

7 Darstellung der methodischen Vorgehensweise (Studiendesign)

Einleitung

Bei der Studie »Gesundheit in Deutschland aktuell 2010« (GEDA 2010) handelt es sich um eine repräsentative Befragung der deutschsprachigen, erwachsenen Wohnbevölkerung in Privathaushalten, die über einen Festnetzanschluss verfügt. Die Befragung wurde zwischen September 2009 und Juli 2010 vom Robert Koch-Institut durchgeführt. Insgesamt wurden 22.050 Interviews erhoben. Die im Rahmen des Gesundheitsmonitorings regelmäßig wiederholte GEDA-Studie ist auf die kontinuierliche Beobachtung von Entwicklungen im Krankheitsgeschehen und im Gesundheits- und Risikoverhalten ausgerichtet und soll dazu beitragen, der Gesundheitsberichterstattung sowie der Gesundheitspolitik zeitnah entsprechende Informationen zur Identifizierung von Gesundheitstrends in der Bevölkerung oder in Bevölkerungsgruppen zu liefern. Aufgrund der Stichprobengröße können auch regionalisierte oder tief gegliederte Zusammenhangsanalysen vorgenommen werden (Kurth et al. 2009).

Der Survey schließt an die seit 2003 durchgeführten telefonischen Gesundheitssurveys (»GSTelo3« bis »GSTelo6«, RKI 2005) an und wird in gleicher Weise erhoben, d. h. die Interviews werden computergestützt per Telefon (CATI: computer assisted telephone interviewing) durchgeführt. Es wurde ebenfalls angestrebt, soweit wie möglich bewährte Frageinstrumente beizubehalten, um Zeitreihen erstellen zu können.

Telefonische Befragungen bieten eine Reihe von Vorteilen, aber auch einige Nachteile, die bei der Studiendurchführung berücksichtigt werden müssen. Der Vorteil telefonischer Befragungen liegt in den relativ geringen Kosten, der Möglichkeit, aktuelle Themen zeitnah zu befragen, der programmierten Filterführung der Fragen, einer effizienten Feldarbeit und einer effizienten Qualitätskontrolle. Da durch die feine Streuung der Nummern über alle Regionen hinweg Klumpungseffekte vermieden werden, ist eine ökonomische Fallzahlplanung möglich. Zudem können relativ preiswert regionale Aufstockungen vorgenommen werden. Die Nachteile telefonischer Befragungen bestehen in dem ausschließlich akustischen Kontakt zwischen Befragten und Interviewern und den daraus resultierenden Einschränkungen bei der Instrumentenauswahl und der fehlenden Möglichkeit, Messergebnisse zu erheben. Außerdem sinkt in der Bevölkerung die Akzeptanz von telefonischen Befragungen

und spezielle Bevölkerungsgruppen sind schlecht erreichbar. Insbesondere steigt die Zahl der Personen, die nur noch über mobile Telefone zu erreichen ist (vgl. Häder et al. 2009). In Zusammenarbeit mit GESIS Mannheim arbeitet das Robert Koch-Institut an Lösungsansätzen der Probleme, die aus dem Auswahlrahmen resultieren.

Seit Mitte der 1990er-Jahre liegt der Anteil an Telefoninterviews an empirischen Befragungen konstant bei über 40 %, bei über die Jahre sinkenden Anteilen persönlicher Interviews und steigenden Anteilen von Online-Befragungen. Im Jahr 2010 lag der Anteil der Online-Interviews mit 38 % erstmal über dem der telefonischen Interviews mit nur noch 35 % (vgl. www.adm-ev.de).

Stichprobendesign

Den Stichprobenrahmen für die Studie GEDA 2010 bildet die volljährige, deutschsprachige Wohnbevölkerung in Privathaushalten der Bundesrepublik Deutschland, die 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 ist (ADM/BVM 1998; Heckel 2001). Der Ausgangspunkt dieses Verfahrens besteht darin, aus dem Pool öffentlich zugänglicher Nummernverzeichnisse eine Bestandsliste sämtlicher in der Bundesrepublik vorhandener Festnetzanschlüsse zu erstellen. Dabei 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 (sogenannte Non-Pubs) mit gleicher Wahrscheinlichkeit erreichen zu können, haben Gabler und Häder versucht, den in den USA bei telefonischen Befragungen verwendeten Waksberg-Ansatz (Waksberg 1978) auf die bundesrepublikanischen Verhältnisse zu übertragen. 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, Doppelinträge 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 sogenannte 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 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 Nummern erst im Laufe der Feldarbeit ergibt. 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 zuletzt Geburtstag hatte (Last-Birthday-Methode). Ein Vorteil des beschriebenen Stichprobendesigns liegt zudem darin, dass mit dem Auswahlrahmen nach dem Gabler-Häder-Design auch regionale Klumpungseffekte nahezu ausgeschlossen sind und bei genügend großer Gesamtzahl auch kleinräumig regionalisierbare Analysen durchgeführt werden können (vgl. Abschnitt 5 in diesem Band).

Regionale Aufstockung

Die Möglichkeit regionalisierbarer Stichproben des Telefonnummernsamples wurde im Rahmen von GEDA 2010 genutzt. Das Saarland hat sich wie bereits im Rahmen von GEDA 2009 mit einer Aufstockung der Fallzahlen beteiligt. Ziel ist es, auf Landesebene repräsentative Aussagen zur Häufigkeit von Krankheiten und Gesundheitsstörungen sowie zum Gesundheitsverhalten treffen zu können, die nach Geschlecht und drei Altersgruppen stratifizierbar sind. Hierfür wurde im Vorfeld der benötigte Stichprobenumfang berechnet¹ und durch GESIS Mannheim nach dem bereits beschriebenen Verfahren zufällig gezogen. Für das Saarland bedeutete dies, zusätzlich 2.024 Interviews zu erbringen. Die Stichproben wurden zusammen mit dem Haupt-sample in einzelnen Tranchen über den Verlauf der Feldlaufzeit eingespielt, um saisonale Effekte weitestgehend auszuschließen.

Auswahl der Befragungsperson

Das für die GEDA-Studie 2010 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 in GEDA 2010 als Ergebnis der Stichprobenziehung eine Menge von zu kontaktierenden Rufnummern vor (ca. 250.000), bei denen zunächst nicht bekannt ist, welche tatsächlich existieren und welche davon Anschlüsse von Privathaushalten sind². Nach dieser Klärung erfolgte dann beim ersten Kontakt mit der zufällig angetroffenen Person die Ermittlung der Zielperson im Haushalt nach der »Last-Birthday-Methode«. Gemeint ist damit, dass mit derjenigen volljährigen Person im Haushalt, die als letzte Geburtstag hatte, das Interview geführt werden soll.

¹ Um eine Prävalenz von 50 Prozent mit einem Konfidenzintervall von maximal +/- 5 Prozent schätzen zu können, bedarf es 385 Probanden pro Zelle. Um also Aussagen für drei Altersgruppen und Geschlecht treffen zu können, werden mindestens 2.310 Probanden für eine bestimmte Region benötigt. Ausgehend von den erwarteten Fallzahlen für die Bundesländer in GEDA 2010 wurde die Anzahl der zusätzlich zu erbringenden Interviews für das Ländermodul geschätzt.

² Zum Begriff des Privathaushalts vgl. auch (Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. ADM et al. 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).

