

Methoden und Design des telefonischen Gesundheitssurveys 2003

Mittlerweile gehören telefonische Befragungen international zum Standard der Gesundheitsforschung [1]. Auch in Deutschland liegen inzwischen zumeist regional oder thematisch begrenzte Erfahrungen mit telefonischen Befragungen vor [2, 3, 4].

Zu den Vorteilen telefonischer Befragungen zählt u. a., dass sie es ermöglichen, aktuelle bevölkerungsrepräsentative Daten über Veränderungen in gesundheitsrelevanten Bereichen zeitnäher und mit geringerem Kostenaufwand zu gewinnen als mit anderen Erhebungsverfahren. Mit dem Gabler-Häder-Design steht für die Bundesrepublik Deutschland ein nationaler Auswahlrahmen zur Verfügung, der auch regionalisierte Analysen erlaubt [5]. Telefonische Befragungen sind darüber hinaus in der Felddurchführung effizienter, weil nur selten zu Hause anzutreffende Bevölkerungsgruppen mit geringerem Aufwand mehrfach kontaktiert werden können. Die Unterbringung der Interviewer in einem zentralen Telefonstudio ermöglicht zudem eine wirksamere Kontrolle der Interviewdurchführung und verbessert damit die Datenqualität, vor allem gegenüber schriftlichen oder Face-to-face-Erhebungsformen.

Den genannten Vorteilen telefonischer Befragungen stehen aber auch Nachteile gegenüber. So eignen sich nicht alle Inhalte oder Operationalisierungen aus schriftlichen Befragungen gleichermaßen für den Einsatz im Telefoninterview, denn Letzteres basiert ausschließlich auf einem akustischen Kontakt zwischen Interview-

er und Proband. Dies stellt hohe Anforderungen an die Entwicklung des Fragebogens mit einer häufig notwendigen Anpassung etablierter Instrumente an diese Erhebungsform. Darüber hinaus stellen Telefonbefragungen besondere Ansprüche an das Erhebungspersonal, da neben den inhaltlichen Gesichtspunkten auch die Auswahl und Ermittlung der Zielpersonen sowie die stichprobengerechte Dokumentation der Anwahlergebnisse in der Eröffnungsphase des Interviews souverän beherrscht werden müssen und ein hohes Abstraktionsvermögen verlangt wird, um die Perzeption der Befragtensituation adäquat berücksichtigen zu können. Entsprechend sind u. a. eine intensive Schulung der Erhebungskräfte und die kontinuierliche Überprüfung während des Erhebungsprozesses erforderlich. Weiterhin können spezifische Bevölkerungsgruppen (z. B. Migranten, soziale Randgruppen, junge hochmobile Erwachsene oder Personen, die nur noch über Handys verfügen) über das Telefon im Festnetz nur unzureichend erreicht werden.

Diese Aspekte sind bei der Konzeption, Durchführung und Interpretation der Ergebnisse telefonischer Befragungen zu berücksichtigen. Welche Vorkehrungen und Maßnahmen getroffen wurden, um den ersten bundesweiten telefonischen Gesundheitssurvey des Robert Koch-Instituts (RKI) gemäß den gewohnten Qualitätsansprüchen umzusetzen, wird im Folgenden in den wesentlichen Grundzügen dargestellt.

Methodische Aspekte der Fragebogenentwicklung

Die Qualität des Erhebungsinstrumentes ist für das Gelingen nicht nur von Telefonsurveys von entscheidender Bedeutung. Die Auswahl der in den GSTelo3 einbezogenen Themen und Dimensionen war inhaltlich und strategisch begründet, wie bei Ziese [6] und Ziese et al. [7] dargestellt. Daneben bedurfte die Entwicklung des Fragebogens der Berücksichtigung verschiedener Aspekte, die mit dem Erhebungsmodus zusammenhängen. Denn Telefoninterviews stellen teilweise andere Anforderungen an das Erhebungsinstrument als postalische Befragungen oder Face-to-face-Interviews. Diese betreffen den Aufbau und die Struktur des Fragebogens (sog. Dramaturgie) sowie die einzelnen Fragen und deren Antwortkategorien. Zum Beispiel lassen sich einige Operationalisierungen nicht ohne Anpassung bzw. Modifikation aus einer Erhebungsmethode in eine andere übertragen. Die hohe Relevanz der methodischen Aspekte der Instrumentenentwicklung für die Datenqualität, Item-Nonresponse und Ausschöpfung verlangte ihre adäquate Berücksichtigung bei der Entwicklung des CATI- (Computer Aided Telephone Interviewing-)Fragebogens [8, 9, 10, 11, 12, 13].

