

Telefonstichproben

Sabine Häder

Zentrum für Umfragen, Methoden und Analysen, Mannheim

Zusammenfassung

Der Aufsatz behandelt theoretische Grundlagen und gibt praktische Tipps für das Ziehen von Zufallsstichproben für Telefonumfragen in Deutschland.

Summary

The paper presents theoretical foundations and practical recommendations for drawing probability samples for telephone surveys in Germany.

**ZUMA How-to-Reihe, Nr. 6
2000**

1. Warum und wann Telefonumfragen?

Der Anteil von telefonisch durchgeführten Interviews an den quantitativen Umfragen in Deutschland hat im letzten Jahrzehnt stark zugenommen. In den Mitgliedsinstituten des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute werden gegenwärtig z.B. etwa 40% aller Interviews per Telefon durchgeführt – 1990 waren dies noch 22%. Aber auch an zahlreichen Universitäten wurden inzwischen Telefonlabors eingerichtet und die Studenten erwerben nun praktische Kenntnisse der empirischen Sozialforschung, indem sie an telefonischen Lehrforschungsprojekten teilnehmen. Für die Ausbreitung der Telefonumfragen sprechen eine Reihe von Vorteilen gegenüber Face-to-face-Umfragen, dem bisherigen „Königsweg“ der Sozialforschung:

- Die Interviews können leicht regional gestreut werden, Klumpungen (d.h. die Durchführung mehrerer Interviews pro Sample Point) aus Kostengründen sind nicht notwendig. Dadurch werden die Aussagen präziser.
- Aufgrund der relativ geringen Kosten sind viele Kontaktversuche möglich. Dies verringert den Anteil der nicht erreichten Zielpersonen.
- Die Arbeit der Interviewer kann durch die zentrale Supervision besser kontrolliert werden.
- Die anonyme Interviewsituation verringert häufig die Bedenken der Zielpersonen, sich auf ein Interview einzulassen.
- Bei CATI (Computer Assisted Telephone Interviewing) kommt es gegenüber herkömmlichen Face-to-face-Befragungen (ohne Computer) zu Fehlerreduktionen, da in die Programme logische Kontrollen implementiert werden können. Außerdem können die Daten schnell aufbereitet werden und die Ergebnisse stehen schon nach kurzer Zeit zur Verfügung.

Insgesamt gelten Telefonumfragen – verglichen mit persönlich-mündlichen Befragungen - als zeit- und kostensparend. Natürlich sind sie nicht für jeden Forschungsgegenstand geeignet. Sollen z.B. differenzierte Beurteilungs- und Einstellungsskalen eingesetzt, visuelle Hilfsmittel verwendet oder lange Befragungen durchgeführt werden, sind andere Befragungsmodi vorzuziehen (vgl. Porst 2000).

Im folgenden sollen verschieden Abläufe bei der Ziehung von Zufallsstichproben für Telefonumfragen beschrieben werden. Diese ermöglichen - wenn sie ordentlich geplant und die ausgewählten Zielpersonen auch zur Mitarbeit bereit sind - zuverlässige Schlüsse von der Stichprobe auf die zugrunde liegende Gesamtheit.

2. Voraussetzung für die Stichprobenziehung: Definitive Abgrenzung der Grundgesamtheit

Zu Beginn jeder Stichprobenerhebung – also auch der für eine telefonische Umfrage – ist zunächst eine exakte Abgrenzung der Grundgesamtheit notwendig. Das bedeutet, dass eine sachliche, räumliche und zeitliche Definition der Menge von Elementen (Zielpersonen) vorgenommen werden muss, für die die aus der Stichprobe gewonnenen Aussagen gültig sein sollen (Grundgesamtheit oder Population).