Damit soll eine Zufallsauswahl auf Haushaltsebene erreicht werden. Nur wenn die Zielperson direkt erreicht wird, darf nach vorheriger mündlicher Einwilligung sofort ein Interview durchgeführt werden. Andernfalls ist ein geeigneter Rückruftermin zu vereinbaren, an dem die ermittelte Zielperson voraussichtlich zu erreichen ist.

Die Zielpersonermittlung oder gar die Durchführung des Interviews mit der Zielperson gelingt nur selten beim ersten Anwahlversuch einer Telefonnummer. Dies kann häufig erst durch mehrmalige Kontaktversuche festgestellt werden. Es ist davon auszugehen, dass sich die schwer unter ihrer Festnetznummer erreichbaren Personen systematisch von den leichter zu Erreichenden unterscheiden (Häder 1996). Es muss daher ein Mittelweg gefunden werden zwischen Effizienz der Anwahlversuche und dem Anspruch, möglichst für jede Nummer einen definitiven Status bestimmen zu können. Aus diesem Grund wurde die maximale Anzahl an Kontaktversuchen relativ hoch auf 15 festgelegt. In einzelnen Fällen wurden diese 15 Anwahlversuche aber zum Teil überschritten, um beispielsweise unterbrochene Interviews zeitnah beenden zu können. Die Verwaltung der Rufnummern mit dem jeweils von den Interviewern zu protokollierenden Anwahlergebnis (Disposition Code), in dem gegebenenfalls auch Kommentare zur Zielpersonermittlung enthalten sind, übernimmt dabei das in der verwendeten Software VOXCO 5.4.4.5 enthaltene Anruf- und Rückrufmanagement, das den jeweiligen Anforderungen entsprechend konzipiert und programmiert werden muss.

Anruf- und Rückrufmanagement

Bei der Konzeption des Anruf- und Rückrufmanagements sind sowohl die Erfordernisse der Interviewer hinsichtlich einer schnellen und einfachen Zuordnung der Ergebnisse als auch die Anforderungen an eine hohe Ausschöpfung der Stichprobe zu berücksichtigen³ (vgl. Abschnitt zur Responseberechnung). Über die Kodierung des Anwahlergebnisses (Disposition Codes) erfolgt auch die Steuerung der Wiedervorlage der Rufnummern nach vorher festgelegten Regeln (Call-Back-Rules). Weiterhin wird über die Vergabe der Disposition Codes auch die Ergebnisdarstellung der Ausschöpfung und damit auch die Berechnung der Ausschöpfungsquote festgelegt. Von zentra-

³ So wird beispielsweise differenziert, ob eine Verweigerung durch die – zufällig – erreichte Kontaktperson erfolgte, bevor die Zielperson im Haushalt ermittelt werden konnte oder ob die Zielperson verweigerte. Die Ergebnisse werden jeweils entsprechend codiert.

ler Bedeutung hierbei ist die Unterscheidung, ob sich Anrufergebnisse den Kategorien »neutraler« bzw. »nicht neutraler« Ausfälle zuordnen lassen. Bei der Konzeption des Anrufmanagements, das über einen Zeitraum von 10 Monaten die Kontrolle über 250.000 Rufnummern ermöglichen soll, musste des Weiteren berücksichtigt werden, dass Nummern zu unterschiedlichen Wochentagen und Uhrzeiten zur Wiedervorlage erscheinen⁴. Über die Call-Back-Verwaltung wird schließlich auch 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 Motivation der Interviewer/-innen zu erhalten, ist es daher in Maßen notwendig, »frische« Nummern zuzulassen. Deren Anteil sollte jedoch möglichst gering gehalten werden, da sie gegen Ende der Feldphase nicht mehr in gleichem Maße systematisch ausgeschöpft werden können, wie früher eingespielte Telefonnummern.

Anders als in den USA liegen im deutschsprachigen Raum bisher keine einheitlichen Standards für die Entwicklung von Disposition Codes vor⁵. Diese wurden daher auf Basis der Erfahrungen mit den vorherigen telefonischen Gesundheitssurveys am Robert Koch-Institut für GEDA entwickelt. Dabei wurden so weit wie möglich die Richtlinien der American Association of Public Opinion Research (AAPOR 2011) angewandt.

Feldverlauf

Die Datenerhebung erfolgte im CATI-Labor des Robert Koch-Institutes, welches über insgesamt 40 Interviewer- und drei Supervisorarbeitsplätze verfügt. Die Feldzeit der Studie GEDA 2010 begann am 22. September 2009 und endete am 10. Juli 2010. In diesem Zeitraum wurden insgesamt 22.076 Personen repräsentativ für die

⁴ Wenn beispielsweise bereits mehrere Kontaktversuche bei einer Telefonnummer durchgehend das Ergebnis »Freizeichen« erzielt haben, ist trotzdem nicht eindeutig zu sagen, ob der Anschluss vergeben ist und wenn ja, ob es sich nicht um einen Firmenschluss handelt. Nur durch den mehrmaligen Wiederanruf zu verschiedenen Uhrzeiten und Wochentagen kann das Risiko minimiert werden, eine Zielperson nicht zu erreichen.

⁵ Für die USA gibt es Richtlinien der American Association of Public Opinion Research (AAPOR 2011). Die in etwa vergleichbaren Einrichtungen der ADM-Institute in Deutschland oder GESIS in Mannheim haben keine derartigen Empfehlungen publiziert. Die umfangreiche Literatur über Telefonsurveys geht auf diesen Aspekt in der Regel nicht ein.

deutschsprachige Wohnbevölkerung ab 18 Jahren befragt, von denen letztlich 22.050 Interviews verwertbar waren⁶.

Für die Realisierung der Interviews waren 138 Interviewerinnen und Interviewer an 223 Arbeitstagen in 353 Schichten (eine Schicht dauert ca. vier Stunden) im Einsatz. Um Intervieweffekte zu minimieren, wurde bei der Zusammensetzung des Interviewerstamms auf ein möglichst ausgewogenes Geschlechterverhältnis, sowie eine gewisse Heterogenität bezüglich des Alters geachtet.

Telefoniert wurde wochentags zwischen 10 und 14 Uhr⁷ mit Ausnahme des Donnerstags sowie von Montag bis Freitag zwischen 16:30 und 20:30 Uhr⁸. Die Samstagsschicht fand zwischen 14 und 18 Uhr statt. Jede Schicht wurde von einem wissenschaftlichen Supervisor bzw. Supervisorin, unterstützt von mindestens einer studentischen Hilfskraft mit Supervisionsaufgaben, geleitet.

Im Durchschnitt dauerte ein Interview abzüglich der Kontaktabstimmungsphase und Überzeugungsarbeit vor Interviewbeginn 31 Minuten. Vor Beendigung des Interviews wurde die Bereitschaft, für weitere Befragungen zu gesundheitsrelevanten Themenbereichen kontaktiert zu werden, erfasst.

Verwendete Software

Die telefonischen Befragungsdaten werden im RKI mit Hilfe der Software »VOXCO Interviewersuite 5.4.4.5« erhoben. Die Auswahl der aktuell zu kontaktierenden Telefonnummern, der Anwahlvorgang und die Wiederholung der Kontaktversuche werden vollautomatisch durch die Software gesteuert. Die Interviewer haben keinen Einfluss auf das System. Diese Software zeichnet sich durch eine sehr gute grafische Darstellung für die Interviewer sowie ein sehr komplexes Anrufmanagement aus.

