleren und oberen Sozialschichten an telefonischen Befragungen bei (Schulte 1997; Porst et al. 1998; Wüst 1998). Für die Auswertungen und Analyse der Daten des GSTelo3 folgt daraus, dass bei schichtsensitiven Fragestellungen die soziale Schicht immer in die Modellierung einbezogen werden sollte.

Tabelle 12a  
Gesundheitsbezogene Merkmale und Kontakthäufigkeit

Anrufe	Subjektive Gesundheit					Ges. Einschränkung			Chron. Erkrankung	
	Sehr gut	Gut	Mittel- mäßig	Schlecht	Sehr schlecht	Stark eingeschr.	Nicht eingeschr.		Ja	Nein
Bis 5	21,5%	51,6%	20,6%	5,0%	1,3%	9,6%	24,6%	65,8%	39,2%	60,8%
5–9	23,3%	55,9%	15,7%	4,4%	0,7%	7,8%	21,0%	71,2%	33,3%	66,7%
Mehr als 9	27,2%	56,5%	13,4%	2,5%	0,4%	3,6%	21,7%	74,7%	34,3%	65,7%
Gesamt	22,0%	52,4%	19,6%	4,8%	1,1%	9,0%	24,0%	67,0%	38,2%	61,8%
	0,6%	0,8%	-1,0%	-0,2%	-0,1%	-0,6%	-0,6%	1,2%	-1,0%	1,0%
Signifikanz (chi <sup>2</sup> )	P = 0,000					P = 0,000			P = 0,000	

Tabelle 12a  
Gesundheitsbezogene Merkmale und Kontakthäufigkeit

Anrufe	Body-Mass-Index adjustiert und gruppiert nach BGSg8					
	Fehlend	Untergewicht (<19)	Normalgewicht (19–>25)	Leichtes Übergewicht (25–<27,5)	Mittleres Übergewicht (27,5–< 30)	Adipositas (>=30)
Bis 5	2,1%	1,1%	37,9%	24,0%	16,5%	18,3%
5–9	2,1%	1,5%	44,3%	25,2%	12,7%	14,1%
Mehr als 9	0,8%	0,4%	45,0%	21,7%	15,7%	16,3%
Gesamt	2,1%	1,1%	39,1%	24,1%	16,0%	17,7%
	-0,1%	0,0%	1,2%	0,0%	-0,5%	-0,6%
Signifikanz (chi <sup>2</sup> )	P = 0,000					

#### 4.10 Zusammenfassung

Die Ergebnisse zur Response und Teilnahmebereitschaft am GSTelo3 lassen eine hohe Übereinstimmung mit den Befunden, die aus der Literatur zu telefonischen Befragungen entnommen werden können, erkennen. Es erweist sich, dass diese Methode nicht allein wegen der geringeren Erhebungskosten und der größeren Zeitnähe gut für den Aufbau eines kontinuierlichen Gesundheitsmonitoring geeignet ist, sondern auch wegen der leichten Verfügbarkeit eines nationalen Auswahlrahmens, der regionalisierbare

Analysen ermöglicht. Das setzt allerdings eine entsprechende Sorgfalt bei der Durchführung der Erhebung voraus. Dies betrifft nicht nur die Intensität der Sampleausschöpfung und die Feldsteuerung sondern beinhaltet den gesamten Erhebungsprozess, beginnend mit der Erstellung des Erhebungsinstruments unter Berücksichtigung methodenspezifischer Aspekte, der Schulung der eingesetzten Erhebungskräfte sowie der feldbegleitenden qualitätssichernden Maßnahmen, die in eigenen Beiträgen dieses Bandes ausführlicher behandelt werden (vgl. Borch, Rieck 2005).