Entsprechend wurde für den GSTelo3 eine Reihe spezieller Maßnahmen, angefangen bei der Fragebogendramaturgie bis hin zu Administrationsanweisungen, umgesetzt. Hierzu ein Beispiel: Um ver-

gleichweise komplexe Frageformulierungen aus schriftlichen Fragebögen früherer Studien, insbesondere aus dem Bundes-Gesundheitssurvey 1998, fortschreiben zu können, wurden sie für den GSTel03 in einfachere Fragen zerlegt. Lange Antwortlisten aus schriftlichen Bögen wurden aufgespalten. Eine ausführliche Darstellung der bei der Fragebogenentwicklung berücksichtigten methodischen Aspekte findet sich bei Rieck et al. [13]. Zu den programmierten Plausibilitätskontrollen vgl. Borch/Rieck [14].

Feldarbeit

Neben einem gelungenen Erhebungsinstrument entscheidet die Qualität der Feldarbeit über den Erfolg einer Studie. Sie setzt eine adäquate Schulung der Mitarbeiter und eine kontinuierliche Qualitätssicherung voraus.

Im Anschluss an 2 ausführliche Pretests begann am 2. September 2002 die Feldzeit des GSTel03, die am 14. März 2003 endete. An 112 Tagen wurden von insgesamt 81 Interviewern¹ mit 45.000 Telefonnummern des Basissamples über 165.000 Anwahlvorgänge durchgeführt. Insgesamt wurden dabei 8362 Frauen und Männer befragt, 8318 Interviews waren verwertbar.

Die durchschnittliche Länge der Interviews im GSTel03 betrug zirka 21 Minuten. Die Dauer variierte jedoch für einzelne Befragtengruppen erheblich: Sie lag zwischen ca. 18 Minuten (18- bis 20-Jährige) bis 30 Minuten (über 80-Jährige) und je nach Geschlecht zwischen ca. 20 Minuten (Männer) und 22 Minuten (Frauen). Bei Vorliegen spezifischer Erkrankungen stieg die durchschnittliche Interviewdauer erheblich, da in diesem Fall filterbedingt mehr Fragen zu beantworten waren [13].

Die Datenerhebung erfolgte im neu geschaffenen CATI-Labor des RKI, das 19 Interviewerarbeitsplätze und einen Supervisorarbeitsplatz umfasst. Die Anrufe wurden wochentags zwischen 16 und 20 Uhr durchgeführt. Zusätzlich fanden einige Vormittags- und Samstagseinsätze statt,

¹ Die Geschlechtsverteilung im Interviewerstamm war ausgewogen, zum größten Teil handelte es sich um Studierende sozialwissenschaftlicher Fächer im Alter zwischen 20 und 30 Jahren.

Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz 2005 · 48:1224–1230
DOI 10.1007/s00103-005-1154-y
© Springer Medizin Verlag 2005

M. Kohler · A. Rieck · S. Borch

Methoden und Design des telefonischen Gesundheitssurveys 2003

Zusammenfassung

Telefonische Befragungen werden inzwischen auch in Deutschland zunehmend im Bereich der Gesundheitsforschung eingesetzt. Das Robert Koch-Institut hat im Jahr 2003 einen ersten bundesweiten telefonischen Survey zur Verbreitung vorwiegend chronischer Erkrankungen beendet (GSTel03), dessen Ergebnisse in diesem und dem folgenden Heft vorgestellt werden. Telefonische Surveys stellen eigene, neue Anforderungen an die Planung, Umsetzung und Durchführung gegenüber anderen Erhebungsformen. In diesem Beitrag werden

die wichtigsten methodischen Aspekte behandelt, die zur Gewährleistung einer hohen Datenqualität im Rahmen des GSTel03 ein- und umgesetzt wurden. Dazu zählen das Stichprobendesign, die Fragebogenentwicklung, die Felddurchführung einschließlich der Rekrutierung und Schulung der Erhebungskräfte sowie die Darstellung der Ausschöpfung und Response.

Schlüsselwörter

Telefonischer Gesundheitssurvey · CATI · Design · Response · Datenerhebung

Method and design of the Telephone Health Survey 2003

Abstract

Telephone surveys are increasingly being used in public health research, also in Germany. In 2003 the Robert Koch Institute completed the first nationwide telephone survey predominantly on conditions and prevalence of selected chronic diseases (GSTel03). Its findings are presented in this publication. The tasks of planning, implementation and realization of procedures used in telephone surveys differ in many aspects from other modes of data collection. This article deals with the most important methodological aspects that were implemented

and used in the course of the GSTel03 to ensure a high level of data quality. These include sample design, questionnaire development and the process of data collection (including the recruitment and training of the staff). Finally steps of data preparation for analysis including weighting procedures are described and calculation of contact and cooperation rates are reported.

Keywords

Telephone Health Survey · CATI · Design · Response · Data collection

um die Responserate für die in den oben genannten Zeiten schlecht erreichbaren Personen zu erhöhen.