Qualitätssicherung im Feldverlauf

Ein zentraler Aspekt bei der Durchführung wissenschaftlicher telefonischer Befragungen ist das Einhalten einer standardisierten Messsituation. Um diesem Anspruch gerecht zu werden, ist eine kontinuierliche Qualitätssicherung (QS) unverzichtbar. Die Qualität der Datenerhebung wird fortlaufend anhand definierter Kriterien kontrolliert, um daraus konkrete Maßnahmen für die Qualitätssicherung abzuleiten.

Tabelle 7.1
Eckdaten aus dem Feld

Interviewer/-innen	
Anzahl	138 (82 Frauen und 56 Männer)
Durchschnittsalter	41 Jahre (zwischen 20 bis 70 Jahren)
Probanden	
Anzahl	22.050 (12.483 Frauen und 9.567 Männer)
Durchschnittsalter	48 Jahre (zwischen 18 bis 99 Jahren)
Quantitative Parameter	
Durchschnittliche Interviewdauer	31 Minuten
Telef. Wiederbefragungsbereitschaft	82,8 %
Anzahl Verweigerungen auf Zielpersonenebene	14.477
Anzahl angerufene Nummern	215.513
Anzahl der Anrufe insgesamt	908.883
Ø benötigter Anrufe pro Interview	41,2

⁶ 26 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 Alters-, Sprach- oder Bildungsbarrieren).

⁷ Ab dem 03. Mai 2010 wurden die Frühschichten auf zwei Schichten reduziert, die jeweils montags und mittwochs stattfanden.

⁸ Ausweitung der Anrufzeit auf 16:00 Uhr bis 20:30 Uhr zwischen dem 15.03. und 11.06.2010

Hierbei müssen zwei Arten von Vorgehensweisen kombiniert werden:

- ▶ Strategien zur Identifikation von Verbesserungsbedarfen
- ▶ Strategien zur Implementierung von Verbesserungen in den Routinebetrieb.

Dabei erfolgt eine inhaltliche Trennung zwischen quantitativer und qualitativer Qualitätssicherung. Die quantitative QS vergleicht rein zahlenmäßig die von den jeweiligen Interviewerinnen und Interviewern erzielten Resultate wie z. B. die Interviewdauer zum Durchschnitt, wogegen bei der qualitativen Qualitätssicherung das Beobachten der Interviewerinnen und Interviewer hinsichtlich der sachgemäßen Durchführung der Interviews im Vordergrund steht. Insgesamt wird der Einhaltung der qualitativen Kriterien in der Qualitätssicherung ein höherer Stellenwert beigemessen als derjenige der quantitativen Kriterien.

Die Interviewführung wird regelmäßig durch die Supervisorinnen und Supervisoren unter Verwendung einer Checkliste bewertet. Diese ist angelehnt an diejenige des Behavioral Risk Factor Surveillance System (CDC 2010). Sie wurde für die Bedingungen der Telefonischen Gesundheitssurveys des Robert Koch-Instituts adaptiert und eingesetzt.

Hinsichtlich der Einhaltung der Qualität der Interviews während der Durchführung werden folgende Kriterien angewandt:

- ▶ Korrekter technischer Umgang mit dem Fragebogen⁹
- ▶ Vollständige, aussagekräftige und datenschutzgerechte Dokumentation der Zielperson
- ▶ Vollständige Dokumentation bei mangelnder Qualität eines Interviews
- ▶ Korrekte Zielpersonenermittlung
- ▶ Standardisierte Interviewführung
- ▶ Angemessenes Tempo beim Vorlesen der Fragen
- ▶ Angemessenes »Auftreten« gegenüber den Befragten
- ▶ Vollständiges Vorlesen der Fragen und der Antwortkategorien
- ▶ Kodierverhalten im Rahmen der Disposition-Codes

Datenschutz

Alle Befragungen des Robert Koch-Instituts unterliegen der strikten Einhaltung der datenschutzrechtlichen Bestimmungen des Bundesdatenschutzgesetzes (BDSG). Das Konzept von GEDA 2010 wurde dem Bundesbeauftragten für den Datenschutz und die Informationsfreiheit zur Prüfung vorgelegt. Es wurden keine Bedenken gegen die Durchführung der Studie geäußert.

Gemäß § 5 des BDSG und der Datenschutzordnung ist es untersagt, personenbezogene Daten oder andere sachbezogene Daten und Informationen unbefugt zu nutzen oder zu verarbeiten oder an Dritte weiterzugeben. Damit wird der Einzelne davor geschützt, durch den Umgang mit den personenbezogenen Daten in seinem Persönlichkeitsrecht verletzt zu werden.

Die Verwaltung personenbezogener Angaben aus GEDA-Daten erfolgt getrennt von den Angaben der Befragung zentral in einer passwortgeschützten Datenbank, auf die nur besonders dazu berechtigte Mitarbeiterinnen und Mitarbeiter Zugriff haben. Die Daten werden vollständig anonym ausgewertet.

Voraussetzung für die Erhebung der Daten ist die datenschutzrechtliche Information der Befragten vor Beginn des Interviews. Die teilnahmebereiten Personen werden darüber aufgeklärt, dass die Teilnahme an der Befragung freiwillig ist, diese jederzeit abgebrochen werden kann und einzelne Fragen nicht beantwortet werden müssen. Des Weiteren wird darüber informiert, dass bei Nichtteilnahme an der Studie keine Nachteile entstehen und die Auswertung anonym erfolgt, also persönliche Angaben (Telefonnummer) und Befragungsergebnisse getrennt aufbewahrt werden. Bei Personen, die sich im Rahmen von GEDA dazu bereit erklären, zu einem späteren Zeitpunkt erneut für Befragungen des RKI kontaktiert zu werden, wird zusätzlich das Einverständnis eingeholt, Namen und gegebenenfalls die Adresse für eine erneute Kontaktaufnahme zu speichern. Diese Personen werden auch informiert, dass sie diese Einwilligung jederzeit widerrufen können.

Die im Feld tätigen Supervisorinnen und Supervisoren sowie die Interviewerinnen und Interviewer werden über die datenschutzrechtlichen Auflagen und den Umgang mit personenbezogenen Daten belehrt und unterzeichnen eine Datenschutzerklärung.

⁹ Hiermit ist z. B. der korrekte Umgang mit der Funktion zur Unterbrechung und Wiederaufnahme des Interviews gemeint.

Response

Das Gesundheitsmonitoring am Robert Koch-Institut ist bemüht, die Erhebungsinstrumente und Methoden der Studien soweit wie möglich an international etablierten Kriterien zu orientieren (Kurth et al. 2009). Die Berechnung und Darstellung der Response orientiert sich auch aus diesem Grund an den Vorgaben der American Association for Public Opinion Research (AAPOR 2011) und ermöglicht eine sehr differenzierte Beurteilung der Datenqualität. Ausgangsbasis für die Responseberechnung ist ein Datensatz, der alle in der Feldzeit durchgeführten Telefonanrufe enthält (ein sog. »callbasierter Datensatz«). Jedes Anrufergebnis (wie »Interview durchgeführt«, »Freizeichen«, »Verweigerung durch Zielperson« etc.) wird dabei anhand eines Disposition-Codes erfasst. Zur Berechnung der Ausschöpfung des Nummernsamples muss jede Telefonnummer der Stichprobe einer der folgenden Antwortgruppen nach Klassifikation der AAPOR zugeordnet werden.

