

Schriftenreihe der ASI – Arbeitsgemeinschaft
Sozialwissenschaftlicher Institute

Sabine Häder · Michael Häder
Patrick Schmich *Hrsg.*

Telefonumfragen in Deutschland



Springer VS

Schriftenreihe der ASI – Arbeitsgemeinschaft Sozialwissenschaftlicher Institute

Reihe herausgegeben von

F. Faulbaum, Duisburg, Deutschland

S. Kley, Hamburg, Deutschland

B. Pfau-Effinger, Hamburg, Deutschland

J. Schupp, Berlin, Deutschland

J. Schröder, Mannheim, Deutschland

C. Wolf, Mannheim, Deutschland

Reihe herausgegeben von

Frank Faulbaum
Universität Duisburg-Essen

Stefanie Kley
Universität Hamburg

Birgit Pfau-Effinger
Universität Hamburg

Jürgen Schupp
DIW Berlin

Jette Schröder
GESIS – Leibniz-Institut für
Sozialwissenschaften

Christof Wolf
GESIS – Leibniz-Institut für
Sozialwissenschaften

Weitere Bände in der Reihe <http://www.springer.com/series/11434>

Sabine Häder · Michael Häder
Patrick Schmich
(Hrsg.)

Telefonumfragen in Deutschland

 Springer VS

Hrsg.

Sabine Häder
GESIS – Leibniz-Institut für
Sozialwissenschaften
Mannheim, Deutschland

Patrick Schmich
Robert Koch-Institut
Berlin, Deutschland

Michael Häder
Technische Universität Dresden
Dresden, Deutschland

ISSN 2625-9427

ISSN 2625-9435 (electronic)

Schriftenreihe der ASI – Arbeitsgemeinschaft Sozialwissenschaftlicher Institute

ISBN 978-3-658-23949-7

ISBN 978-3-658-23950-3 (eBook)

<https://doi.org/10.1007/978-3-658-23950-3>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer VS

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2019

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Springer VS ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Inhaltsverzeichnis

Sabine Häder & Michael Häder

Vorwort 3

Bärbel-Maria Kurth & Patrick Schmich

Vorwort Robert Koch-Institut 7

Sabine Häder & Michael Häder

1 Einleitung und Motivation 11

Hagen von Hermann

2 Das Total Survey Error Modell in telefonischen Befragungen 17

Siegfried Gabler & Sabine Häder

3 Repräsentativität: Versuch einer Begriffsbestimmung 35

Sabine Häder & Matthias Sand

4 Telefonstichproben 45

Tobias Gramlich, Martin Liebau & Jürgen Schunter

5 Erhebung und Verwendung von Para- und Metadaten 81

Martin Liebau, Jürgen Schunter, Ronald Schurath & Rainer Schwarz

6 Fragebogenkonstruktion bei telefonischen Befragungen 113

*Matthias Wetzstein, Marika Varga, Johannes Lemcke, Luise Richter,
Jennifer Allen, Martin Liebau & Patrick Schmich*

7 Feldphase 153

Hagen von Hermann

8 Rolle und Einfluss von Interviewern in telefonischen Umfragen 193

Michael Häder & Robert Neumann

9 Datenqualität 241

Tobias Gramlich, Martin Liebau & Jürgen Schunter

10 Die Berechnung von Responseraten 293

Matthias Sand & Siegfried Gabler

11 Gewichtung von (Dual-Frame -) Telefonstichproben 349

Tobias Gramlich & Sabine Häder

12 Methoden- und Feldberichte 393

Jennifer Allen & Johannes Lemcke

13 Ausschreibung einer Telefonstudie 405

Anhang 1

Beschreibung der verwendeten Erhebungen 425

Anhang 2

Empfehlungen zu Inhalten von Methoden- und Feldberichten 435

Stichwortverzeichnis 473

Die Autorinnen und Autoren dieses Bandes 477

Vorwort

Unter dem Titel „Fragwürdige Umfragen: Von wegen repräsentativ“ berichten Björn Christensen und Sören Christensen am 16. Oktober 2015 im Spiegel: „Selbst Telefonumfragen können verzerrt sein, wenn ein Teil der Angerufenen die Teilnahme verweigert und einfach auflegt. Ob sich die Ergebnisse sinnvoll auf alle Deutschen übertragen lassen, hängt sehr von der Fragestellung ab. Häufig hilft nur der gesunde Menschenverstand, um die Ergebnisse einer Umfrage richtig bewerten zu können.“ Das zentrale Ziel dieses Bandes ist es, die angesprochene Beliebigkeit in der Bewertung von Ergebnissen aus Telefonumfragen zu verringern und dem gesunden Menschenverstand ein wenig wissenschaftliche Hilfestellung zu leisten.

Der vorliegende Band verdankt sein Entstehen einer ganzen Reihe von Telefonumfragen zugewandten Personen und Institutionen. An erster Stelle soll die Deutschen Forschungsgemeinschaft (DFG) genannt werden, die die Förderung des MiMoSe-Projektes übernommen hatte. Zwei anonyme DFG-Gutachter trugen mit ihren Hinweisen dazu bei, das Profil des Projekts zu schärfen. Das Manuskript entstand als ein Ergebnis dieser Projektförderung. Unser Dank richtet sich auch an die Arbeitsgemeinschaft Sozialwissenschaftlicher Institute (ASI), die das Erscheinen dieses Bandes in ihrer Reihe ermöglicht hat.

Viele weitere Personen haben bei dem Projekt mitgearbeitet und können sich für dessen Erfolg verantwortlich fühlen. Nahezu alle jene finden sich in diesem Band als Autorin oder als Autor wieder. Wichtig erscheint uns der Hinweis, dass Inhalt und Struktur des Bandes von allen gemeinsam diskutiert und auch alle Entwürfe in der Gruppe beraten wurden. In einer Reihe von Abschnitten in diesem Band werden die Ergebnisse aus unterschiedlichen Studien berichtet. Die Aufbereitung der verschiedenen Datensätze ist ebenfalls das Ergebnis gemeinsamer Bemühungen im Rahmen der Arbeiten am MiMoSe-Projekt. Es verdient auch erwähnt zu werden, dass die Gruppe der Verfasser der vorliegenden Texte aus dem akademischen Bereich (GESIS und TU Dresden), aus dem Bereich der Datennutzer (Robert Koch Institut Berlin) sowie aus dem Bereich der Erhebungsinstitute (USUMA

GmbH) stammt. Sie alle bringen ihre spezifischen Erfahrungen mit telefonischen Befragungen ein.