## 4.11 Literatur

- AAPOR (2004) Standard Definitions: Final Disposition Codes and Outcome Rates for Surveys. Lenexa, Kansas, American Association for Public Opinion Research
- ADM/BVM (1998) Richtlinie für telefonische Befragungen. In: Gabler S, Häder S, Hoffmeyer-Zlotnik HJP (Hrsg) Telefonstichproben in Deutschland. Westdeutscher Verlag, Opladen, S 181–187
- Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. (ADM), Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V. (ASI), Statistisches Bundesamt (1999) Demografische Standards. Statistisches Bundesamt, Wiesbaden
- Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. (ADM), Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V. (ASI), Statistisches Bundesamt (2004) Demografische Standards. Statistisches Bundesamt, Wiesbaden
- Borch S, Rieck A, Blümel M (2004) Operationshandbuch – Erster telefonischer Bundes-Gesundheitssurvey 2003, Robert Koch-Institut, Berlin
- Borch S, Rieck A (2005) Feldarbeit (in diesem Band)
- Böltken F (1976) Auswahlverfahren – Eine Einführung für Sozialwissenschaftler. Stuttgart, Teubner
- CASRO (1982) On the Definition of Response Rates – A Special Report of the CASRO Task Force. New York, Council of American Survey Research Organizations
- CDC (1998) Behavioral Risk Factor Surveillance System Users Guide. Atlanta, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention
- Couper MP, Baker RL, Bethlehem J, Clark CZ, Martin J, Nicholls II WL, O'Reilly JM (1998) Computer assisted survey information collection. New York, John Wiley & Sons
- DFG (1999) Qualitätskriterien der Umfrageforschung. Denkschrift. Berlin, Deutsche Forschungsgemeinschaft
- Fischer R, Meyer N, Weitkunat R, Crispin A, Schotten K, Überla K (2001) Bevölkerungsbezogenes Gesundheitsmonitoring in Bayern mit computerassistierten Telefoninterviews; Das Gesundheitswesen 63 (Sonderheft 2): 123–129
- Gabler S, Häder S (1997) Wirkung von Gewichtungen bei Face-to-Face und Telefonstichproben. Eurobarometer-Experiment 1994. In: Gabler S, Hoffmeyer-Zlotnik HJP (Hrsg) Stichproben in der Umfragepraxis. Westdeutscher Verlag, Opladen, S 221–245
- Gabler S, Häder S (1999) Generierung von Telefonstichproben mit TelSuSa. ZUMA-Nachrichten 44: 138–143
- Gabler S, Häder S (2002) Telefonstichproben – Methodische Innovationen und Anwendungen in Deutschland. Münster/New York/München
- Gabler S, Hoffmeyer-Zlotnik HJP, Krebs D (1994) Gewichtungen in der Umfragepraxis. Westdeutscher Verlag, Opladen
- Gabler S, Hoffmeyer-Zlotnik HJP (1997) Stichproben in der Umfragepraxis. Westdeutscher Verlag, Opladen
- Glemser A (2002) ADM-Telefonstichproben in der Praxis. In: Gabler S, Häder S (Hrsg) Telefonstichproben – Methodische Innovationen und Anwendungen in Deutschland. Münster/New York/München, S 46–58
- Häder S (1996) Wer sind die »Nonpubs«? Zum Problem anonymer Anschlüsse bei Telefonumfragen. ZUMA-Nachrichten 39: 45–68
- Häder S, Gabler S (2000) Überlegungen zur Anwendung von RLD-Verfahren bei Telefonumfragen in Deutschland. In: Hüfken V (Hrsg) Methoden in Telefonumfragen. Westdeutscher Verlag, Wiesbaden, S 33–48