Ein wesentliches Kriterium der Zuordnung ist, ob die generierten Anwahlnummern jeweils ein Element der Grundgesamtheit sind oder nicht. Die Grundgesamtheit für GEDA 2010 ist die volljährige, deutschsprachige Wohnbevölkerung in Privathaushalten der Bundesrepublik, sofern sie über Festnetzanschlüsse erreichbar ist.

Tabelle 7.2
Arten von endgültigen Anrufergebnissen in telefonischen Surveys nach AAPOR

Hauptgruppe/Untergruppe	Code
Interview	
Vollständiges Interview	Interview (I)
Unvollständiges Interview	Partial Interview (P)
Kein Interview	
Verweigerung, Abbruch, Offene Terminvereinbarungen	Refusal (R)
Kein Kontakt	Non Contact (NC)
Andere Ursachen	Other (O)
Unbekannt ob in Grundgesamtheit	
Unbekannt ob Haushalt	Unknown Household (UH)
Unbekannt ob Haushalt in Grundgesamtheit	Unknown Other (UO)
Nicht in Grundgesamtheit	
Nicht in Grundgesamtheit	Not Eligible (E)

Insgesamt werden vier Haupt- und acht Untergruppen von Anrufergebnissen unterschieden, die in Tabelle 7.2 mit ihrem jeweiligen AAPOR Code dargestellt sind. Den Interviews (I) oder Teilinterviews (P) werden drei Gruppen gegenübergestellt: Nummern, die nicht in die Grundgesamtheit gehören, also qualitätsneutrale Ausfälle (E), Personenkontakte, die nicht zu Interviews führen, hierzu gehören Verweigerungen von Zielpersonen oder Kontaktpersonen im Haushalt oder offene Terminierungen zu Feldende (R, NC, O) sowie Nummern, in denen nicht bekannt ist, ob sie zur Grundgesamtheit gehören oder nicht (UH, UO).

Unvollständige Interviews (P) werden in Anlehnung der Standards in AAPOR (AAPOR 2011: 13ff.) definiert als abgebrochene Interviews, in denen 50 % bis 80 % der inhaltlichen Fragekomplexe sowie alle relevanten soziodemografischen Fragen beantwortet wurden. Der Kurzfragebogen der telefonischen Gesundheitssurveys am RKI enthält dagegen lediglich 11 Fragen und damit deutlich weniger Fragen als nötig, um dieser Definition zu entsprechen. Die demografischen Informationen des Hauptfragebogens befinden sich zudem am Ende des Fragebogens, so dass diese Kategorie »Partial Interview« in GEDA keine Entsprechung hat. Die erfassten Kurzfragebögen werden für die Berechnungen der Response nach dem Schema der AAPOR den Verweigerungen auf Zielpersonenebene zugeordnet (vgl. auch Tabelle 7.3).

Für die Zuordnung der Nummern aus dem Nummernsample zu den genannten Gruppen ist eine Unterscheidung zwischen temporären und endgültigen Anrufergebnissen notwendig. Temporäre Ergebnisse (z. B. Freizeichen oder Terminvereinbarungen) ziehen in der Regel weitere Anrufe nach sich. Endgültige Ergebnisse (z. B. Interview oder Geschäftsanschluss) werden dagegen nicht mehr kontaktiert und gelten als abschließend bearbeitet. Treten endgültige Anrufergebnisse auf, können diese bei der Responseberechnung direkt zugeordnet werden. Es gibt allerdings eine Reihe von Telefonnummern, für die zum Feldende lediglich temporäre Ergebnisse vorliegen; für sie lässt sich der endgültige Status nicht direkt klären. Gründe für temporäre Anrufergebnisse liegen etwa in der begrenzten Feldzeit oder in einer Begrenzung der Anrufe pro Telefonnummer. In der GEDA-Studie 2010 war die Feldzeit auf 10 Monate und die Anzahl der Kontakte

in der Regel auf 15 pro Nummer begrenzt.¹⁰ Die Empfehlungen der AAPOR (AAPOR 2011) sehen vor, bei Nummern ohne endgültiges Anrufergebnis eine Betrachtung der nummernbezogenen Anrufsequenzen durchzuführen, um ihren endgültigen Status zu bestimmen. Das Vorgehen bei der Zuweisung der endgültigen Anrufresultate für die Responseberechnung in der GEDA-Studie 2010 orientierte sich an diesen allgemeinen Vorgaben der AAPOR.

Eine zentrale Vorgabe bei der Bearbeitung von Anrufergebnissen ist, dass Kontakte mit Personen höher als technische Informationen gewertet werden müssen. Liegen mehrere widersprüchliche temporäre Anrufergebnisse vor, soll das Ergebnis mit dem höchsten Informationsgehalt und der höheren Wertigkeit für die Responseberechnung zugeordnet werden. So kann es beispielsweise sein, dass bei einem Anwahlversuch ein Kontakt mit der zu befragenden Person hergestellt und ein Termin vereinbart wird (R). Im weiteren Verlauf aber lauten die Resultate »10: Freizeichen« oder »14: Fax« (UH bzw. E). Die höhere Wertigkeit des Ergebnisses R im Vergleich zu den Ergebnissen UH und E würde zu einer Zuordnung der Nummer zu den Verweigerungen/offenen Terminvereinbarungen zum Feldende führen, es wird also nicht das letzte Anrufergebnis zur Responseberechnung herangezogen.

Insgesamt konnten in der GEDA-Studie 2010 die in Tabelle 7.3 dargestellten endgültigen und temporären Anrufergebnisse auftreten. Bezüglich der Verweigerungen bzw. offenen Terminvereinbarungen (R) wird zwischen Zielpersonen (R_{ZP}) und Kontaktpersonen (R_{KP}) unterschieden. Sind alle Telefonnummern einem endgültigen Resultat zugeordnet worden, so werden diese in eine von AAPOR für diesen Zweck erstellte Standardtabelle eingegeben.¹¹ Die Tabelle berechnet unterschiedliche Ausschöpfungs- und Responsequoten (Response Rate, Cooperation Rate, Refusal Rate und Contact Rate).

¹⁰ Die Begrenzung der Anzahl der Anrufe resultiert aus den Erfahrungen mit früheren Befragungen (ein Großteil der endgültigen Anrufergebnisse wird bereits mit dem fünften Anruferziel) und soll einen angemessenen Mittelweg zwischen nicht zu hoher Belästigung der potenziellen Teilnehmerinnen und Teilnehmer und dem Ziel einer möglichst hohen Ausschöpfung des Nummernsamples darstellen. So zeigen Ergebnisse aus anderen Studien, dass die optimale Anzahl von Anrufen wahrscheinlich größer als 5 aber kleiner als 20 Anrufe pro Telefonnummer ist (Peytchev et al. 2009). Lagen Terminabsprachen mit den Zielpersonen vor, wurde in Einzelfällen auf Entscheidung der Feldleitung aber auch häufiger bei einer Nummer angerufen.

¹¹ Die aktuelle Tabelle kann über die Internetadresse: http://www.aapor.org/Resources_for_Researchers.htm (Version 3.1, November 2010) bezogen werden.