Mit den „Telefonumfragen in Deutschland“ wenden wir uns an verschiedene Leserkreise. Im Blick haben wir vor allem Forscher mit ganz unterschiedlichen professionellen Hintergründen, die ein Interesse an Informationen über die soziale Wirklichkeit haben. Dies können beispielsweise Politikwissenschaftler, die sich für Wahlprognosen interessieren, Soziologen, die etwas über den Wertewandel erfahren möchten, Erziehungswissenschaftler, die an den Determinanten von Bildungserfolgen interessiert sind, Mediziner, die sich mit der Einstellung der Bevölkerung zum Impfen auseinandersetzen, Juristen, die die Hintergründe abweichenden Verhaltens erforschen oder Betriebswirtschaftler, die etwas über die Absatzchancen neuer Produkte erfahren möchten, sein. Sie alle sind potenzielle Nutzer telefonischer Befragungen und damit häufig auch Kunden entsprechender Umfrageinstitute. Vermutlich verfügen sie jedoch nur über relativ geringe Kenntnisse zum Funktionieren dieses Instruments. Weiterhin dachten wir beim Leserkreis an Studierende, die sich im Rahmen ihrer Ausbildung mit telefonischen Befragungen beschäftigen. Schließlich wenden wir uns auch an die Kolleginnen und Kollegen aus den Umfrageinstituten, die für ihre Kunden telefonische Befragungen anbieten. Wir wollen ihnen eine Hilfestellung bieten, um ihrer Klientel die erforderlichen Informationen, angefangen bei der Ausschreibung einer Studie, über die relevanten Aspekte der Fragebogenerstellung und zum Verständnis der Auswahlstrategie, bis hin zum Inhalt der Methodenberichte, zu geben.

Die Struktur des Buches verlangt es nicht unbedingt, dass alle Abschnitte der Reihe nach gelesen werden. Die einzelnen Texte enthalten Informationen zu spezifischen Sachverhalten und wir haben uns bemüht, diese so zu verfassen, dass sie – ähnlich wie in einem Handbuch – für den Leser jeweils für sich verständlich sind. Im Anhang finden sich dann weitere Hinweise vor allem für die Dokumentation telefonischer Studien. Diese sind wiederum für Sekundäranalysen von großer Bedeutung.

Wir möchten uns ganz herzlich bei Bettina Zacharias (GESIS, Team Publikationen) für die umsichtige und geduldige Fertigstellung des Manuskripts bedanken.

Die Herausgeber sowie die Autorinnen und Autorin würden sich freuen, wenn mit diesem Band die Grundlagen für erforderliche methodische Standards bei Telefonumfragen gelegt werden können. Diese sollten als Orien-

tierung dienen, um diesem Instrument zu einer angemessenen Qualität zu verhelfen – und damit eher auf eine theoretische und praktische Fundierung als den gesunden Menschenverstand zu vertrauen.

Vorwort Robert Koch-Institut

Das Robert Koch-Institut hat als Public Health-Institut die Aufgabe, den Gesundheitszustand und das Gesundheitsverhalten der in Deutschland lebenden Bevölkerung zu beobachten, Trends bzw. Trendwenden festzustellen und diese Erkenntnisse adressatengerecht zu kommunizieren. So kann unter anderem eine Evidenzbasierung für gesundheitspolitische Maßnahmen geschaffen und im Idealfall folglich auch deren Auswirkung beobachtet werden. Die hierfür zur Verfügung stehenden Datenquellen sind für Infektionskrankheiten reichhaltiger als für die nichtübertragbaren Krankheiten. Die Lücken, die für letztere durch die routinemäßig erfassten Informationen gelassen werden, wurden am Institut seit über 30 Jahren durch repräsentative Untersuchungssurveys geschlossen. Der Vorteil, dass dafür die Teilnehmenden nicht nur befragt, sondern auch standardisiert untersucht werden, hat den Preis eines höheren Zeit- und Kostenaufwandes. Bestimmte Fragen brauchen eine schnellere Antwort, als sie von den Untersuchungssurveys gegeben werden können. Daher wurde um die Jahrtausendwende am Robert Koch-Institut damit begonnen, telefonische Surveys in Eigenregie durchzuführen. Mit dem ersten offiziell ausgewerteten telefonischen Survey im Jahre 2003 (GsTELO3) war eine effektive und schnelle Methode zur Informationsgewinnung etabliert. Diese wurden dann fester Bestandteil des ab 2008 durch das RKI durchzuführenden Gesundheitsmonitorings (Kurth et al. 2009). Durch die Einrichtung eines eigenen CATI-Labors hatte das RKI die Möglichkeit, die Informationsgewinnung von der wissenschaftlichen Konzipierung über die Umsetzung und Durchführung bis zur Datennutzung in einer Hand zu behalten und entsprechende Erfahrungen in allen drei Bereichen zu gewinnen. Auch wenn heute die Befragungsstudien des Gesundheitsmonitorings (anders als die Untersuchungssurveys) größtenteils nach außen vergeben werden, bleiben diese Erfahrungen bei der Gestaltung der Schnittstellen der drei Bereiche von großem Vorteil.

Zwar hat das Telefon als Instrument zur Erfassung von Informationen mittlerweile an Marktanteilen verloren (laut ADM liegt der Anteil quantitativer telefonischer Befragungen ihrer Mitgliedsinstitute 2017 bei 29%),

dennoch bleiben telefonische Befragungen ein wirkungsvolles Instrument zur Informationsgewinnung. Dies liegt auch daran, dass in den letzten Jahrzehnten viel in die methodische Forschung zu diesem Erhebungsmodus investiert wurde und dass dadurch eine fundierte Basis zur Nutzung des telefonischen Modus vorliegt. Dieser Umstand kann als besonderer Vorteil des telefonischen Erhebungsmodus gesehen werden. So sind für diesen Modus grundlegende Fragen des Stichprobendesigns, der Fragebogengestaltung und der Qualitätssicherung bereits beantwortet.

Wie bereits erwähnt versetzt der telefonische Erhebungsmodus das RKI in die Lage, schnell und effizient Daten und die damit verbundenen Informationen für die wissenschaftliche beziehungsweise politische Diskussion zu generieren. Die strukturierten Prozesse ermöglichen eine effiziente Steuerung der jeweiligen Studien. Das RKI setzt mit dem Format der Ad-hoc Studien auch weiterhin auf den telefonischen Modus. Die Ad-hoc Studien stellen eine effiziente und schnelle Erweiterung zu den großen und aufwendigen Monitoringstudien des RKI dar (vgl. Schmich et al. 2018). Dabei setzt das RKI auf eine enge Zusammenarbeit in Form eines Rahmenvertrags mit einem externen Berliner Markt- und Sozialforschungsinstitut. Dieses Vorgehen erlaubt dem RKI, den Prozess der Datenerhebung durch eigene Supervisorinnen und Qualitätssicherungsmaßnahmen transparent zu gestalten.

Das Ziel dieses Sammelbandes ist es, eine umfassende Aufstellung der Prozesse zu bieten, die mit telefonischen Befragungen zusammenhängen, und Qualitätskriterien zu formulieren. Dies ist umso bedeutsamer, da die stetig wachsende Nachfrage nach evidenzbasierten Daten in Verbindung mit einem ebenfalls steigenden Kostendruck öffentlicher Institutionen in den letzten Jahren zu einem erheblichen Druck auf Firmen im Segment der Markt- und Meinungsforschung geführt hat. So häufen sich Meldungen in unterschiedlichen Medien (siehe u.a. Spiegelonline vom 1.2.2018: <http://www.spiegel.de/wirtschaft/unternehmen/manipulation-in-der-marktforschung-wie-umfragen-gefaelscht-werden-a-1190711.html>), in denen dezidiert Fälschungen von Umfragen thematisiert werden. Diese Fälschungen sind nicht zuletzt auf mangelnde Transparenz der Arbeit der Erhebungsinstitute, aber auch auf die mangelnde Erfahrung der Auftraggeber zurückzuführen. Für die Außenvergabe von telefonischen Befragungen und die dazu erforderlichen Ausschreibungen muss letztlich der gesamte Prozess der Datenerhebung im Vorfeld durchdacht und damit auch bekannt sein. Potentielle Auftraggeber (inhaltlich Forschende) benötigen hierfür als

Handreichung eine Art „Checkliste“ für die Prozesse und Qualitätsparameter, die beachtet werden müssen. Diesem Bedürfnis trägt der vorliegende Band kompakt und in komprimierter Form Rechnung, was für interessierte Forschende in diesem Feld effizienter sein könnte als eine Vielzahl von Einzelpublikationen.