Die Telefoninterviews wurden mittels der CATI-Software Interviewer Suite 4.3 durchgeführt. Die Grundlage bildete ein programmierter Fragebogen, der vergleichbar mit computergestützten Interviews sowohl eine Datenkontrolle während der Befragung als auch die Ablaufsteuerung mittels programmierter Filter übernimmt. Die CATI-Software steuert und verwaltet auch die An- und Rückrufe, d. h. so genannte call backs. Dies sind Anrufe mit einem noch offenen abschließenden Kontakt mit einer Zielperson im Haushalt, die entsprechend programmierter Regeln automatisch durch das Programm ausgewählt wurden (auto dialing). Anwahlfelder sind damit nahezu ausgeschlossen, was eine saubere Stichprobenrealisierung gewährleistet [13, 14, 15].

Für die Betreuung und Qualitätssicherung der Datenerhebung waren an jedem Feldtag 2 Supervisoren und ein Mitglied der wissenschaftlichen Feldleitung zuständig. Die Aufgaben der studentischen Supervisoren bestanden vor allem in der Anleitung der Interviewer während der Datenerhebung, in der Qualitätssicherung der Interviewführung sowie in der Planung des Interviewereinsatzes. Die wissenschaftliche Feldleitung war insbesondere verantwortlich für die Optimierung des Erhebungsinstrumentes, die Stichprobenkonzeption, die Schulung der Interviewer und Supervisoren sowie für die Qualitätssicherung und Koordination der Feldarbeit [14].

Mitarbeiterschulung und Qualitätssicherung

Das Verhalten der Interviewer und Supervisoren beeinflusst die Datenqualität, die Ausschöpfung und die Item-Nonresponse [10, 17]. Daher wurden beide Personengruppen intensiv auf ihre Aufgaben vorbereitet [14]. In 3-tägigen Eingangsschulungen mit einem Umfang von 15 Stunden, in denen sich Theorie- und Praxisblöcke abwechselten, wurden die Interviewer bezüglich des technischen Umgangs mit der CATI-Software, der Einhaltung des Datenschutzes, des Verhaltens in der Kontaktpphase und im weiteren Verlauf des Interviews

sowie der inhaltlichen Handhabung des Erhebungsinstrumentes unterwiesen. Die studentischen Supervisoren wurden frühzeitig, bereits während der Planungsphase in das Projekt einbezogen, da die Mitarbeit bei den Feldvorbereitungen auch eine implizite Schulungsfunktion erfüllt. Die konkreten Tätigkeiten im Kontext der Supervision wurden während des Pretestes und bei der Vorbereitung des Hauptfeldes als „training on the job“ erlernt [14].

Um die Qualität über die gesamte Dauer der Feldzeit sichern zu können, bedarf es neben der Vermittlung entsprechender Standards in den Eingangsschulungen einer kontinuierlichen Qualitätssicherung. Grundlage dafür war das Operationshandbuch [16]. Es stellt die Anleitung zum standardisierten Vorgehen bei der Studierendurchführung dar und dient als Schulungsmaterial und Nachschlagewerk. Für die Qualitätssicherung war auch die kontinuierliche Dokumentation durch die Supervisoren, z. B. in einem Logbuch, wichtig. Außerdem verlangt die Optimierung der Feldarbeit die kontinuierliche und systematische Reflexion des Vorgehens. In diesem Kontext hat sich auch der Austausch zwischen Supervisoren und Feldleitung in wöchentlichen Teamsitzungen bewährt. Eine ausführliche Beschreibung der qualitätssichernden Maßnahmen für den GSTelo3 findet sich bei Borch/Rieck [14].

Stichprobendesign und Anrufmanagement

Die Stichprobenkonzeption des GSTelo3 soll repräsentative Aussagen über die volljährige, deutschsprachige Wohnbevölkerung in Privathaushalten der Bundesrepublik Deutschland ermöglichen. Vom Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim wurde für diesen Zweck das Gabler-Häder-Design [5] entwickelt, das einen nationalen Auswahlrahmen für Telefonbefragung zur Verfügung stellt. Es liefert auch die Basis für das ADM-Mastersample (ADM: Arbeitskreis Deutscher Marktforschungsinstitute) und bildet damit die Grundlage fast aller bundesweit repräsentativen Telefonbefragungen.

Das Gabler-Häder-Verfahren beruht darauf, aus dem Pool öffentlich zugänglicher Nummernverzeichnisse eine Bestandsliste sämtlicher in der BRD vorhandenen Fest-

netzanschlüsse zu erstellen. Aus diesem um Sondernummern und offensichtliche gewerbliche Anschlüsse bereinigten Nummernpool werden die letzten beiden Ziffern abgeschnitten, wodurch so genannte Nummernblöcke entstehen. Diese werden abschließend wieder durch die Ziffernfolgen „00“ bis „99“ ergänzt und bilden den Auswahlrahmen für bevölkerungsrepräsentative Telefonstichproben in der Bundesrepublik Deutschland. Aus dieser hypothetischen Grundgesamtheit² wird dann uneingeschränkt zufällig die erforderliche Anzahl von Anwahlnummern gezogen. Die Repräsentativität auf Personenebene wird durch eine zweite Auswahlstufe erreicht, in der die Zielperson in Mehrpersonenhaushalten nach einem Zufallsverfahren ermittelt wird.