Tabelle 7.3
Zuordnung der Disposition Codes aus GEDA 2010 zu den Kategorien von AAPOR

Resultat	Code	Zuordnung
Hauptfragebogen komplett	97	I
Kurzfragebogen beendet	98	R _{ZP}
Endgültiger Abbruch	99	R _{ZP}
Abbruch durch Interviewer	91	R _{ZP}
Unterbrechung während des Interviews	8	R _{ZP}
ZP hat gerade keine Zeit	72	R _{ZP}
KP verweigert stellvertretend für ZP	66	R _{KP}
KP verweigert stellvertretend für den Haushalt – ZP nicht bekannt	57	R _{KP}
KP lässt nicht ausreden/wimmelt ab	51	R _{KP}
Auskunft KP: ZP nicht da/ hat gerade keine Zeit	62	R _{KP}
Auskunft KP: ZP während Feldzeit nicht erreichbar	65	R _{KP}
Anrufbeantworter privat (inkl. Mailbox)	13	NC
ZP ist verstorben	69	O
ZP ist nicht in der Lage ein Interview durchzuführen (Krankheit, Behinderung etc.)	75	O
Auskunft KP: ZP nicht in der Lage (Krankheit, Behinderung etc.)	64	O
KP ist nicht auskunftsfähig (z. B. Kind...)	55	UO
KP hat gerade keine Zeit für ZP-Ermittlung	56	UO
Technische Verständigungsprobleme mit KP	53	UH
Besetzt	11	UH
Freizeichen	10	UH
Auskunft KP: ZP verzogen, neue Tel.-Nr. ist bekannt	67	E
Auskunft KP: ZP verzogen, neue Tel.-Nr. ist nicht bekannt	68	E
KP versteht nicht ausreichend deutsch	52	E
Auskunft KP: ZP versteht nicht ausreichend deutsch	63	E
ZP versteht nicht ausreichend deutsch	73	E
FAX/Modem	14	E
Leitung tot - nichts tut sich	30	E
Display-Anzeigen	32	E
Nummer nicht vergeben/Kein Anschluss	41	E
Sonstige Ansage – bitte eintragen!	44	E
Anschluss ist vorübergehend nicht erreichbar	43	E
Teilnehmer hat neue Nummer (Ansage)	42	E
KP auf dem Handy erreicht	54	E
Geschäftsanschluss	24	E
Geschäftsanschluss (Warteschleife; AB)	81	E
Kein Privathaushalt (z. B. Krankenhaus, Behörde...)	23	E
Keine ZP im Haushalt (z. B. unter Altersgrenze)	22	E
Keine ZP im Haushalt: sonstige Gründe	25	E

KP = Kontaktperson, ZP = Zielperson

Ergebnisse

In der GEDA-Studie 2010 gab es insgesamt 22.050 erfolgreich durchgeführte, vollständige Interviews. Die Anteile der Anrufergebnisse sind in Tabelle 7.4 dargestellt. Qualitätsneutrale Ausfälle (Nummern, die kein Element der Grundgesamtheit sind) machen mehr als die Hälfte des Nummernsamples aus.

Tabelle 7.4
GEDA 2010 – Verteilung der Rufnummern auf die Kategorien von AAPOR

Antwortgruppe	Code	Anzahl	%
Telefonnummernsample	–	215.449	(100,0%)
Vollständiges Interview	I	22.050	(10,2%)
Unvollständiges Interview	P	0	
Verweigerung/Abbruch/ offene Termine insgesamt	R	40.527	(18,8%)
Verweigerung/Abbruch/ offene Termine durch ZP	R _{ZP}	17.442	(8,1%)
Kein Kontakt	NC	6.458	(3,0%)
Andere Ursachen	O	1.440	(0,7%)
Unbekannt ob Haushalt in Grundgesamtheit	UO	354	(0,2%)
Unbekannt ob in Grund- gesamtheit	UH	16.430	(7,6%)
Nicht in Grundgesamtheit	E	128.190	(59,5%)

In Tabelle 7.5 sind die nach AAPOR berechneten Ausschöpfungsquoten dargestellt.

Die Responserate stellt den Anteil der realisierten Interviews an allen Telefonnummern der Grundgesamtheit dar, bereinigt um qualitätsneutrale Ausfälle.

Die Kooperationsraten stellen den Anteil der realisierten Interviews an allen kontaktierten Personen der Grundgesamtheit dar. Hierbei bleiben zusätzlich zu den qualitätsneutralen Ausfällen auch diejenigen Anwahlnummern unberücksichtigt, die beispielsweise zu Frei- oder Besetzzeichen führen.

Da für GEDA 2010 keine gesonderten Outcome Rates für Teilinterviews (partial interviews) berechnet werden, bleiben diese Raten in der entsprechenden Tabelle unberücksichtigt.

Für GEDA 2010 liegt die nach AAPOR berechnete Response Rate 3 bei 28,9 %. Response Rate 3 gibt den Anteil der realisierten Interviews im Verhältnis zu allen wahrscheinlichen Haushalten der Grundgesamtheit wieder. Für diejenigen Telefonnummern, bei denen zu Feldende unklar bleibt, ob sie Elemente der Grundgesamtheit sind oder nicht (UH+UO), wird dabei eine Schätzung vorgenommen, wie viele von diesen der Grundgesamt-

Tabelle 7.5
Outcome Rates nach AAPOR in der GEDA-Studie 2010

Rate/ Variante	Berechnung	2010
Eligibility Rate		
e	$(I+P+R+NC+O) / (I+P+R+NC+O+E)$	35,5%
Response Rate		
1	$I / (I+P+R+NC+O+UH+UO)$	25,3%
3	$I / (I+P+R+NC+O+e(UH+UO))$	28,9%
Cooperation Rate		
1	$I / (I+P+R+O)$	34,4%
3	$I / (I+P+R)$	35,2%
Refusal Rate		
1	$R / (I+P+R+NC+O+UH+UO)$	46,4%
2	$R / (I+P+R+NC+O+e(UH+UO))$	53,0%
3	$R / (I+P+R+NC+O)$	57,5%
Contact Rate		
1	$(I+P+R+O) / (I+P+R+O+NC+UH+UO)$	73,4%
2	$(I+P+R+O) / (I+P+R+O+NC+e(UH+UO))$	83,8%
3	$(I+P+R+O) / (I+P+R+O+NC)$	90,8%
Cooperation Rate at respondent level		
RKI	$I / (I+P+R_{ZP})$	55,8%

heit angehören (AAPOR 2011, S. 45). Diese Schätzung wird den erhobenen Daten entnommen, und zwar als Anteil der Haushalte der Grundgesamtheit an allen Nummern, für die der Status, ob sie der Grundgesamtheit angehören oder nicht, festgestellt werden konnte (eligibility rate).

Die Cooperation Rate 3 als Anteil der realisierten Interviews an allen erreichten Haushalten der Grundgesamtheit liegt bei 35,2 %. Bezieht man den Anteil der Interviews auf die Zielpersonenkontakte, so konnte bei 55,8 % der zu befragenden Personen ein Interview durchgeführt werden (siehe Tabelle 7.5). Diese Rate lag für GEDA 2009 bei 51,2 % und hat sich damit leicht verbessert. Trotz verbesserter Kooperationsrate und Verweigerungsrate (Refusal Rate 1 sinkt von 48,8 % auf 46,4 %) gegenüber GEDA 2009, stagniert die Responserate insgesamt, was auf eine schlechtere Erreichbarkeit zurückzuführen ist (Contact rate 1 sinkt von 75,4 % auf 73,4 %). Für eine schlechtere Erreichbarkeit spricht auch die gestiegene durchschnittliche Anzahl von Anrufen, die pro Interview erbracht werden müssen. Diese ist von ca. 34 (GEDA 2009) auf 41 (GEDA 2010) gestiegen.