Das RKI ist im Rahmen des Gesundheitsmonitorings nach wie vor auf den telefonischen Mode als Erhebungsmethode angewiesen und hat in den letzten 10 Jahren solche Befragungen mit ca. 100.000 Personen über diesen durchgeführt. Die, zum Teil durch ein hauseigenes Telefonlabor realisierten, Studien haben eine große Menge an Prozessdaten generiert. Um diese aufzubereiten und zu analysieren, wurde von Anfang an eine enge Kooperation mit externen Institutionen wie z.B. der TU Dresden oder der GESIS angestrebt. Wir sind stolz, unsere Ergebnisse in diesem Band einem methodisch interessierten oder inhaltlich forschenden Adressatenkreis zugänglich machen zu können und unsere Erfahrungen mit den Schnittstellen von Theorie und Praxis zu teilen.

Literatur

- ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. <https://www.adm-ev.de/zahlen/?L=1%2525252525252527#c245> (Zugriff am 25.03.2018)
- Schmich, P., Lemcke, J., Zeisler, M. L., Müller, A., Allen, J. & Wetzstein, M. (2018). Ad-hoc-Studien im Robert Koch-Institut. *Journal of Health Monitoring* (Manuskript in Vorbereitung).
- Kwasniewski, N., Maxwill, P., Seibt, P. & Siemens, A. (2018). Wie Umfragen gefälscht und Kunden betrogen werden. *Spiegel Online* vom 01.02.2018. Erhältlich auch unter: <http://www.spiegel.de/wirtschaft/unternehmen/manipulation-in-der-marktforschung-wie-umfragen-gefaelscht-werden-a-1190711.html>
- Kurth, B.-M., Lange, C., Kamtsiuris, P. & Hölling, H. (2009). Gesundheitsmonitoring am Robert Koch-Institut Sachstand und Perspektiven. *Bundesgesundheitsblatt* 52, 557-570. DOI 10.1007/s00103-009-0843-3



1 Einleitung und Motivation

Es existieren schon diverse deutsche und internationale Publikationen zur Methodik von Telefonumfragen. Wozu wird also ein weiteres Werk benötigt? In der Diskussion um Erhebungsmethoden melden sich gegenwärtig verstärkt Protagonisten, die eine Verlagerung weg von Zufallsstichproben, darunter insbesondere auch für telefonische Befragungen, und hin zu non-probabilistischen Auswahlen skizzieren. Dies wird mit sinkenden Responderaten, aber auch technischen Veränderungen begründet. Tatsächlich muss gegenwärtig von einer starken Dynamik bei der technischen Entwicklung der Telefonie ausgegangen werden. Die entsprechenden Stichworte lauten hier: immer häufigeres bzw. ausschließliches Telefonieren über das Internet (IP-Telefonie) sowie verstärkte asynchrone Kommunikation, bei der unangekündigte Anrufe inzwischen oftmals als unhöflich gelten (Spehr, 2017). Beides stellt den Modus Telefonbefragungen vor Herausforderungen. Die Autorinnen und Autoren des vorliegenden Bandes meinen jedoch, dass die in der Vergangenheit bewährten Telefonumfragen in naher Zukunft keinesfalls vor dem Aus stehen, sondern im Gegenteil noch akribischer erforscht werden sollten, um Zeit für die theoretische und praktische Entwicklung alternativer Befragungsmethoden zu gewinnen und in der Zwischenzeit trotzdem akkurate Umfrageergebnisse garantieren zu können. Ein Beitrag zur weiteren methodischen Durchdringung des Prozesses telefonischer Befragungen soll hier geleistet werden.

Die Autorinnen und Autoren beziehen ihre Erkenntnisse aus verschiedenen Quellen:

Zunächst wurde vor einiger Zeit im Arbeitskreis Mobilsample, einem Zusammenschluss von Statistikern und Methodikern aus der akademischen und kommerziellen Umfrageforschung, diskutiert, dass es an verschiedenen Institutionen teilweise recht umfangreiche, aber bislang ungenutzte Datensätze aus Telefonbefragungen gibt. Gerade in diesem Befragungsmodus fallen bei der Erhebung Para- und Metadaten an, die aus methodischer Sicht relevante Informationen zur Befragung enthalten.

Eine weitere Motivationsquelle ist mit der Tatsache verbunden, dass telefonische Befragungen der deutschen Allgemeinbevölkerung derzeit nur noch sinnvoll parallel über das Festnetz und über den Mobilfunk erhoben werden können. Damit handelt es sich um einen Mixed-Device-Ansatz, der aus methodischer Sicht ebenfalls eine besondere Herausforderung darstellt und Anlass zu neuen Betrachtungen gibt. Auf diesem Gebiet sind die Autorinnen und Autoren dieses Bandes führend in der deutschen Umfrageforschung.

Telefonische Befragungen werden in der Bundesrepublik zumeist von Telefonlabors geführt, die von kommerziellen Umfrageinstituten betrieben werden und die – beispielsweise mit den Predictive Dialern sowie mit umfangreichen Interviewerstäben – über eine anspruchsvolle technische und personelle Ausrüstung verfügen. Solche Institute besitzen damit ebenfalls ein hohes Maß an sehr spezifischer Kompetenz bei der Umsetzung telefonischer Befragungen. Damit lag die Einbeziehung dieser Kompetenz bei der Erstellung des vorliegenden Bandes nahe.

Der Verbund dieser drei Erkenntnisquellen führte zu einem erfolgreichen Projektantrag an die Deutsche Forschungsgemeinschaft. Dieser beinhaltete die gemeinsame Bearbeitung der oben genannten Fragestellungen unter der Maxime, für anwendungsorientierte Forscher Empfehlungen zum Einsatz von Telefonbefragungen geben zu können. Dafür erschienen insbesondere Sekundäranalysen mit den vorliegenden Datensätzen als die geeignete Strategie. Bearbeitet wurde das Projekt „Qualitätsmanagement bei Telefonbefragungen im Rahmen von Mixed Mode Surveys (MiMoSe)“ schließlich von Kolleginnen und Kollegen von GESIS - Leibniz Institut für Sozialforschung in Mannheim, vom Institut für Soziologie an der Technischen Universität Dresden, vom Robert Koch-Institut in Berlin und vom Markt- und Sozialforschungsinstitut USUMA GmbH in Berlin.

Im Verlauf der gemeinsamen Arbeit stellte sich heraus:

- Telefonische Befragungen über das Festnetz und den Mobilfunk sind auch weiterhin ein attraktiver Erhebungsmodus, der gegenüber anderen Modi zahlreiche Vorteile bietet. Relativ geringe Kosten, eine gute Kontrolle der Interviewsituation und eine fehlende territoriale Klumpung der geführten Interviews sind dafür wichtige Argumente. In der Abbildung 1 wird deutlich, welche Rolle telefonische Befragungen im Vergleich zu anderen Modi bei den ADM-Instituten bisher gespielt haben. Die Autorinnen und Autoren

rechnen mit einer stetigen Fortführung der angegebenen Trends und damit einer Berechtigung telefonischer Befragungen auch in naher Zukunft.