Das für den GSTelo3 gewählte stichprobentheoretische Konzept bestimmt auch maßgeblich das Vorgehen während der Feldarbeit. Als Ergebnis der Stichprobenziehung liegen im GSTelo3 45.000 zu kontaktierende Rufnummern vor, von denen zunächst nicht bekannt ist, ob sie tatsächlich existieren bzw. welche davon Anschlüsse von Privathaushalten sind. Nach Klärung des Letzteren wurde mit der zufällig angetroffenen Person die Zielperson im Haushalt gemäß der Next-Birthday-Methode ermittelt und ggf. ein Interview durchgeführt. Selbstverständlich gelingt dies nur selten beim ersten Anwahlvorversuch. In der Regel sind mehrmalige Kontaktversuche erforderlich.

Die Verwaltung der Rufnummern mit dem jeweils von den Interviewern zu protokollierenden Anwahlergebnis in den so genannten disposition codes übernimmt das Anruf- und Rückrufmanagement, das den jeweiligen Anforderungen entsprechend konzipiert und programmiert werden muss. Dies geschieht in einem eigenen Modul der CATI-Software. Hier werden jedem Ergebniscode Regeln zugeordnet, die festlegen, wann die jeweilige Rufnummer zur Wiedervorlage kommt. Diese Call-back-Regeln bestimmen u. a., zu welchem Zeitpunkt (nach 10 Minuten, einem Tag, mehreren Tagen etc.) eine mit einem Call-back-Status versehene Telefon-

² Für das Jahr 2001 besteht der Gesamtrahmen aus 71,7 Millionen Nummern, von denen 30,7 Millionen tatsächlich eingetragen sind [18].

nummer wieder in den Pool anzurufender Datensätze eingespielt wird.

Anders als in den USA liegen im deutschsprachigen Raum bisher keine einheitlichen Standards für die Entwicklung des Rückrufmanagements und der anzuwendenden disposition codes vor [19, 20]. Sie mussten daher eigens für den GSTelo3 entwickelt werden. Die Dokumentation der Anwahlvorgänge in der call history dient v. a. der Ermittlung der Ausschöpfungsquote, erfüllt aber noch andere Zwecke im Rahmen der feldbegleitenden Qualitätssicherung. So lässt sich z. B. aus ungewöhnlichen Häufungen bestimmter Anwahlergebnisse ein Nachschulungsbedarf für einzelne Interviewer erkennen.

Ausschöpfung und Response

Gegenüber registerbasierten Auswahlverfahren ermöglichen zufallsgenerierte Nummernsamples keine Bestimmung einer Rücklaufquote im herkömmlichen Sinn, da sich die Stichprobenbasis – Personen in Privathaushalten – erst im Vollzug der Feldarbeit realisiert. In der Literatur über Telefonsurveys wird daher statt vom Rücklauf auch von Kooperationsraten oder Ausschöpfungsquoten gesprochen.

Bei der Aufbereitung und Verdichtung der Anwahlergebnisse in der call history zur Bestimmung der Ausschöpfungsquote ist die Unterscheidung zwischen verlaufs- und ergebnisbezogener Betrachtung wichtig. So spiegelt der Endzustand der disposition codes in der call history möglicherweise nicht den tatsächlichen Zustand bei der Zuordnung der generierten Anwahlnummern nach dem Kriterium „Element der Grundgesamtheit“ wider. Das ist z. B. der Fall, wenn bei einem Anruf ein Termin mit der Zielperson vereinbart werden konnte, dieser aber nicht eingehalten wurde und in der Folge unter der Telefonnummer nur noch ein Fax angetroffen wird. Bei ausschließlicher Berücksichtigung des Endzustandes würde in diesem Fall eine falsche Zuordnung zu den stichprobenneutralen Ausfällen vorgenommen. Daher sind vor der Ermittlung der Ausschöpfungsquote die letzten Anwahlergebnisse unter Berücksichtigung des gesamten Feldverlaufs in einen Endzustand zu überführen [21]. In Übereinstimmung mit den Regeln der American Association for Public

Opinion Research (AAPOR) wurden im GSTelo3 die Ergebnisse der call history nach folgenden Gesichtspunkten verdichtet, wenn kein Abbruch vorlag oder kein Interview realisiert wurde:

- Endresultate durch das letzte Anrufergebnis mit einer Zielperson im Haushalt ersetzen, sonst
- letztes Ergebnis ersetzen durch letzten Kontakt mit einer beliebigen Kontaktperson eines Privathaushalts.