Schlussfolgerungen

Für die GEDA-Studie 2010 hat sich die Ausschöpfung auf Zielpersonenebene im Vergleich zu GEDA 2009 leicht verbessert, im Vergleich zu den telefonischen Gesundheitssurveys 2005 (48,5 %) und 2006 (44,9 %) konnte sie dagegen deutlicher gesteigert werden. Die erzielte Ausschöpfung bewegt sich im Bereich dessen, was in den letzten Jahren für andere telefonische Umfragen in Deutschland oder den USA berichtet wurde: So wurden im Rahmen einer methodischen Studie 132 sozialwissenschaftliche Studien mit insgesamt über 300.000 Interviews, die im Zeitraum 1990 bis 2007 in Deutschland durch das infas-Institut erhoben wurden, im Hinblick auf das Antwortverhalten untersucht (Aust, Schröder 2009). Zu den einbezogenen Untersuchungen gehörten u. a. die deutsche Stichprobe der SHARE-Studie (Survey of Health, Ageing and Retirement in Europe), der Alterssurvey und der Deutsche Jugendsurvey. Die berichtete Nettoausschöpfung (vergleichbar mit der Response Rate 2 laut AAPOR 2011) lag für diese Studien im Bereich von 25 % bis 40 % (telefonische Studien ohne vorherige schriftliche Kontaktaufnahme).

Im Zeitverlauf wurde zudem eine Verringerung der Kooperationsraten beobachtet. Ergebnisse, die auf einen kontinuierlichen Rückgang der allgemeinen Antwortrate deuten, liegen außerdem auch aus den USA für den Zeitraum 1979 bis 2003 vor (Curtin et al. 2005). Demnach ist die Antwortrate (Response Rate 2) seit 1979 jährlich um durchschnittlich 0,95 Prozentpunkte zurückgegangen. Holbrook, Krosnick und Pfent fassen 114 telefonische Studien zusammen, die von führenden amerikanischen Umfrageinstituten zwischen 1996 und 2005 durchgeführt worden sind. Die durchschnittliche Responserate (Response Rate 3 – hier also ohne Teilinterviews) wird mit 30 % angegeben. Im beobachteten Zeitraum ergibt sich durch regressionsanalytische Berechnungen ein Rückgang in der Responserate von 0,2 % pro Monat. Im Zeitraum zwischen 1996 bis 2005 konnte insgesamt ein dramatischer Rückgang um 16,4 % festgestellt werden (Holbrook et al. 2008).

Die Ausschöpfung der Stichprobe in der GEDA-Studie 2010 liegt daher angesichts des insgesamt schwierigeren Umfeldes für telefonische Surveys im erwarteten und aus der Literatur bekannten Rahmen. Angesichts des allgemeinen Trends der sinkenden Antwortbereitschaft in Telefonumfragen werden für die GEDA-Studie aber alternative Stichprobenzugänge sowie die Realisierbarkeit von Mixed-Mode-Datenerhebungen untersucht.

Das Verfahren zur Responseberechnung in der GEDA-Studie 2010 orientiert sich an international etablierten Standards. Die ausgewiesenen Schätzwerte sind dabei vergleichsweise konservativ, da auch jene Anrufe, die durch ein schnelles Auflegen der Kontaktperson noch während der Vorstellung der jeweiligen Interviewer vorzeitig beendet wurden als Verweigerung auf Haushaltsebene zugeordnet wurden (vgl. Tabelle 7.3, Code 51).

Insgesamt kann angesichts der mit anderen Studien vergleichbaren Response Rate 3 und der aufwändigen Poststratifizierung (Gewichtung) von einer guten Datenqualität in der GEDA-Studie 2010 ausgegangen werden.

Gewichtung

Telefonische Bevölkerungsumfragen haben allgemein mit sinkenden Teilnehmerquoten (vgl. Abschnitt zur Response) zu kämpfen. Nichtteilnahme führt häufig zu systematischen Verzerrungen in der Stichprobe, so dass die erhobenen Daten nicht mehr repräsentativ für die Wohnbevölkerung sind. Zusätzlich kann die steigende Verbreitung der Mobiltelefonie und den damit einhergehenden sinkenden Ausstattungsgrad privater Haushalte mit Festnetztelefonen zu nicht übereinstimmenden Grund- und Auswahlgesamtheiten führen. Sowohl Nonresponse als auch Noncoverage können demnach die Unter- oder Überrepräsentation einzelner Bevölkerungsgruppen in der Nettostichprobe bewirken (Gabler, Häder 1997).

Um die oben genannten Verzerrungen auszugleichen und für die erwachsene Bevölkerung in Deutschland repräsentative Aussagen treffen zu können, wurde im Rahmen von GEDA ein Gewichtungsverfahren implementiert, das sich unter anderem an ein im Behavioral Risk Factor Surveillance System (BRFSS) angewandten Verfahren orientiert (CDC 2010).

Ziel der Gewichtung ist es, die relative Wichtigkeit der einzelnen Befragten zu ändern, um hierdurch das Nettostichprobenprofil hinsichtlich bestimmter aus der Grundgesamtheit ausgewählter bekannter Variablen anzupassen.

Designgewichtung

Für die Berechnung der Stichprobengewichte wird auf der ersten Stufe das Stichprobendesign berücksichtigt. Die Designgewichte sind umgekehrt proportional zur Auswahlwahrscheinlichkeit der Befragten. Bei GEDA 2010 handelt es sich

um ein zweistufiges Auswahlverfahren, bei dem in der ersten Stufe Haushalte über Telefonnummern zufällig nach dem Gabler-Häder-Design (Gabler, Häder 1999) ausgewählt werden. In der zweiten Stufe wird dann in jedem ausgewählten Haushalt eine Zielperson unter den über achtzehnjährigen Hausmitgliedern nach der Last-Birthday-Methode ausgewählt und diese um ein Interview gebeten. Entsprechend wird die Auswahlwahrscheinlichkeit sowohl durch die Anzahl der Telefonnummern je Haushalt als auch durch die Anzahl der volljährigen Personen im Haushalt beeinflusst. Somit ergibt sich nach einer Normierung auf die Anzahl n der interviewten Personen folgendes Designgewicht für die Daten einer Person im Haushalt k :

$$d_k = \frac{n}{\sum_{i=1}^n \frac{m_i}{t_i}} \cdot \frac{m_k}{t_k}$$

mit

m_k : Anzahl der Personen im Haushalt k

t_k : Anzahl der Telefonnummern im Haushalt k

Bei diesem Design-Gewicht d_k können durch eine hohe Zahl von Telefonnummern pro Haushalt bzw. durch viele Personen pro Haushalt sehr große bzw. sehr kleine Gewichte auftreten, die möglicherweise unrealistisch sind. Deshalb ist hier eine Dämpfung dieser extremen Gewichte vorgenommen worden, indem die Anzahl der Telefonnummern t_k als auch die Anzahl der Personen m_k pro Haushalt bei der Bildung der Designgewichte ab einer gewissen Höhe gekappt worden sind. Hierdurch bleiben die Gewichte von Zielpersonen in Haushalten, die diese Kappungsgrenzen nicht überschreiten, unberührt. Bei GEDA 2010 sind für die Telefonnummern pro Haushalt zwei Kategorien gebildet worden (a: eine Telefonnummer pro Haushalt und b: mehr als eine Telefonnummer pro Haushalt). Aus der Anzahl der Personen über 18 Jahre pro Haushalt wurden drei Kategorien gebildet (a: eine Person pro Haushalt, b: zwei Personen pro Haushalt und c: mehr als 2 Personen pro Haushalt). Das gekappte Designgewicht hat eine Spannweite von 0,33 bis 2,00 (siehe Tabelle 7.6).