- Heckel C (2001) Erstellung der ADM-Telefonauswahlgrundlage. ZUMA-Workshop »Methodische Probleme bei der Stichprobenziehung und Realisierung«, Aschpurwitz + Behrens GmbH, Mannheim
- Heckel C (2002) Erstellung der ADM-Telefonauswahlgrundlage. In: Gabler S, Häder S (Hrsg) Telefonstichproben – Methodische Innovationen und Anwendungen in Deutschland. Münster/New York/München, S 11–31
- Heyde C (2002) Das ADM-Telefonstichproben-Modell. In: Gabler Häder (Hrsg) Telefonstichproben – Methodische Innovationen und Anwendungen in Deutschland. Münster/New York/München, S 32–45
- Latza U, Stang A, Bergmann M, Kroke A, Sauer S, Holle R, Kamtsiuris P, Terschüren C, Hoffmann W (2004) Zum Problem der Response in epidemiologischen Studien in Deutschland (Teil I), Das Gesundheitswesen (66): 326–336
- McCarty C (2003) Differences in Response Rates Using Most Recent Versus Final Dispositions in Telephone Surveys; Public Opinion Quarterly 67: 396–406
- Murza G, Faulbaum F, Deutschmann M, Dietmair I, Simon K (2003) Telefonische Bevölkerungsbefragungen als Instrument kommunaler Gesundheitsberichterstattung. Zeitschrift für Gesundheitswissenschaften 11 (Heft 2): 131–145
- Porst R, Ranft S, Ruoff B (1998) Strategien und Maßnahmen zur Erhöhung der Ausschöpfungsquoten bei sozialwissenschaftlichen Umfragen. Ein Literaturbericht, ZUMA Arbeitsbericht 98
- Rieck A, Borch S, Kohler M (2005) Methodische Aspekte der Fragebogenentwicklung (in diesem Band)
- Rösch G (1994) Kriterien der Gewichtung einer nationalen Bevölkerungsstichprobe. In: Gabler S, Hoffmeyer-Zlotnik HJP, Krebs D (Hrsg) Gewichtungen in der Umfragepraxis. Westdeutscher Verlag, Opladen
- Rothe G (1994) Wie (un)wichtig sind Gewichtungen? Eine Untersuchung am Allbus 1996. In: Gabler S, Hoffmeyer-Zlotnik HJP, Krebs D (Hrsg) Gewichtungen in der Umfragepraxis. Westdeutscher Verlag, Opladen, S 62–87
- Schulte W (1997) Telefon- und Face-to-Face-Umfragen und ihre Stichproben. Allgemeine Bevölkerungsumfragen in Deutschland. In: Stichproben in der Umfragepraxis. Gabler S, Hoffmeyer-Zlotnik HJP, Krebs D (Hrsg) Westdeutscher Verlag, Opladen, S 148–195
- Statistisches Bundesamt (2003) Ausstattung privater Haushalte mit Informations- und Kommunikationstechnik – Ergebnisse der Einkommens- und Verbrauchsstichprobe 1998 und 2003. DESTATIS, Statistisches Bundesamt Deutschland
- Stolzenberg H (2002) Bundes-Gesundheitssurvey 1998 – Public Use File BGS98 – Dokumentation des Datensatzes. Berlin, Robert Koch Institut
- Voigt LF, Koepsell TD, Daling JR (2003) Characteristics of Telephone Survey Respondents According to Willingness to Participate: American Journal of Epidemiology 157(1): 66–73
- Voxco (2001) INTERVIEWER PC Software for Computer Assisted Surveying – Reference Manual Version 4.2. Montreal, Canada
- Waksberg J (1978) Sampling Methods for Random Digit Dialing; Journal of the American Statistical Association 73: 40–46