Als Ergebnis liegen nun bereinigte call results vor, die eine exaktere Bestimmung der Ausschöpfungs- und Responsequoten ermöglichen.

Werden die 9426 durchgeführten Interviews auf die 15.918 Zielpersonenkontakte bezogen, ergibt sich mit etwas über 59% eine bei vergleichbaren Telefonbefragungen im deutschsprachigen Raum übliche Kooperationsrate (■ **Tabelle 1**). Die Ausschöpfungsquote liegt mit knapp 53% ebenfalls im üblichen Bereich. Der Gesundheitsmonitor München [2] berichtet eine Ausschöpfungsquote von 52%, das Münchener Lebensstilpanel [3] kommt – bei einem aufwändigeren Rekrutierungsverfahren mit vorherigem Anschreiben der Probanden – auf eine Quote von 67,2%, und das Landesinstitut für den Öffentlichen Gesundheitsdienst NRW berichtet über Ausschöpfungsquoten zwischen 50 und 60% [4].

Datenaufbereitung und Gewichtung

Die Rohdaten des GSTelo3 wurden nach Abschluss der Feldarbeit einer umfangreichen Plausibilitätsprüfung unterzogen und nach Bereinigung an die Verteilung der Grundgesamtheit der erwachsenen Bevölkerung in Privathaushalten angeglichen. Vor dieser Anpassungsgewichtung muss jedoch zunächst eine Kompensation der designbedingten Verzerrungen vorgenommen werden.

Aus der Anlage des Stichprobendesigns des Gabler-Häder-Verfahrens als implizite Haushaltsstichprobe, das in der ersten Auswahlstufe zufallsgesteuert einen Haushalt auswählt und in der zweiten Stufe innerhalb dieses Haushaltes die jeweilige Befragungsperson nach der Next-birthday-Methode ermittelt, ergibt sich ein An-

passungsbedarf an die daraus folgende unterschiedliche Auswahlwahrscheinlichkeit der Zielperson in Mehrpersonenhaushalten.

In Verbindung mit der Zunahme der ISDN-Anschlüsse haben Haushalte mit mehreren Privatnummern eine höhere Wahrscheinlichkeit, befragt zu werden. Diese erhöhte Auswahlwahrscheinlichkeit war daher durch den Kehrwert der „Anzahl der privat genutzten Festnetzanschlüsse im Haushalt“ zu kompensieren. Das Designgewicht ergibt sich damit als Kehrwert der unterschiedlichen Auswahlwahrscheinlichkeiten oder als Quotient aus der Anzahl der erwachsenen Haushaltsmitglieder und der Anzahl der privat genutzten Festnetzanschlüsse im Haushalt.

Auf Basis der Designgewichtung wurde in der zweiten Gewichtungsstufe eine Anpassung an die Ergebnisse der amtlichen Statistik mit Stand vom 31.12.2001 vorgenommen. Die Grundlage dafür bilden tief gestaffelte Tabellen der Alters- und Geschlechtsverteilung pro Bundesland. Sie wurden vom Statistischen Bundesamt zur Verfügung gestellt und ermöglichen die Aufbereitung der Referenzdaten in einer den Anforderungen des Datensatzes entsprechenden Weise. Das Redressmentgewicht konnte daher als einfache Soll-durch-Ist-Gewichtung der 72 aus Altersgruppen, Geschlecht und Region gebildeten Gewichtungszellen vorgenommen werden. Eine ausführlichere Darstellung zur Gewichtung findet sich bei Kohler 2005 [15].

Non-Response und Samplenutzung

Merkmale von Non-Respondern

Die Ausschöpfungsquote sagt, unabhängig von ihrer Höhe, noch nichts über das Ausmaß möglicher Selektivität aus [22]. Daher stellt sich die Frage: Was wissen wir über die Non-Responder und welche Quellen zur Beantwortung stehen zur Verfügung? Erste Hinweise auf systematische Selektionseffekte ergeben sich aus der Dokumentation der Anwahlergebnisse, die sich in den disposition codes (■ **Tabelle 1**) wieder finden. So konnte mit 252 kontaktierten Zielpersonen wegen sprachlicher Probleme kein Interview durchgeführt werden.