Tabelle 7.6
Statistische Kennwerte der normierten Design-Gewichte bei GEDA 2010 (n = 22.050)

	Min.	1. Qu.	Median	Mittelwert	3. Qu.	Max.	Effizienz	≤ 0,3	≥ 3
Designgewicht GEDA 2010	0,33	0,65	0,98	1,00	1,31	2,00	84 %	0	0

Die Güte einer Gewichtung kann durch mehrere Maße bestimmt werden. Kriterien hierfür sind Verteilungsparameter wie Spannweite, Streuung und Quantile der Gewichte. Ein wichtiges Gütekriterium ist das Effizienzmaß

$$E = \frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n d_i^2}$$

d_i : Gewicht des Probanden i

das sich seinem Maximum 1 annähert, wenn die Gewichte alle nahezu den Wert 1 haben. Eine hohe Effizienz besagt also eine geringe Streuung der Gewichte, die dann auch relativ kleine Stichprobenfehler bedingen. Allerdings kann der Nonresponse-Bias größer sein als der Stichprobenfehler, so dass das Effizienzmaß nicht überbewertet werden sollte.

Anpassungsgewichtung

Die Anpassung der Gewichte an die Bevölkerungszahlen erfolgte durch Poststratifizierung nach Bundesländern, Geschlecht, Altersklassen (Datengrundlage: Bevölkerungsfortschreibung des Statistischen Bundesamtes zum Stand 31.12.2008) und drei Bildungsstufen nach der internationalen Klassifikation ISCED (International Standard Classification of Education, Schroedter et al. 2006), die die schulische und berufliche Bildung einbezieht (Datengrundlage: Mikrozensus 2008). Bei der Anpassungsgewichtung über Poststratifizierung wird zunächst die Population in H nicht überlappende Straten aufgeteilt, für die die Bevölkerungszahlen bekannt sind. In der Stichprobe werden dann in jedem Stratum die Gewichte so verändert, dass die geschätzte Anzahl mit der externen Angabe übereinstimmt.

Bei der gleichzeitigen Poststratifizierung nach den genannten Merkmalen wäre eine hohe Anzahl von Straten entstanden (16 Bundesländer, 2 Geschlechter, 6 Altersklassen und 3 Bildungsstufen: $16 \times 2 \times 6 \times 3 = 576$), die in der Stichprobe nicht alle belegt waren. Aus diesem Grund ist ein dreistufiges Vorgehen gewählt worden, mit dem das Problem der nicht besetzten Straten vermie-

Tabelle 7.7
Kennwerte der Gewichte bei GEDA 2010 (n = 22.050)

	Min.	Median	Mittelwert	Max.	≤ 0,3	≥ 3
Gewichte	0,27	0,74	1,00	5,34	10,3 %	4,0 %

den werden konnte. In der ersten Stufe wurde die Stichprobe nach 7 Altersklassen und Geschlecht poststratifiziert, d. h. die Designgewichte wurden an die Besetzung dieser Straten in der Population angepasst. In der zweiten Stufe sind die modifizierten Designgewichte weiter an die drei Bildungskategorien innerhalb von 6 Altersklassen und Geschlecht adjustiert worden. Die Wahl der unterschiedlichen Altersstrukturierung bei den Anpassungsstufen ermöglichte eine bessere Anpassung an die Altersverteilung in der Bevölkerung als das bei einheitlichen Altersklassen der Fall gewesen wäre. In der dritten und letzten Stufe wurden die Gewichte nochmals an die Bevölkerungszahlen der Bundesländer, 6 Altersklassen und Geschlecht angepasst, um die Konsistenz der Schätzungen innerhalb dieser Teilpopulationen zu gewährleisten. Bei den letzten beiden Anpassungsstufen sind, um extreme Gewichte zu vermeiden, die adjustierten Gewichte bei dem 0,05 Quantil bzw. bei dem 99,5 Quantil abgeschnitten worden, und auf den Wert des entsprechenden Quantils gesetzt worden. Nach jeder Anpassungsstufe wurden die Stichprobengewichte wieder auf die Anzahl der durchgeführten Interviews (n = 22.050) normiert.

Dieses dreistufige Verfahren ist iterativ solange wiederholt worden, bis sich die Gewichte zwischen erster und dritter Anpassungsstufe nur noch wenig (Summe der Differenzen zwischen den Gewichten der ersten und dritten Anpassungsstufe $\leq 1/1.000$ des Stichprobenumfangs) unterschieden haben. Durch das Iterationsverfahren wurde erreicht, dass die Stichprobe nahezu an die Randverteilungen von allen drei Klassifizierungen angepasst werden konnte.

Nach Design- und Anpassungsgewichtung war das minimale Gewicht um dem Faktor 20 kleiner als das größte Gewicht (siehe Tabelle 7.7). Die Aussagen von der Person mit dem maximalen Gewicht tragen zwanzigmal mehr zu einer Schätzung bei, als die von der Person mit dem geringsten Gewicht. Etwa 10 % der Gewichte haben einen Wert kleiner oder gleich 0,3, 4 % der Gewichte liegen im Wertebereich 3 oder darüber.

Datenaufbereitung

Beim Datenmanagement der Erhebungsdaten stand die Standardisierung der Verfahren zur Aufbereitung, Prüfung und Bereinigung der Daten im Vordergrund. Die etablierten Methoden zur Datenaufbereitung und Datenqualitätssicherung wurden weitgehend standardisiert und automatisiert und durch Datenbanktools zur Verwaltung und Dokumentation von Erhebungsinstrumenten und Qualitätssicherungsmaßnahmen ergänzt. Zur Gewährleistung eines einheitlichen Vorgehens bei Datenänderungen wurden allgemeine Regeln zur Datenprüfung und -korrektur vereinbart.

Im Zuge der Datenaufbereitung wurden alle im Verlauf der Felderhebung erzeugten Prozessinformationen wie z. B. Anmerkungen der Interviewer zur Qualität der Interviews, technische Probleme während der Erhebungszeit, Dokumentationen der Supervisoren etc. zusammengeführt und dahingehend geprüft, ob sich daraus Korrekturen für den Datensatz ergeben. Als Ergebnis dieser Prüfung wurden nachträglich 26 Interviews entfernt (vgl. Abschnitt zum Feldverlauf).