Beispiele:

- Die Grundgesamtheit bilden alle wahlberechtigten Personen in der Bundesrepublik Deutschland, die zum Zeitpunkt der Befragung in Privathaushalten leben.
- Die Grundgesamtheit umfasst alle Personen, die zum Zeitpunkt der Befragung 60 Jahre oder älter sind und in einem Privathaushalt in einer der zehn kreisfreien Städte des Regierungsbezirks Düsseldorf leben.
- Die Grundgesamtheit bilden alle deutschsprachigen Frauen, die in der Bundesrepublik Deutschland in Privathaushalten leben und seit dem 01.01.1992 ein Kind bekommen oder adoptiert haben. Ausgenommen sind Auszubildende, Schülerinnen und Studentinnen.

Die Definition der Grundgesamtheit leitet sich natürlich aus dem jeweiligen Untersuchungsziel der Studie ab. Die im ersten Beispiel gegebene Definition ist z. B. für die Erhebung von Wahlabsichten geeignet. Das zweite Beispiel stammt aus einem Projekt, das an der Universität Düsseldorf durchgeführt wurde (vgl. Knesebeck/Lüschen 2000). Gegenstand war die Ermittlung des Gesundheitszustands älterer Bürger in einer bestimmten Region. Im dritten Beispiel schließlich erforschte das Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit, wie sich die Änderung des Gesetzes über den Erziehungsurlaub aus dem Jahr 1986 auf die Berufsverläufe von Frauen in Deutschland ausgewirkt hat. Es ist offensichtlich, dass zur Bearbeitung dieses Themas die ersten beiden Definitionen nicht sinnvoll gewesen wären.

Ergänzt werden müsste der Vollständigkeit halber noch – da es sich in allen drei Fällen um Telefonstudien handelt – dass die Zielpersonen über einen Festnetzanschluß in ihrem Privathaushalt verfügen müssen. Dieser Zusatz ist aber bei einer Ausstattung von Haushalten mit Festnetztelefonen von gegenwärtig 99% (Statistisches Bundesamt 1999) vernachlässigenswert.

3. Auswahl: Liste vorhanden oder nicht?

Nachdem die Grundgesamtheit definiert ist, stellt sich die entscheidende Frage: Existiert eine Liste (ein Auswahlrahmen, ein Frame), d.h. gibt es eine akzeptable Aufstellung mit allen Elementen der Grundgesamtheit? Wenn dies der Fall ist, erleichtert das die Stichprobenziehung erheblich; wenn nicht, müssen Umwege in Kauf genommen werden.

Ideal wäre eine Liste mit folgenden Eigenschaften:

- Auf der Liste sind alle Zielpersonen mit ihren Telefonnummern einmal und nur einmal aufgeführt, d.h. es fehlen keine Personen und es sind keine Personen mehrfach aufgelistet.
- Die Liste ist hinreichend aktuell, d.h. es kann davon ausgegangen werden, dass die Zielpersonen tatsächlich (noch) über die angegebenen Rufnummern erreichbar sind.
- Die Liste ist für die Erhebung zugänglich und ihre Nutzung nicht zu teuer.
- Die Liste enthält Zusatzinformationen für eine mögliche Schichtung, d.h. Merkmale, die im Zusammenhang mit Untersuchungsmerkmalen stehen dürften.