Tabelle 1

Ausschöpfung des Nummernsamples im GSTel03 – Endzustand der call results

	Absolut	Bruttostich- probe [%]	Nettostich- probe [%]	Personen- kontakt [%]	Zielpersonen- kontakt [%]	Vollständige Interviews [%]
Bruttoansatz	45.819	100,0				
1 Telekom-Ansage/Falsche Nummer	18.604	40,6				
2 Max-Kontakt erreicht	4863	10,6				
3 FAX/Modem	715	1,6				
4 Kein Privathaushalt	2601	5,7				
5 Doppeladresse	–	0,0				
6 Falsche Regionalzelle/Quote voll	9	0,0				
7 Sprachprobleme	252	0,5				
8 Sonstige neutrale Ausfälle	1268	2,8				
<i>Neutrale Ausfälle insgesamt (Summe 1–8)</i>	<i>28.312</i>	<i>61,8</i>				
Bereinigtes Brutto	17.507	38,2	100,0			
9 Teilnehmer nimmt nicht ab, Status noch offen	242	0,5	1,4			
10 Anschluss besetzt	15	0,0	0,1			
11 Anrufbeantworter	293	0,6	1,7			
<i>Personenkontakte insgesamt (Brutto Privathaushalte)</i>	<i>16.957</i>	<i>37,0</i>	<i>96,9</i>	<i>100,0</i>		
12 Terminvereinbarung nicht wahrnehmbar	73	0,2	0,4	0,4		
13 Kein Termin möglich während Feldzeit	313	0,7	1,8	1,8		
14 KP ist beschäftigt, nicht erreichbar	–	0,0	0,0	0,0		
15 KP gibt keine Auskunft: Thema/Projekt	–	0,0	0,0	0,0		
16 KP gibt keine Auskunft: anderer/kein Grund	69	0,2	0,4	0,4		
17 KP verweigert absolut jede Teilnahme	584	1,3	3,3	3,4		
<i>Zielpersonenkontakte insgesamt (Brutto Zielperson)</i>	<i>15.918</i>	<i>34,7</i>	<i>90,9</i>	<i>93,9</i>	<i>100,0</i>	
18 ZP ist beschäftigt, nicht erreichbar	107	0,2	0,6	0,6	0,7	
19 ZP gibt keine Auskunft: Thema/Projekt	–	0,0	0,0	0,0	0,0	
20 ZP gibt keine Auskunft: anderer/kein Grund	–	0,0	0,0	0,0	0,0	
21 ZP verweigert Teilnahme ^a	449	1,0	2,6	2,6	2,8	
22 ZP ist krank, nicht in der Lage	361	0,8	2,1	2,1	2,3	
23 Interview-Verweigerung zu Beginn/Abbruch	5531	12,1	31,6	32,6	34,7	
24 Interview nicht auswertbar	44	0,1	0,3	0,3	0,3	
<i>Systematische Ausfälle insgesamt (Summe 9–11, 12–17, 18–24)</i>	<i>8081</i>	<i>17,6</i>				
Durchgeführte Interviews (Netto Zielperson)	9426	20,6	53,8	55,6	59,2	100,0
Davon Vollinterviews (Netto Stichprobe) ^b	8318	18,2	47,5	49,1	52,3	88,2
Davon Kurzfragebögen	1108	2,4	6,3	6,5	7,0	11,8

^a Nach Statusklärung, ob es sich um einen Privathaushalt handelt, ^b endgültig, nach Bereinigung; KP Kontaktperson, ZP Zielperson.

Weitere 361 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 für die Gründe der Nichtteilnahme bildet jedoch der im programmierten Fragebogen enthaltene

Kurzfragebogen. Bei jeder Kontaktaufnahme mit einer Zielperson, die kein Interview durchführen wollte, 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 zentrale soziodemographische Merkmale sowie einige gesundheitsbezogene Daten erhoben.

Insgesamt gaben 1166 Probanden im Kurzfragebogen Auskunft über ihre Gründe zur Nichtteilnahme. Wie zu erwarten, stehen dabei zeitliche Gründe an erster Stelle. Mit deutlichem Abstand folgt die

grundsätzliche Ablehnung von Befragungen. Erfreulicherweise liegen gesundheitliche Probleme an letzter Stelle, sodass ein gesundheitlicher Bias in der realisierten Stichprobe des GSTelo3 zunächst eher ausgeschlossen scheint. Non-Responder – also Personen, die den Hauptfragebogen nicht beantwortet haben – unterscheiden sich durch höheres Alter und einen leicht erhöhten Frauenanteil von den Befragungsteilnehmern. Erwartungsgemäß liegt auch der Ausländeranteil bei den Non-Respondern etwas höher, wofür v. a. sprachliche Probleme ausschlaggebend sein dürften. Demgegenüber waren weniger Gesunde und durch mindestens eine chronische Erkrankung betroffene Personen eher bereit, an der Befragung teilzunehmen (Details vgl. Kohler [15]).

Einschränkend ist an dieser Stelle noch anzumerken, dass die hier skizzierten Ergebnisse nur erste Hinweise auf eine gruppenspezifische Non-Response geben können, da über den größeren Teil der nicht Teilnahmewilligen keine zusätzlichen Informationen vorliegen.