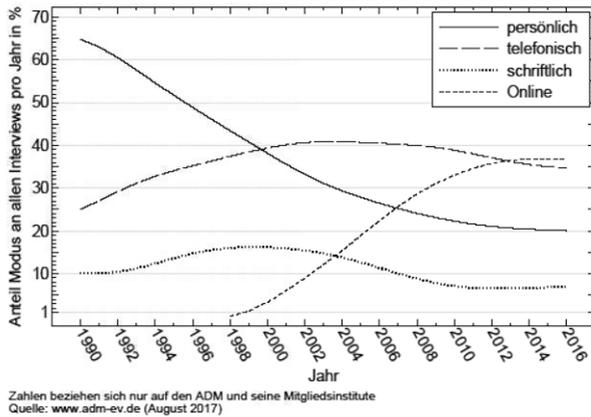


Abbildung 1.1 Relativer Anteil der Befragungsmodi
(eigene Darstellung, Quelle: ADM, 2016)

- Es existieren noch zahlreiche Wissens- und damit Forschungslücken beim Einsatz telefonischer Befragungen. Qualitätsstandards und Best-Practice-Routinen verdienen vor allem Aufmerksamkeit. Wir bemerkten eine große Beliebigkeit bei der Umsetzung telefonischer Befragungen in nahezu allen Phasen des Forschungsprozesses. Die Bestimmung der Response Rate, die gesamte Dokumentation der Methodik, die Kontrolle der Interviewer, die Anwendung von Gewichtungszprozeduren, die Bewertung der Datenqualität, die Erhebung von Para- und Metadaten, die genutzten Anrufzeiten und die Anzahl an Anrufversuchen, um nur einige wesentliche Aspekte zu nennen, zählen dazu.
- Zu konstatieren ist ein Mangel an Literatur mit einem komplexen Überblicks- und Anleitungscharakter für die Nutzer des Instruments Telefonbefragungen. Aufgrund verschiedener (vor allem technischer) nationaler Besonderheiten beim Telefonieren ist dabei ein Fokus auf die Bedingungen in der Bundesrepublik durchaus angebracht.

Daraus resultiert nun unser mit dem vorliegenden Band verbundenes Anliegen. Anhand des umfangreichen Materials¹ soll den Nutzern von Daten aus telefonischen Befragungen, die nicht über tiefere Kenntnisse zu den empirischen Methoden verfügen, die aber Interesse am Zustandekommen der Daten haben, die Grundlagen dieses Verfahrens nahegebracht werden. Gedacht haben wir beispielsweise an Personen, die mit den Daten Sekundäranalysen vornehmen wollen, an Multiplikatoren wie etwa Journalisten oder Pädagogen, an empirisch arbeitende Politik- und Kommunikationswissenschaftler usw. Ihnen möchten wir vor allem zeigen, wie die Daten aus Telefonbefragungen zustande kommen und welche Qualität sie erwarten können. Hilfestellung soll aber auch jenen Personen gegeben werden, die eine eigene telefonische Befragung bei einem Erhebungsinstitut in Auftrag geben möchten. Ein zentraler Gedanke dabei ist, die dafür erforderlichen Kriterien herauszuarbeiten. Weiterhin sollen geeignete Indikatoren für die Bewertung der Qualität von Befunden aus telefonischen Befragungen aufgezeigt werden.

Wir möchten den mit telefonischen Befragungen vertrauten Fachwissenschaftlern Standards vorschlagen, wie mit den diversen Problemen dieses Erhebungsinstruments umgegangen werden kann. Beispiele sind hier die Responseratenberechnung und die Entwicklung von Gewichtungspfezuren.

Es sollen sich Studierende – vor allem der Sozialwissenschaften – mit dieser Grundmethode der empirischen Sozialforschung vertraut machen können. Es wäre ein guter Erfolg, wenn wir Interessenten damit auch einen Berufseinstieg ermöglichen.

Der vorliegende Band besitzt folgenden Aufbau:

Als Klammer für alle Überlegungen dient den Autorinnen und Autoren die Qualitätssicherung. Dazu wird das theoretische Wissen um den Total Survey Error (TSE) herangezogen. Der Anspruch, diesen Fehler zu minimieren, gilt schließlich für alle Phasen der telefonischen Befragung. Dazu müssen

1 Die Analysen in diesem Band basieren im Wesentlichen auf den folgenden drei Studien des RKI: Influenza, GEDA2010 und GEDA2012. Außerdem werden die Ergebnisse von CELLA1 und CELLA2 genutzt. Diese Datenquellen werden in einem gesonderten Abschnitt im Anhang ausführlich vorgestellt.

möglichst alle die Qualität beeinflussenden Faktoren bekannt sein und entsprechend gesteuert werden. Das Ablaufschema einer empirischen Studie bestimmt daher die Struktur der Darstellungen. Mithilfe spezieller Stichprobendesigns werden zunächst die für die Projektbearbeitung geeigneten Untersuchungseinheiten ausgewählt. Im Verlauf der Erhebung im Feld sollten neben den substanziellen Fragestellungen auch bestimmte Meta- und Paradata gewonnen werden, die eine Bewertung der Datenqualität erlauben. Der Fragebogen steuert dann die gesamte telefonische Untersuchung. Während der Feldarbeit erheben die Interviewerinnen und Interviewer die relevanten Daten. Dabei unterliegen sie ebenfalls bestimmten Maßnahmen der Qualitätssicherung, beispielsweise in Form von Schulungen und der Kontrolle. Die erzielten Responseraten sind ein Indikator (von mehreren) für die Qualität der erzeugten Daten. Die Ermittlung dieser Responseraten erfordert ebenfalls besondere Aufmerksamkeit. Dies gilt auch für die Gewichtungungsverfahren. Das Projekt bedarf schließlich einer ordentlichen Dokumentation, um die einzelnen Schritte nachvollziehen zu können. Im Anhang zu diesem Band finden sich schließlich hilfreiche Unterlagen für die Nutzer bzw. für die Auftraggeber einer telefonischen Befragung.

Literatur

Spehr, M. (2017). Wozu noch Festnetz? *Frankfurter Allgemeine Sonntagszeitung* vom 9. Juni 2017. Erhältlich auch unter:
<http://www.faz.net/aktuell/technik-motor/digital/alles-mobil-oder-was-wozu-noch-festnetz-15046198.html>



2 Das Total Survey Error Modell in telefonischen Befragungen

Auf einen Blick

- ▶ Es lassen sich verschiedene Qualitätsdimensionen unterscheiden, die Forscher bei der Planung und Durchführung von telefonischen Studien berücksichtigen sollten.
- ▶ Grundlegend wird die Güte einer Studie anhand der zu erwartenden systematischen und zufälligen Fehler, die die Genauigkeit der Ergebnisse reduzieren, gemessen.
- ▶ Das Total-Survey-Error Modell stellt eine umfassende Systematisierung der möglichen Fehlerquellen dar und erlaubt somit eine gezielte Berücksichtigung im Forschungsdesign.