3.1 *Telefonbuch als Liste für allgemeine Bevölkerungsbefragungen?*

Noch bis 1992 war das Telefonbuch in Deutschland ein akzeptabler Auswahlrahmen für Telefonstichproben (in Westdeutschland), da Telefonteilnehmer bis zu diesem Zeitpunkt mit nur sehr wenigen Ausnahmen verpflichtet waren, ihre Telefonnummer zu veröffentlichen. Inzwischen besteht diese Eintragungspflicht nicht mehr und der Anteil sogenannter „Nonpubs“, d.h. nicht eingetragener Telefonteilnehmer, ist in den letzten Jahren stark angestiegen. Da sich Eingetragene und Nicht-Eingetragene in wesentlichen sozio-demographischen Merkmalen unterscheiden (Nicht-Eingetragene leben z.B. häufiger in Großstädten als in ländlichen Gebieten, sind jünger, haben häufiger eine formal höhere Bildung und sind häufiger geschieden als eingetragene Telefonbesitzer), würde die Ziehung aus dem Telefonbuch zu in der Verteilung dieser und davon abhängiger Merkmale verzerrten Stichproben führen. Deshalb ist das Telefonbuch mittlerweile für die Stichprobenziehung für Befragungen der allgemeinen Bevölkerung der Bundesrepublik oder auch einzelner Gemeinden nicht mehr geeignet. Dies bedeutet aber nicht, dass es überhaupt keine Verwendung bei der Stichprobenplanung mehr finden kann. Für die Konstruktion von Migrantenstichproben (vgl. Humpert/Schneiderheinze 2000) oder die Auswahl von Unternehmen oder Institutionen ist es weiterhin wertvoll.

Kein Problem für derartige Befragungen ist dagegen – wie schon bemerkt - die Ausstattung der Haushalte mit Festnetztelefonen. Diese erreichte in Westdeutschland schon Ende der 80er Jahre ein Niveau von über 90% und erhöhte sich weiter. In Ostdeutschland stieg sie von 17% im Jahr 1989 auf ebenfalls über 95% gegenwärtig. Zur Zeit existiert auch noch kein beachtlicher Trend, das Festnetztelefon zugunsten der ausschließlichen Nutzung des Mobiltelefons abzumelden. Insofern können Telefonstichproben prinzipiell zu einer guten Abbildung der Gesamtbevölkerung oder einzelner Segmente der Bevölkerung führen – wenn die nicht ins Telefonbuch eingetragenen Haushalte berücksichtigt werden.

3.2 *Liste vorhanden*

Mitunter existieren Listen, die die Grundgesamtheit in dem Sinn „akzeptabel“ abbilden, dass der Unterschied zwischen der Grundgesamtheit und der sich aus der Liste ergebenden Auswahlgesamtheit vernachlässigt werden kann. Dies kann z.B. für Mitglieder bestimmter Vereinigungen, Organisationen oder Institutionen sowie Kunden von Firmen der Fall sein. Ein Beispiel soll die Nutzung von Listen als Auswahlgrundlage für Telefonumfragen illustrieren. Es ist denkbar, dass ein Möbelhaus über eine Liste seiner Kunden mit Namen und Telefonnummern verfügt, da bei jedem Kauf eines größeren Möbelstückes die entsprechenden Angaben registriert werden, um beispielsweise Liefertermine telefonisch ankündigen zu können. Wenn diese Firma nun eine Befragung über die Zufriedenheit bei Käufern größerer Möbel durchführen lassen will, kann es die genannten Angaben an ein Marktforschungsinstitut weitergeben. Das ist datenschutzrechtlich unbedenklich, wenn die Angaben ausschließlich für den Geschäftszweck der Kundenzufriedenheitsmessung verwendet werden. Rechtlich nicht zulässig ist dagegen die Kopplung mit Werbekampagnen und dergleichen. Es wird nun zu entscheiden sein, wie lange die Registrierung (bzw. letzte Aktualisierung) der Angaben längstens zurückliegen darf. Diese Entscheidung betrifft allerdings nicht nur die Aktualität der Angaben (die Telefonnummern könnten sich inzwischen geändert haben), sondern auch inhaltliche Erwägungen wie z.B. das Rückerinnerungsvermögen eines Kunden an den Möbelkauf. So ist es z.B. denkbar, dass lediglich alle Angaben aus den letzten beiden Jahren berücksichtigt werden sollen. Die Auswahlgesamtheit bestünde dann aus denjenigen Kunden, die in den letzten beiden Jahren mindestens ein größeres Möbelstück in diesem Möbelhaus erworben und Angaben über ihre Telefonnummer hinterlassen haben. Wenn in der Liste zusätzlich erfasst wurde, welchen Preis

die erworbenen Möbel hatten und/oder wie häufig eine Person Möbel gekauft hat, wären diese Daten als Schichtungsmerkmale verwendbar. Aus den einzelnen, durch die Schichtung gebildeten Zellen würden dann zufällig Telefonnummern gezogen¹ und die entsprechenden Kunden angerufen. Für die zufällige Auswahl von Datensätzen, d.h. Namen mit Telefonnummern von Kunden, können z.B. die folgenden Programme verwendet werden:

- SPSS: Nach dem Einlesen der Datei und der Überprüfung auf Eindeutigkeit, d.h. nur einmaliges Vorkommen jedes Datensatzes, können die folgenden Optionen verwendet werden:

DATA → SELECT CASES → RANDOM SAMPLE OF CASES
→ APPROXIMATELY ■ % OF ALL CASES oder
→ EXACTLY ■ CASES FROM THE FIRST ■ CASES

- STATISTICA und EXCEL: Bei diesen beiden Programmpaketen existieren Funktionen zur Erzeugung von gleichverteilten Zufallszahlen. Diese sollten in einem extra Feld der Datei erzeugt und die Datensätze dann der Größe dieser Zufallszahlen nach geordnet werden. Anschließend können die ersten x Datensätze (Kunden) entnommen werden.

Liegen zusätzlich zu Namen und Telefonnummern auch die Adressen der Kunden vor, ist es möglich, die ausgewählten Kunden durch ein Ankündigungsschreiben auf die geplante Studie vorzubereiten. Inwieweit derartige Vorkontakte vor dem eigentlichen Telefoninterview zu einer erhöhten Antwortbereitschaft beitragen, ist allerdings noch nicht eindeutig geklärt (vgl. Hüfken 2000). Ebenso ungeklärt ist bislang, ob es günstig ist, den zu Interviewenden bei Vorliegen ihrer Adressen den Fragebogen im Vorhinein zuzusenden. Theoretisch bietet sich dies insbesondere bei komplizierteren Fragestellungen an, klare Befunde für eine Erhöhung der Responserate liegen indes nicht vor (vgl. Friedrichs 2000).

3.3 Keine Liste vorhanden

3.3.1 Abschätzung des Umfangs der Grundgesamtheit

Weitaus häufiger tritt der Fall ein, dass für die Grundgesamtheit einer geplanten Studie keine Liste vorliegt, auf der alle potentiellen Zielpersonen mit Namen und Telefonnummern verzeichnet sind. Anders als bei der Stichprobenziehung aus Listen (wo die Grundgesamtheit die Gesamtzahl der eindeutigen Listeneinträge umfasst) muss dann zunächst der Umfang der Grundgesamtheit abgeschätzt werden, um vorhersagen zu können, wie viele Haushalte voraussichtlich kontaktiert werden müssen, um überhaupt eine bestimmte Anzahl von Zielpersonen zu erreichen. Für das erste der obigen drei Beispiele für Grundgesamtheiten lässt sich der Umfang der Population aus Angaben des Statistischen Bundesamtes relativ einfach abschätzen. Zur Bundestagswahl 1998 waren z.B. 60.762.751 Bürger wahlberechtigt, der in der Studie interessierende Anteil an der Gesamtbevölkerung ist also „groß“. Außer in Haushalten, in denen ausschließlich nicht wahlberechtigte Ausländer oder Personen unter 18 Jahren wohnen, wird man in jedem Privathaushalt potentiell wenigstens eine Zielperson ermitteln können. Weitaus schwieriger ist die Ermittlung des Umfangs der Gesamtheiten in den beiden anderen Studien. Für das zweite Beispiel benötigt man eine Zählung der Bevölkerung nach dem Alter in den kreisfreien Städten des Regierungsbezirks Düsseldorf, um den Anteil der Personen, die 60 Jahre oder älter sind, ermitteln zu können. Diese könnten z.B. aus den Einwohnermelderegistern der betreffenden Gemeinden beschafft werden. Da derartige Zählungen aber nicht für Haushalte verfügbar sind, müssen weniger exakte

¹ Auf das Problem des notwendigen Stichprobenumfangs wird weiter unten eingegangen.