Intensität der Samplenutzung

Selektivität bei (telefonischen) Befragungen kann auch aus der systematisch besseren Erreichbarkeit bestimmter Bevölkerungsgruppen resultieren. In diesem Zusammenhang sind insbesondere die Intensität der Sampleausschöpfung und die Kontakthäufigkeit einer Telefonnummer von Bedeutung. Letztere wurde im GSTelo3 aus Qualitätsgründen auf maximal 15 Kontakte festgesetzt. Diese Zahl geht damit deutlich über das übliche Maß hinaus [5, 23].

Die intensivere Sampleausnutzung im GSTelo3 verändert die soziodemografische Zusammensetzung der Befragten in Richtung jüngerer, männlicher Teilnehmer und führt zu einer leichten Zunahme des Anteils der Migranten. Die Anzahl an mittleren und höheren Bildungsgraden steigt ebenfalls mit der Zahl der Kontaktversuche und führt in der Folge zu einer Zunahme des Anteils der höheren Sozialschicht und der privat Versicherten. Darüber hinaus ist ein Anstieg der Zahl der Vollzeitbeschäftigten mit zunehmender Kontakthäufigkeit festzustellen. Offensichtlich

handelt es sich also 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 gesundheitsbezogener Merkmale dieser Teilgruppe, die einen höheren Anteil normalgewichtiger und gesünderer Probanden mit weniger gesundheitlichen Einschränkungen oder chronischen Erkrankungen aufweist.

Zusammenfassende Betrachtung zu Non-Response und Samplenutzung

Die Ergebnisse der Non-Responderanalyse lassen tendenziell eine größere Teilnahmebereitschaft der durch Krankheit betroffenen Bevölkerung am GSTelo3 erkennen. Die intensive Feldarbeit mit einer über das übliche Maß hinausgehenden Sampleausnutzung kompensiert diesen Effekt jedoch zum Teil. Diese Ergebnisse decken sich mit Resultaten von Voigt et al. [24]. Demnach resultieren aus intensiveren Anstrengungen bei der Kontaktabstimmung deutlich bessere Schätzer der bevölkerungsbezogenen Merkmale, die in epidemiologischen Studien von Interesse sind.

Andererseits ergeben sich hinsichtlich der demografischen Zusammensetzung überlagernde Effekte aus der geringeren Teilnahmebereitschaft von Personen mit niedriger Schulbildung und deren Kompensation durch höher Gebildete der Mittel- und Oberschicht aufgrund einer intensiveren Sampleausschöpfung. Dies trägt auch zur Erklärung der bereits andersorts festgestellten höheren Beteiligungsbereitschaft der mittleren und oberen Sozialschichten an telefonischen Befragungen bei [12, 25].

Fazit

Die eingangs genannten Vorteile telefonischer Studien kommen nur zum Tragen, wenn alle methodischen Aspekte dieser Erhebungsmethode angemessen berücksichtigt werden. Beispielsweise wird eine bessere Feldkontrolle in der Praxis nur erreicht, wenn die Interviewer die sorgfältig konzipierten disposition codes in der Schulung adäquat vermittelt bekommen

und ihre Anwendung eine kontinuierliche qualitätssichernde Begleitung erfährt. Viele der genannten Nachteile von Telefonsurveys können durch geeignete Maßnahmen neutralisiert bzw. zumindest abgemildert werden. Wird beispielsweise die Erhebungsmethode bei der Entwicklung des Fragebogens angemessen berücksichtigt, lassen sich die allermeisten Themen sehr gut per Telefon präsentieren.

Die Vorteile telefonischer Interviews kommen erst durch Computerunterstützung (CATI) richtig zum Tragen. Die damit einhergehende Technisierung stellt ihrerseits eigene Anforderungen an das Feldteam und die gesamte technische Infrastruktur. Die unzureichende Berücksichtigung von Teilaspekten kann erhebliche Implikationen für das Ergebnis der Feldarbeit haben. Die Qualität der Feldarbeit ist damit nicht nur von den im engeren Sinn wissenschaftlichen Vorarbeiten und den begleitenden Qualitätssicherungsmaßnahmen abhängig, sondern auch auf kooperative Unterstützung zahlreicher Personen angewiesen, deren „tägliches Geschäft“ anderen Dingen gewidmet ist.

Die gründliche Vorbereitung und gewissenhafte Durchführung der Befragung haben zum Erfolg des GSTelo3 beigetragen. Dieser Erfolg, der im vorliegenden sowie im darauffolgenden Schwerpunkttheft des Bundesgesundheitsblattes in den einzelnen Beiträgen dokumentiert wird, ist letztlich auch vom kontinuierlichen Engagement aller Beteiligten abhängig. Ihnen gebührt deshalb an dieser Stelle besonderer Dank.