Obwohl dem Begriff ‚Qualität‘ in der Regel eine wertende Bedeutung beigemessen wird, dient er in seiner einfachsten Form als wertfreie Beschreibung des Zustandes oder der Beschaffenheit von Produkten und Prozessen. Dabei lassen sich praktisch beliebig viele Eigenschaften oder Dimensionen zur Beschreibung zugrunde legen, was gerade in der Umfrageforschung zu historisch unterschiedlichen Diskursen geführt hat (Groves 2004).

Zunächst lassen sich vier grundlegende Kriterien wissenschaftlicher Standards identifizieren, nach denen sich die Güte empirischer Sozialforschung bemessen lässt. Hierzu zählen die (1) inhaltliche Relevanz der Arbeit, (2) die methodische Strenge in der Durchführung und Analyse, (3) die Einhaltung ethischer Standards sowie (4) die Präsentationsqualität (Döring und Bortz 2016). Die Bewertung der methodischen Strenge (*„scientific rigor“*) in Form von Erhebungsinstrumenten und deren Anwendung, erfolgt dabei aufgrund der drei Gütekriterien *Objektivität, Reliabilität und Validität*. Einfach ausgedrückt bemisst sich die Qualität einer Erhebung danach, wie unabhängig die Messung von Einflüssen der Forscherseite ist, wie reproduzierbar die Ergebnisse sind und inwieweit die Forschungsfrage anhand der Daten beantwortet werden kann (M. Häder 2015).

Während über die Bedeutung dieser Konzepte ein genereller Konsens herrscht und sie zur konzeptionellen Arbeit unerlässlich sind, eignen sie sich nur bedingt zur Bemessung der Qualität in der Umfrageforschung. Dies liegt zum einen an ihrem Anspruch auf Allgemeingültigkeit, der sich universell auf den Forschungsprozess und alle empirischen Methoden bezieht, dabei aber den spezifischen Anforderungen von Surveys nicht gerecht wird; zum anderen an ihrem normativen Charakter, der keine quantifizierbaren und somit auch keine vergleichbaren Aussagen zur Qualität einer konkreten Umfrage erlaubt.

Der Spezialfall der Umfrageforschung und damit auch die Frage nach der Qualitätsdefinition ergibt sich aus den Eigenschaften des Erhebungsinstrumentes selbst – der Umfrage. Auch mit Blick auf die weitere Verwendung in diesem Buch lassen sich Umfragen dabei wie folgt definieren:

- Daten und Ergebnisse beruhen auf den Angaben von Probanden;
- Die Sammlung von Daten erfolgt durch Interviewer, die den Probanden Fragen stellen und deren Antworten notieren, oder durch die Probanden selbst, nachdem sie Fragen gelesen oder gehört haben;
- Informationen werden nur für eine Teilmenge (eine Stichprobe) der untersuchten Population erhoben.

Diese Definition von Groves et al. (2004, S. 3) lässt sich um eine weitere Kondition erweitern soweit aus den Angaben der Stichprobe auf die Lage und Verteilung von Parametern in der Gesamtpopulation geschlossen werden soll:

- Die Auswahl der Stichprobe erfolgt auf Grund eines methodisch-probabilistischen Auswahlprozesses¹.

Die als Inferenzschluss bezeichnete Ableitung von Populationsschätzern aus den Daten der Stichprobe stellt häufig das eigentliche Ziel von Umfragen dar und setzt implizit voraus, dass es für den gesuchten Schätzer einen tatsächlichen oder ‚*wahren Wert*‘ (*,true score*‘) in der untersuchten Grundgesamtheit gibt. Dabei handelt es sich um eine notwendige Annahme, die nicht immer zweifelsfrei unterstellt (Hansen et al. 1951) und die empirisch nur im Vergleich mit den Ergebnissen anderer und (vorzugsweise

1 Für eine noch detailliertere Auseinandersetzung mit dem Survey Begriff und dessen Voraussetzungen siehe Lyberg (2012).

bewährter) Umfragetechniken überprüft werden kann (Deming 1960). Anders als bei deskriptiven Analysen, die absolute Aussagen erlauben (bspw. den Anteil an Frauen in der Stichprobe), schließen die Schätzungen inferenzstatistischer Verfahren auch immer die Angabe von Fehlerbereichen (Konfidenzintervalle) und die Möglichkeit von Signifikanztests mit ein (M. Häder 2015). Damit stellt die Genauigkeit, mit der eine Schätzung den wahren Wert in einer Population abbildet eine zentrale Qualitätsdimension in der Surveyforschung dar.

2.1 Fehlertypen und Mean Squared Error

Die Qualität von Umfragedaten ergibt sich somit aus der Summe der (fehlerhaften) Abweichungen zwischen dem auf der Erhebung beruhenden Schätzer und dem wahren Wert in der Population, wobei sich zwei Fehlertypen unterscheiden lassen:

- Systematische Fehler oder Bias (im Weiteren als Bias bezeichnet) bezeichnet alle Einflussgrößen im Forschungsdesign, die zu einer systematischen Verzerrung des Schätzers gegenüber dem wahren Wert führen und auch bei wiederholten Messungen auftreten. Faulbaum beschreibt ihn daher auch als die „*durchschnittliche Performance* des Schätzers in Bezug auf den [untersuchten] Populationswert“ (2014, S. 441):

$$BIAS(\hat{y}) = E(\hat{y}) - (y) \quad (1.1)$$

Dabei beschreibt y den untersuchten Populationsparameter; \hat{y} den Parameterschätzer für die einzelne Messung und $E(\hat{y})$ den Durchschnitt der Schätzer über wiederholte Messungen bzw. dessen Erwartungswert. In Umfragen lässt sich ein solcher Bias regelmäßig bei der Untersuchung von heiklen Themen beobachten, wobei Angaben zu sozial erwünschten Verhaltensweisen (Wählen gehen oder Spenden) systematisch ‚überberichtet‘ und sozial unerwünschtes Verhalten (Drogenkonsum oder Delinquenz) ‚unterberichtet‘ werden (Biemer 2010; Krumpalet al. 2015). Dabei wird in beiden Fällen jeweils der Mittelwert der Schätzung (nach oben bzw. unten) verschoben, ohne dabei die Varianz des Schätzers zu beeinflussen.

- Der Erwartungswert $E(\hat{y})$ entspricht dabei dem Mittelwert des Schätzers \hat{y} und ergibt sich aus der Summe aller Teilstichproben, wobei $p(s)$ die Aus-

wahrscheinlichkeit und $\hat{y}(s)$ den Schätzer für die Stichprobe s darstellen:

$$E(\hat{y}) = \sum_s p(s) \hat{y}(s) \quad (1.2)$$

- Zufällige Fehler oder Variable Error (im Weiteren als Error bezeichnet) beschreiben die Varianz und damit die *Präzision* eines Schätzers, die sich aus den variierenden Angaben bei wiederholten Erhebungen ergibt. Folglich leitet sich die $VAR(\hat{y})$ aus der mittleren quadrierten Abweichung zwischen den Stichprobenschätzern und dem Erwartungswert ab:

$$VAR(\hat{y}) = \sum_s p(s) [\hat{y}(s) - E(\hat{y})]^2 \quad (1.3)$$

Bei wiederholten Messungen variiert die Zusammensetzung der Stichprobe, wodurch die jeweiligen Schätzer um den (wahren) Populationswert schwanken. Diese Abweichungen heben sich bei der Betrachtung des Erwartungswertes (also dem Mittelwert aller Schätzer) gegenseitig auf², wodurch die Lage des Erwartungswertes unberührt bleibt, während die Varianz steigt und (unter sonst gleichen Bedingungen) die Konfidenzintervalle größer werden.