- Weitkunat R (1998) Computergestützte Telefoninterviews als Instrument der sozial- und verhaltensepidemiologischen Gesundheitsforschung. Berlin
- Wiedenmayer G (2002) Münchener Gesundheitsmonitoring 1999/2000 – Die Gesundheit von Männern und Frauen verschiedener Lebensaltersstufen. Gesundheitsberichterstattung. RGU-III. München, Referat für Gesundheit und Umwelt der Landeshauptstadt München
- Wüst A M (1998) Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften als Telefonumfrage; ZUMA Arbeitsbericht 72



Bibliografische Information Der Deutschen Bibliothek  
Die Deutsche Bibliothek verzeichnet diese Publikation  
in der Deutschen Nationalbibliografie.

#### Herausgeber

Robert Koch-Institut  
Nordufer 20  
13353 Berlin

#### Redaktion

Robert Koch-Institut  
Gesundheitsberichterstattung  
Seestraße 10  
13353 Berlin

#### Autorinnen und Autoren

Martin Kohler  
Angelika Rieck  
Susan Borch  
Dr. Thomas Ziese  
Gesundheitsberichterstattung  
Robert Koch-Institut

E-Mail: [gbe@rki.de](mailto:gbe@rki.de)  
[www.rki.de](http://www.rki.de)  
Tel.: 018 88. 754-34 00  
Fax: 018 88. 754-35 13

#### Satz

Gisela Winter  
Robert Koch-Institut

#### Druck

Oktoberdruck Berlin  
gedruckt auf PROFIsilk, tcf

#### ISBN

3-89606-163-1

*Beiträge zur Gesundheitsberichterstattung*  
Robert Koch-Institut

**Erster telefonischer Gesundheitssurvey  
des Robert Koch-Instituts – Methodische Beiträge**

November 2005

**Berlin: Robert Koch-Institut**  
ISBN 3-89606-163-1

The Telephone Health Survey 2003 (GSTelo3) is the first nationwide survey conducted by the Robert Koch-Institute using computer-assisted telephone interviewing (CATI). The procedures used in telephone surveys differ in many aspects from face-to-face or mail surveys. The articles in this volume discuss methodological aspects of the GSTelo3 related to questionnaire design, data collection procedures, sampling design and response rate.

One article examines the need for a questionnaire that is suitable for the respective survey method and discusses the measures taken to achieve this in the preparation for the GSTelo3. These include questionnaire dramaturgy, filtering and the optimization of the operationalizations; they focus on breaking down complex questions and/or response categories into smaller units, deciding on the sequence of the fixed answer options, and standardizing retrospective time windows.

Another contribution describes the following aspects of data collection procedures: assigned staff, data collection and storage, interviewer and supervisor training, quality assurance on interviewing, training procedures and supervision.

A further article discusses in detail the technical framework, sampling design, call-back management, and, derived from this, the weighting procedures used in the first nationwide telephone health survey. Finally, the results of a nonresponse analysis are presented, and there is a discussion on the effects of the number of attempts made to contact households.

It is shown that telephone surveys are an appropriate and important method to establish a continuous health monitoring system.

Der telefonische Gesundheitssurvey 2003 (GSTelo3) ist die erste bundesweite Studie, die vom Robert Koch-Institut per CATI (computer-assisted telephone interviewing) durchgeführt wurde. Die Vorgehensweise bei computer-gestützten telefonischen Interviews unterscheidet sich in vielen Aspekten von anderen Befragungsformen. Die Beiträge in diesem Band diskutieren methodische Aspekte des GSTelo3 hinsichtlich Fragebogenentwicklung, Feldarbeit sowie Stichprobenplanung und -realisierung. Ein Aufsatz thematisiert die Notwendigkeit eines erhebungsmethodisch angemessenen Fragebogens und die dazu für den GSTelo3 umgesetzten Maßnahmen. Zu diesen zählen Fragebogendramaturgie, Filterführung und Optimierung der Operationalisierungen mit den Schwerpunkten Zerlegung von komplexen Fragen bzw. Antwortkategorien, Festlegung der Reihenfolge für die Antwortvorgaben und Vereinheitlichung retrospektiver Zeitfenster. Ein anderer Beitrag beschreibt folgende Aspekte der Feldarbeit: eingesetztes Personal, Datenerhebung und -haltung, Interviewer- und Supervisorenschulung sowie Qualitätssicherung der Interviewführung, Schulung und Supervision. Ein weiterer Aufsatz geht ausführlich auf die technischen Rahmenbedingungen, das Stichprobendesign, das Anrufmanagement bzw. die Ablaufsteuerung und – daraus abgeleitet – auf die Gewichtung des ersten bundesweiten telefonischen Gesundheitssurveys ein. Abschließend werden die Ergebnisse einer Nonresponderanalyse vorgestellt und die Auswirkungen der Intensität der Sampleaus-schöpfung diskutiert. Es wird gezeigt, dass telefonische Befragungen ein geeignetes und wichtiges Verfahren für ein kontinuierliches Gesundheitsmonitoring sind.

© Robert Koch-Institut

ISBN 3-89606-163-1