Korrespondierender Autor

M. Kohler

Abteilung Epidemiologie
und Gesundheitsberichterstattung,
Robert Koch-Institut,
Postfach 650261, 13302 Berlin
E-Mail: KohlerM@rki.de

Literatur

1. CDC (1998) Behavioral risk factor surveillance system users guide. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention
2. Wiedenmayer G (2002) Münchener Gesundheitsmonitoring 1999/2000 – Die Gesundheit von Männern und Frauen verschiedener Lebensalterstufen. Referat für Gesundheit und Umwelt der Landeshauptstadt München, München
3. Weitkunat R (1998) Computergestützte Telefoninterviews als Instrument der sozial- und verhaltens-epidemiologischen Gesundheitsforschung. Berlin
4. Murza G, Faulbaum F, Deutschmann M et al. (2003) Telefonische Bevölkerungsbefragungen als Instrument kommunaler Gesundheitsberichterstattung. *Z Gesundheitswissenschaften* 11:131–145
5. Gabler S, Häder S (1999) Generierung von Telefonstichproben mit TeLSuSa. *ZUMA-Nachrichten* 44:138–143
6. Ziese T (2005) Der telefonische Gesundheitssurvey 2003 – Baustein zum Gesundheitsmonitoringssystem für Deutschland. In: Robert Koch-Institut (Hrsg) Erster telefonischer Gesundheitssurvey des Robert Koch-Instituts – Methodische Beiträge. Beiträge zur Gesundheitsberichterstattung. Eigenverlag, Berlin
7. Ziese T et al. (2005) Der telefonische Gesundheitssurvey 2003 als Instrument der Gesundheitsberichterstattung des Bundes. *Bundesgesundheitsblatt* 48:48
8. Dillman DA (1978) Mail and telephone surveys. The total design method. Wiley, New York
9. Frey JH, Kunz G, Lüschen G (1990) Telefonumfragen in der Sozialforschung. Westdeutscher Verlag, Opladen
10. Friedrichs J (1990) Gesprächsführung im telefonischen Interview. In: Forschungsgruppe Telekommunikation (Hrsg) Telefon und Gesellschaft, Band 2. Spiess, Berlin, S 413–425
11. Schwarz N, Strack F, Hippler HJ, Bishop G (1991) The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychol* 5:193–212
12. Wüst AM (1998) Die allgemeine Bevölkerungsumfrage der Sozialwissenschaften als Telefonumfrage. *ZUMA-Arbeitsbericht* 98/04. ZUMA, Mannheim
13. Rieck A, Borch S, Kohler M (2005) Methodische Aspekte der Fragebogenentwicklung. In: Robert Koch-Institut (Hrsg) Erster telefonischer Gesundheitssurvey des Robert Koch-Instituts – Methodische Beiträge. Beiträge zur Gesundheitsberichterstattung, Eigenverlag, Berlin
14. Borch S, Rieck A (2005) Feldarbeit. In: Robert Koch-Institut (Hrsg) Erster telefonischer Gesundheitssurvey des Robert Koch-Instituts – Methodische Beiträge. Beiträge zur Gesundheitsberichterstattung. Eigenverlag, Berlin
15. Kohler M (2005) Stichprobenkonzeption und Stichprobenrealisierung. In: Robert Koch-Institut (Hrsg) Erster telefonischer Gesundheitssurvey des Robert Koch-Instituts – Methodische Beiträge. Beiträge zur Gesundheitsberichterstattung. Eigenverlag, Berlin
16. Borch S, Rieck A, Blümel M (2004) Operationshandbuch erster Telefonischer Bundes-Gesundheitssurvey 2003. Robert Koch-Institut, Berlin
17. Billiet J, Loosveldt G (1988) Improvement of the quality of responses to factual survey questions by interviewer training. *Public Opin Quarterly* 52:190–211
18. 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
19. Council of American Survey Research Organizations (CASRO) (1982) On the definition of response rates – a special report of the CASRO Task Force. Council of American Survey Research Organizations, New York
20. American Association for Public Opinion Research (AAPOR) (2004) Standard Definitions: Final Disposition Codes and Outcome Rates for Surveys. Lenexa, Kansas
21. McCarty C (2003) Differences in response rates using most recent versus final dispositions in telephone surveys. *Public Opin Quarterly* 67:396–406
22. DFG (1999) Qualitätskriterien der Umfrageforschung. Deutsche Forschungsgemeinschaft, Berlin
23. Fritz J, Hartmann J, Sacher M (2004) Arbeitsmarkt in Deutschland – Analysen zu Ausfallprozessen in der Piloterhebung zum ILO-Erwerbsstatus. *Wirtschaft Statistik* 6:649–662
24. Voigt LF, Koepsell TD, Daling JR (2003) Characteristics of telephone survey respondents according to willingness to participate. *Am J Epidemiol* 157:66–73
25. Schulte W (1997) Telefon- und Face-to-Face-Umfragen und ihre Stichproben. *Allgemeine Bevölkerungsumfragen in Deutschland*. In: Gabler S, Hoffmeyer-Zlotnik JHP (Hrsg) Stichproben in der Umfragepraxis. Westdeutscher Verlag, Opladen, S 148–195