Die Unterscheidung zwischen den beiden Fehlerarten ist zunächst theoretischer Natur und beruht auf der Annahme, dass Erhebungen wiederholt werden können, wobei die Varianz des Schätzers in Abhängigkeit der variablen Fehler über die einzelnen Erhebungen schwankt, die durch den Bias verschobene Lage des Schätzers hingegen, gleich bleiben wird (Groves 2004; Groves et al. 2004). Der Bias leitet sich somit direkt aus den spezifischen Eigenschaften der jeweiligen Umfrage ab, die in der Literatur als die ‚*essential survey conditions*‘ bezeichnet werden (Hansen et al. 1961). Dadurch lässt er sich in seinem Umfang auch nur mittels zusätzlicher Informationen, wie bspw. durch Vergleich mit einer anderen Umfragetechnik ermitteln („*golden standard*“).

Die Summe aus dem variablen Fehler (Error) und dem Quadrat des systematischen Fehlers (Bias) ergibt den Mean Squared Error (MSE) bzw. die *mittlere quadratische Abweichung*. Dieser stellt die Gesamtheit aller Fehler-

2 Ergibt die Summe der Abweichungen zwischen dem wahren Populationswert und den jeweiligen Schätzern nicht Null liegt somit offensichtlich ein Bias vor.

quellen auf einen spezifischen Schätzer und somit den zentralen Indikator für die Qualität der Erhebungsdaten dar. Er ergibt sich aus der Summe der Varianz des Schätzers und dem Quadrat des Bias:

$$MSE(\hat{y}) = E[(\hat{y} - y)^2] = \sum_s p(s) [\hat{y}(s) - y]^2 = BIAS(\hat{y})^2 + VAR(\hat{y}) \quad (1.4)$$

Während Schätzer ohne Bias als unverzerrt oder erwartungstreu (unbiased estimator) und Schätzer mit geringer Varianz als präzise (precise estimator) bezeichnet werden, sind Bemühungen zur Steigerung der Qualität immer auf die Reduktion von beiden Fehlerarten ausgerichtet. Umso niedriger der MSE ist, desto geringer ist die Streuung der Messwerte um den Schätzer und umso stärker nähert sich der Schätzer dem wahren Populationswert an (Groves 2004).

Dabei lässt sich der MSE immer nur für ein spezifisches Item oder eine Teststatistik berechnen, was vor allem einen Vergleich identischer Items über unterschiedliche Erhebungen nahe legt. Während der Bias in der Regel nur indirekt bestimmt werden kann und im Verhältnis zum Varianzterm häufig vernachlässigt wird, steigt der Einfluss des Bias mit dem Umfang der Stichprobe an und kann diese bei großen Stichprobenzahlen entsprechend dominieren (Kish 1987). Obwohl der MSE eine Grundgröße in der Qualitätsmessung von Umfragedaten darstellt, wird er aufgrund der mangelnden Informationen über den Bias nur selten vollständig bestimmt (Schnell 2012).

2.2 Fehlerquellen und Total Survey Error

Dies liegt nicht zuletzt daran, dass sich zahlreiche Fehlerquellen im Survey-Prozess identifizieren lassen, die jeweils sowohl Bias als auch Error hervorrufen und somit die Qualität beeinflussen können. Das MSE-Modell (Andersen et al. 1979; Biemer und Lyberg 2003; Groves 2004; Hansen et al. 1961; Kish 1965) systematisierte diese erstmalig in einem gemeinsamen theoretischen Rahmen, was die methodische Integration von Konzepten aus unterschiedlichen wissenschaftlichen Disziplinen – vorwiegend dem Sampling Error aus den statistischen Methoden und Problemen der Messung von Einstellungen und Verhalten aus der Psychometrie – erlaubte.

Um dies zu berücksichtigen schlug Groves (Groves 2004) neben der Unterscheidung von Fehlertypen (Error / Bias) auch eine Unterteilung der Feh-

lerquellen vor, je nachdem ob sie aus einem Mangel an Beobachtung in der Zielpopulation (*Errors of nonobservation*) oder aus einer fehlerhaften Messung (*Errors of observation*) resultieren. Während Erstere sich vorwiegend der Frage widmet, ob die Stichprobe tatsächlich zur Abbildung der Grundgesamtheit taugt, beschäftigen sich Letztere mit der Adäquatheit der Messung selbst. Diese Unterscheidung wurde (wenn auch implizit) von Biemer und Lyberg (2003; Groves et al. 2004; Weisberg 2005) zur Definition des *Total Survey Errors*³ (TSE) übernommen:

$$\text{Total Survey Error} = \text{Sampling Error} + \text{Nonsampling Error} \quad (1.5)$$

Da die einzelnen Fehlerquellen in den folgenden Kapiteln näher beschrieben werden, erfolgt hier nur jeweils eine kurze Definition:

Der *Sampling Error* (oder Stichprobenfehler) beschreibt alle Fehler die daraus resultieren, dass lediglich eine Auswahl aus der zugrundeliegenden Population untersucht wird, wohingegen der *Nonsampling Error* Fehler betrachtet, die bei der Erhebung und Verarbeitung der Daten entstehen und sich in die folgenden Fehlerquellen aufteilen lassen:

Der *Specification error* (Spezifikationsfehler) behandelt das grundlegende Problem der Validität von Umfrageergebnissen und beschreibt die Diskrepanz zwischen dem theoretischen Untersuchungsobjekt und dem durch die Operationalisierung gemessenen Konstrukt.

Der *Coverage Error* beschreibt Fehler, die aus einer mangelnden Übereinstimmung des *Auswahlrahmens* (oder *sampling frames*) mit der untersuchten Grundgesamtheit resultieren. So ist beispielsweise nur noch ein geringer Anteil an vergebenen Telefonanschlüssen in zentralen Registern eingetragen. Würde man nun eine Stichprobe auf Grundlage des Telefonbuches ziehen, erhielten Anschlüsse, die nicht eingetragen sind, eine Auswahlwahrscheinlichkeit von Null (undercoverage). Dies kann dann zu einem Bias führen, wenn die untersuchte Größe mit der Wahrscheinlichkeit, sich nicht im Telefonbuch eintragen zu lassen, korreliert.

3 Wären dieser Begriff von Anderson et al. (1979) populär gemacht und seit dem regulär verwendet wurde (Groves 2004; Lyberg et al. 1997) wandelte er sich in seiner Bedeutung nach dem Buch von Biemer und Lyberg (2003) von einer allgemeinen Beschreibung der Summe aller Fehlerquellen (und damit einem Synonym für den MSE) zu einem eigenen Qualitäts- und Methodenbegriff.