Die Datensatzaufbereitung erfolgte im epidemiologischen Datenzentrum des Robert Koch-Instituts. Hier wurden in einem ersten Schritt, den Vorgaben des Datenschutzes entsprechend, alle personenbezogenen Informationen wie Name, Telefonnummer und Adresse (sofern vorhanden) aus den Erhebungsdaten entfernt. Anschließend wurden die Daten auf Vollständigkeit und logische Korrektheit überprüft. Alle datenbezogenen Prozesse, wie z. B. Korrekturen, Transformationen und Freitextkodierungen erfolgen grundsätzlich per Syntax und sind damit dokumentiert, wiederholbar und ggf. auch korrigierbar. Eine unveränderte Kopie der Originaldaten wird aufbewahrt.

Die Identifizierung von Datenqualitätsproblemen erfolgt durch die Analyse von Inhalt und Struktur einzelner Attribute im Zusammenhang mit Datentypen, Wertebereichen, Verteilungen und Varianzen einzelner Attributwerte. Für die Datenqualitätsanalysen werden in der Regel Verfahren der deskriptiven Statistik (Verteilungsanalysen, Ausreißertests etc.) sowie des Data Mining (regelbasierende oder Cluster-Analysen und Entscheidungsbaumverfahren) eingesetzt. Letztlich bestimmt aber der Mensch die Qualität der Analyseprozesse: Nur durch eine enge Zusammenarbeit zwischen IT- und Datenqualitätsexperten

und qualifizierten Datennutzern aus den Fachabteilungen können die Ergebnisse der Datenprüfung richtig interpretiert und anschließend die notwendigen Regeldefinitionen oder Prozessänderungen eingeleitet werden.

Folgende Arbeitsschritte wurden im Rahmen der Datenaufbereitung und Qualitätssicherung durchgeführt:

- ▶ Identifizierung von Ausschlussfällen
- ▶ Gewichtung der Daten
- ▶ Prüfung der Erhebungssitem
- ▶ Generierung neuer aggregierter Variablen
- ▶ Kodierung von Freitexten
- ▶ Prüfung komplexer Filterfragen auf Funktionalität
- ▶ Definition von benutzerdefinierten Missings
- ▶ Prüfung und ggf. Umsetzung von Interviewerbemerkungen (Notes)
- ▶ Dokumentation der Datenqualitätsicherungsmaßnahmen
- ▶ Bereitstellung des Datensatzes als Public Use File

Die im Rahmen der Datenaufbereitung und Prüfung gewonnenen Erkenntnisse werden dokumentiert und fließen in der Konzeptionsphase zukünftiger Studien in den Entwicklungsprozess neuer Fragebögen mit ein, um die Ergebnisqualität der GEDA-Studiendaten kontinuierlich zu verbessern.

Literatur

- American Association for Public Opinion Research (AAPOR) (2011): Standard Definitions – Final Dispositions of Case Codes and Outcome Rates for Surveys. Revised 2011
http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/ContentDisplay.cfm&ContentID=3156
- ADM/BVM (1998) Richtlinie für telefonische Befragungen. Telefonstichproben in Deutschland. In: Gabler S, Häder S, Hoffmeyer-Zlotnik JHP (Hrsg) Westdeutscher Verlag, Opladen, S 181–187
- Aust F, Schröder H (2009) Sinkende Stichprobenausschöpfung in der Umfrageforschung – ein Bericht aus der Praxis. In: Weichbold M, Bacher J, Wolf C (Hrsg) Umfrageforschung – Herausforderung und Grenzen. VS Verlag für Sozialwissenschaften, S 195–212
- Borch S, Rieck A, Blümel M (2004) Operationshandbuch – Erster telefonischer Bundes-Gesundheitssurvey 2003. Bundesgesundheitsbl – Gesundheitsforsch – Gesundheitsschutz. RKI, Berlin
- Centers for Disease Control and Prevention (CDC) (2010) Behavioral Risk Factor Surveillance System 2009. Summary Data Quality Report Atlanta, U.S. Department of Health and Human Services
- Curtin R, Presser S, Singer E (2005) Changes in Telephone Survey Nonresponse over the Past Quarter Century. *Public Opin Q* 69(1): 87–98
- De Leeuw E, de Heer W (2002) Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In: Groves RM, Dillman DA, Eltinge JL et al. (Hrsg) Survey Nonresponse. Wiley, New York, S 41–54
- Gabler S, Häder S (1997) Wirkung von Gewichtungen bei Face-to-Face und Telefonstichproben. Eurobarometer-Experiment 1994. In: Gabler S, Hoffmeyer-Zlotnik JHP (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
- Häder S (1996) Wer sind die »Nonpubs«? Zum Problem anonymer Anschlüsse bei Telefonumfragen. *ZUMA-Nachrichten* 39: 45–68
- Häder S, Gabler S, Heckel C (2009) Stichprobenziehung, Gewichtung und Realisierung. In: Häder M, Häder S (2009) Telefonbefragungen über das Mobilfunknetz. VS Verlag für Sozialwissenschaften, Wiesbaden, S 23ff
- Heckel C (2001) Erstellung der ADM-Telefonauswahlgrundlage. ZUMA-Workshop »Methodische Probleme bei der Stichprobenziehung und Realisierung«. Aschpurwis + Behrens GmbH, Mannheim
- Holbrook AL, Krosnick JA, Pfent A (2008) The Causes and Consequences of Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms: 499–528. In: Lepkowski JM, Tucker C, Brick JM et al. (Hrsg) Advances in Telephone Survey Methodology. John Wiley & Sons, New Jersey
- Kurth BM, Lange C, Kamtsiuris P et al. (2009) Gesundheitsmonitoring am Robert Koch-Institut. *Bundesgesundheitsbl – Gesundheitsforsch – Gesundheitsschutz* 52 (5): 557–570
- Peytchev A, Baxter RK, Carley-Baxter LR (2009) Not All Survey Effort is Equal: Reduction of Nonresponse Bias and Nonresponse Error. *Public Opin Q* 73 (4): 785–806
- Robert Koch-Institut (Hrsg) (2005) Erster telefonischer Gesundheitssurvey des Robert Koch-Instituts – Methodische Beiträge. Beiträge zur Gesundheitsberichterstattung des Bundes. RKI, Berlin
- Schroedter JH, Lechert Y, Lüttinger P (2006) Die Umsetzung der Bildungsskala ISCED-1997 für die Volkszählung 1970, die Mikrozensus-Zusatzerhebung 1971 und die Mikrozensus 1976–2004 (Version 1). ZUMA-Methodenbericht 2006/08
- Waksberg J (1978) Sampling Methods for Random Digit Dialing. *Journal of the American Statistical Association* 73: 40–46

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
Abteilung für Epidemiologie und
Gesundheitsberichterstattung
Dr. Cornelia Lange, Dr. Thomas Ziese
General-Pape-Straße 62-66
12101 Berlin

Bezugsquelle

Die »Beiträge zur Gesundheitsberichterstattung
des Bundes« können kostenlos bezogen werden.

E-Mail: gbe@rki.de

www.rki.de/gbe

Tel.: 030-18754-3400

Fax: 030-18754-3513

Zitierweise

Robert Koch-Institut (Hrsg) (2012) Daten und Fakten:
Ergebnisse der Studie »Gesundheit in Deutschland aktuell 2010«.
Beiträge zur Gesundheitsberichterstattung des Bundes. RKI, Berlin

Grafik/Satz

Gisela Winter, Kerstin Möllerke
Robert Koch-Institut

Druck

Ruksaldruck, Berlin

ISBN

978-3-89606-213-0