Nonresponse Error tritt hingegen auf, wenn Mitglieder der Stichprobe einzelne Angaben (*Item-Nonreponse*) oder das gesamte Interview (*Unit-Nonresponse*) verweigern oder aus anderen Gründen nicht teilnehmen können. Dabei ist immer dann von einem Nonresponse-Bias auszugehen, wenn sich die Gruppe der Verweigerer oder ‚Nicht-Erreichten‘ systematisch von den Teilnehmenden unterscheidet. Während Verweigerung insbesondere bei sensiblen Themen vorliegen, lassen sich auch triviale Beispiele benennen, die leicht zu vermeiden sind. So werden Umfragen zum Reiseverhalten von Familien mit schulpflichtigen Kindern innerhalb der Sommerferien zu anderen Ergebnissen kommen, als zu Beginn des Schuljahres.

Der *Measurement Error* gehört zu den am umfangreichsten studierten Fehlerquellen in Surveys und beschreibt alle Einflüsse, die zu einer Differenz zwischen den wahren Werten und den dokumentierten Angaben führen können. Dabei wird zwischen dem Einfluss von Befragten (Respondent Error), Interviewern (Interviewer Error), dem Erhebungsinstrument (Instrument Error) sowie dem Erhebungsmode (Mode Error) unterschieden.

Letztlich umfasst der *Processing Error* (Prozess- und technische Fehler) alle Einflüsse, die im Rahmen der Datenerhebung (Data Entry), der Datenaufbereitung (Vercodung, Analyse und Gewichtung) und der Darstellung erfolgt.

Anhand der Vielzahl an Einflussfaktoren wird deutlich, dass eine Kontrolle aller Fehlerquellen praktisch unmöglich ist und auch nachträgliche Korrekturen Verzerrungen nicht vollständig beheben können. Dies liegt nicht zuletzt daran, dass die einzelnen Fehlerquellen sich gegenseitig beeinflussen, was eine ex-post Identifikation nur in Grenzen zulässt (West und Olson 2010). Darüber hinaus geben Biemer und Lyberg zu bedenken, dass die einzelnen Fehlerquellen ein zumeist ungleiches Risiko für die Einführung von systematischen und zufälligen Fehlern beinhalten (vgl. Tabelle 2.1).

Tabelle 2.1 Risiko von Survey Error und Survey Bias aus unterschiedlichen Fehlerquellen

MSE Komponente	Risiko für zufälligen Fehler (Error)	Risiko für systematischen Fehler (Bias)
Sampling Error	Hoch	Niedrig
Specification Error	Niedrig	Hoch
Frame Error	Niedrig	Hoch
Nonresponse Error	Niedrig	Hoch
Measurement Error	Hoch	Hoch
Processing Error	Hoch	Hoch

Der ungleiche Einfluss von Fehlerquellen auf Fehlerarten lässt sich analog auch in einer Anpassung von der Berechnungsformel des Mean Squared Errors ausdrücken. Dabei werden nur die Fehlerformen berücksichtigt, mit denen ein hohes Risiko für die Schätzung von Parametern einhergeht (Biemer und Lyberg 2003, S. 59):

$$\begin{aligned}
 MSE(\hat{y}) &= BIAS^2 + VAR \\
 &= \left(BIAS_{Specification} + BIAS_{Nonresponse} + BIAS_{Frame} + BIAS_{Measurement} \right. \\
 &\quad \left. + BIAS_{processing} \right)^2 + Var_{Sampling} + Var_{Measurement} + Var_{Processing}
 \end{aligned}
 \tag{1.6}$$

2.3 Total Survey Error – vom Produkt zum Prozess

Während in Bezug auf die Produktqualität (also die Genauigkeit der Schätzung) es keinen Unterschied zwischen Mean Squared Error und Total Survey Error gibt und diese Begriffe tatsächlich synonym verwendet werden, erlaubte das TSE-Modell die Zuordnung von Fehlerquellen zu den einzelnen Arbeitsschritten im Forschungsprozess und ermöglicht damit vor allem Anwendern einen leichteren Zugang zum Thema Survey-Qualität. Darüber hinaus berücksichtigt es die Interdependenzen zwischen den einzelnen Bestandteilen des MSE und erlaubt somit eine ganzheitliche Betrachtungen der Surveyqualität über den kompletten Forschungsprozess hinweg (Groves et al. 2004). Das TSE-Modell wird inzwischen als integraler Bestandteil

eines umfassenderen Paradigmas gesehen, bei dem die Optimierung der Schätzgenauigkeit um eine Reihe von weiteren Qualitätsdimensionen ergänzt wird (Biemer 2010; Lyberg 2012). Dieses als Total Survey Quality (TSQ) bezeichnete Konzept entwickelte sich als späte Reaktion auf zwei lange vernachlässigte Herausforderungen.

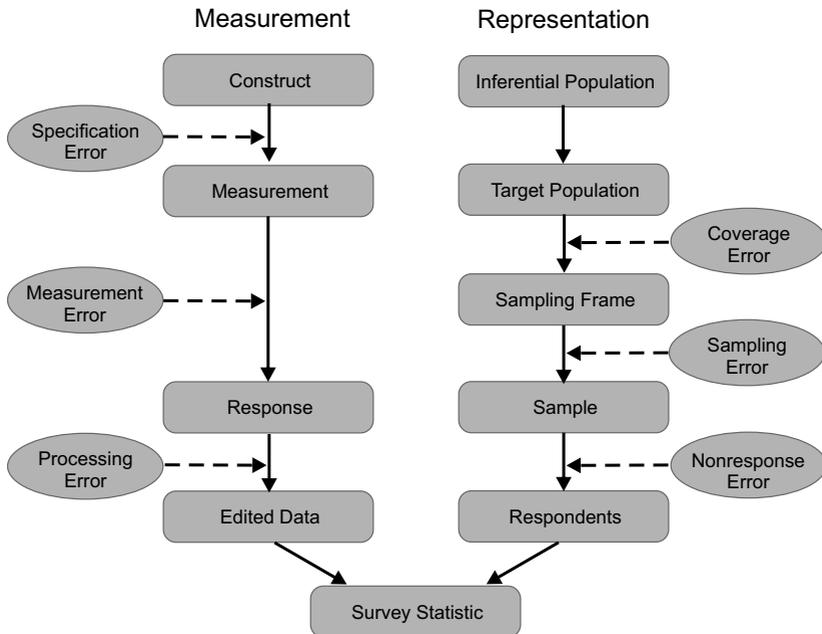


Abbildung 2.1 Lebenszyklus einer Umfrage im Total Survey Error Modell (Groves und Lyberg 2010, S. 856)

Zum einen bestand schon lange ein Konsens darüber, dass es in der Surveypraxis an einer Anwender- oder Nutzerperspektive mangelt. So mahnten bereits Mahalanobis (1965) oder Dalenius (1985), dass die Relevanz von Umfragen vom Grad ihrer faktischen Nutzung abhängt und diese ‚fitness to use‘ eben letztlich durch den Nutzer oder Auftraggeber bestimmt werden (Juran und Godfrey 1998; Krejci 2010). Zum anderen hielt mit der Nutzerperspektive umgehend auch eine zweite relevante Größe Einzug: Die Rolle

der Kosten und die Gegenüberstellung des Forschungsbudgets mit der damit realisierbaren Qualität (Groves 2004; Lyberg und Stukel 2017).

Dieses Effizienzdenken wurde wiederum durch zwei ebenfalls jüngere Entwicklungen befördert. Während erste wissenschaftliche Methoden der Qualitätskontrolle sich bereits in den 1950er Jahren ausbildeten (Juran und Godfrey 1998), entwickelten sie sich zum Ende des vergangenen Jahrhunderts zu eigenständigen Qualitätssystemen - wie dem Total Quality Management (TQM) oder Six Sigma - weiter, die sich auf beliebige Arbeitsprozesse übertragen ließen (Biemer 2010; Biemer et al. 2014; Lyberg und Stukel 2017). Zeitgleich erlaubte erst der technische Fortschritt eine evidenzbasierte Kontrolle und Steuerung von Surveys durch die Erhebung, Speicherung und Analyse relevanter Prozessdaten (Couper 1998; Kreuter 2013; Kreuter et al. 2010).

2.4 Umfrage-, Para-, Meta und Hilfsdaten (Auxiliary Data)

Erst mit dem Einsatz computergestützter Erhebungsmodi (CATI, Web, CASI) wurden Prozessinformationen verfügbar, die bis dahin nur mit besonderem zusätzlichem Aufwand ermittelt werden konnten. Dazu zählen beispielsweise die Kontaktprotokolle von Interviewern (inkl. Datum, Uhrzeit und Anrufergebnis), Zeitmarken bei der Eingabe von Antworten oder beim Verbleib auf Seiten von Webumfragen. Während Couper (1998) diese technischen Nebenprodukte ursprünglich als *Paradaten* bezeichnete, hat sich der Begriff inzwischen zu einem Sammelbecken für all jene Daten entwickelt, die zur Beschreibung des Erhebungsprozesses an sich herangezogen werden können (Kreuter 2013). Hierzu zählen inzwischen auch Einschätzungen und Beobachtungen von Interviewern während der Feldphase oder die Aufzeichnung der Interaktion zwischen Interviewer und Befragten selbst - also Informationen deren Erhebung mit einem Extra an Aufwand und Kosten verbunden sind.

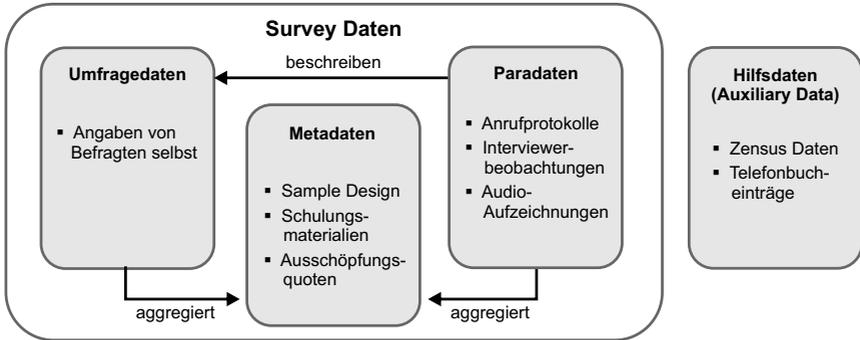


Abbildung 2.2 Datentypen in Umfragen

Die Relevanz von Paradaten im Rahmen des TSE-Modells lässt sich darüber hinaus auf zwei Eigenschaften zurückführen. Zum einen, sind sie bereits während der Feldphase und in einer beinahe beliebigen Detailtiefe⁴ verfügbar, sodass auf ihrer Grundlage direkt Einfluss auf die weitere Erhebung genommen werden kann (Kreuter et al. 2010). Zum anderen, fallen bestimmte Paradaten, wie bspw. die Anrufprotokolle, sowohl für Studienteilnehmer als auch Nicht-Teilnehmer an, was insbesondere zur Kontrolle und Korrektur von Nonresponse dient (Kreuter und Olson 2013).

Neben den Para- und Umfragedaten selbst lassen sich noch zwei weitere Datenquellen identifizieren, die inhaltlich nicht immer leicht zu unterscheiden sind und ebenfalls eine wichtige Rolle in der Surveyforschung spielen. Als *Metadaten* werden dabei Informationen bezeichnet, die die Daten der Umfrage und deren Struktur beschreiben. Sie umfassen neben Beschreibungen von Forschungs- und Sample Design, insbesondere auch Kenngrößen, die sich aus der Aggregation von Umfrage- und Paradaten ableiten lassen. Eine zentrale Metainformation stellen dabei beispielsweise Ausschöpfungs- und Verweigerungsraten dar, die sich aus den Dispositionscores der einzelnen Anrufergebnisse einer Umfrage ableiten. Eine vierte und zumeist nicht in den Prozess der Datenerhebung involvierte Kategorie stellen ‚Auxiliary Data‘ (Hilfsdaten) dar, die zumeist nach Abschluss der Erhebung und aus

4 So können beispielsweise Zeitmarken für einzelne Fragen, Fragenblöcke oder das gesamte Interview gespeichert werden.

externen Quellen (Zensus- oder Verwaltungsdaten, Registerauszüge oder Ähnliches) an die Daten angespielt werden.

2.5 Die Prozessqualität

Die Daten einer telefonischen Befragung werden in den Umfrageinstituten mithilfe eines komplexen Prozesses gewonnen. Zwischen der Vergabe eines Auftrages an das Umfrageinstitut und der Lieferung des vom Auftraggeber gewünschten Produktes liegen dabei zahlreiche Arbeitsschritte. „Die Art und Weise, wie die einzelnen Schritte des Umfrageprozesses realisiert werden, ist maßgeblich für deren Qualität und damit für die Qualität der gesamten Umfrage einschließlich der aus ihr resultierenden Daten; d.h. die Optimierung einer Umfrage beinhaltet die Optimierung aller Schritte und daher die Optimierung der Prozessqualität“ (Faulbaum 2014, S. 450).

Die Erhebungsinstitute sind im Rahmen des Qualitätsmanagements für die Kontrolle der Qualität der einzelnen Schritte verantwortlich. Für den Datennutzer ist es dagegen kaum möglich, sich ein fundiertes Bild zu jedem dieser einzelnen Schritte zu machen. Für das Qualitätsmanagement sind zahlreiche und sehr unterschiedliche Facetten relevant. Hierzu zählen die Weiterbildung der Mitarbeiter, insbesondere der Interviewer, aber auch die Personalentwicklung des Erhebungsinstituts im Allgemeinen, die kontinuierliche Erhebung von Prozessdaten, wie etwa der Response-Rate im Verlauf einer Umfrage, die Etablierung von Standards, beispielsweise für die Dokumentation der telefonischen Kontakte (Dispositions-codes), die Einhaltung von Standards zur Kooperation mit dem Auftraggeber der Studie, die Einhaltung des Datenschutzes, die Entwicklung und die Wartung der Ausstattung des Erhebungsinstituts mit Hard- und Software, die Kontrolle der Einhaltung des geplanten zeitlichen Ablaufs der Studie, die Kostenkontrolle, die Kommunikation und die Unterstützung der Datennutzer bei ihrer Arbeit mit den erzeugten Daten, die Erhebung der Nutzerzufriedenheit, die Analyse der Kundenwünsche und vieles andere mehr.

Lyberg (2012, S. 114ff.) unterscheidet zwischen der Produktqualität, der Prozessqualität und der Qualität der Organisation. Er ordnet den Qualitätsebenen jeweils spezifische Kontrollmechanismen sowie Indikatoren zu (vgl. Tabelle 2.2).