Hochrechnungen vorgenommen werden, z.B. mit Hilfe des Mikrozensus, der eine Schätzung der durchschnittlichen Zahl der Personen in Haushalten dieser Altersgruppe ermöglicht. Auf Grundlage des Mikrozensus wurde auch für die dritte Beispielstudie ermittelt, dass der Anteil von Haushalten mit Kindern unter 9 Jahren in Westdeutschland etwa 13% beträgt, in Ostdeutschland ca. 11%. Wenn die interessierenden Segmente in der Gesamtbevölkerung derart klein sind, ist die Erhebung mit einem wesentlich höheren zeitlichen und finanziellen Aufwand verbunden, denn die zur Grundgesamtheit gehörenden Haushalte (in denen wenigstens eine Zielperson lebt) müssen erst durch ein Screening herausgesucht werden. Dies muss bei der Planung der Studie unbedingt berücksichtigt werden.

1.3.2 Das Gabler-Häder-Design für allgemeine Bevölkerungsbefragungen

Um sowohl eingetragene als auch nicht eingetragene Telefonnummern in die Stichprobe einzuschließen, wäre es zunächst denkbar, das aus den USA bekannte Random Digit Dialling – die einfache Zufallsziffernanwahl – anzuwenden. Dazu müsste für alle 5200 (oder auch nur interessierende ausgewählte) Ortsnetzbereiche in Deutschland jeweils der Range zwischen der niedrigsten und der höchsten eingetragenen Telefonnummer bestimmt werden. Dieser würde die Menge der als Telefonnummern denkbaren Ziffernfolgen umfassen, aus denen die Stichprobe zu ziehen wäre. Dazu ein Beispiel: Für den Ortsnetzbereich „06321“ ist die niedrigste Nummer eines privaten Haushalts 2001, die höchste 6790152. Insgesamt sind für diesen Ortsnetzbereich gegenwärtig etwa 21.000 Nummern im Telefonbuch veröffentlicht. Wenn man nun eine Stichprobe aus dem bestehenden Range zöge, dann würde man nur in 0.3% der Fälle auf eine eingetragene Nummer treffen. Selbst wenn ein Anteil von nicht eingetragenen Nummern von 20% (gegenwärtiger Bundesdurchschnitt) unterstellt wird, läge die Trefferquote („Hitrate“) unter einem Prozent. Damit dürfte klar sein: Dieser Ansatz ist in Deutschland aufgrund des komplizierten Nummernsystems nicht praktikabel, da entschieden zu zeit- und kostenintensiv.

Eine andere, ebenfalls in den USA entwickelte Lösung des Problems der Einbeziehung von Nummern auch nicht eingetragener Telefonteilnehmer ist das sogenannte „Randomize Last Digits“. Hierbei werden Nummern aus dem Telefonbuch gezogen und anschließend ihre letzten beiden Stellen durch zufällig erzeugte Ziffern ersetzt. Bei diesem Vorgehen sind allerdings die Inklusionswahrscheinlichkeiten, d.h. die Chance für die Telefonnummern, in die Stichprobe zu gelangen, nicht gleich. Sie hängen davon ab, wie viele weitere Nummern im gleichen 100er Block eingetragen sind. Ein 100er Block ist dabei definiert als der Stamm einer Telefonnummer, der nach dem Abschneiden der beiden letzten Stellen verbleibt. Er umfasst die Menge aller unterschiedlichen Ziffernfolgen, die sich durch die zufällige Ersetzung dieser letzten beiden Ziffern generieren lassen – also genau 100. So liegt z.B. die Telefonnummer 5129815 im Block 51298xx, der alle Ziffernfolgen von 5129800 bis 5129899 beinhaltet. Um die ungleichen Inklusionswahrscheinlichkeiten beim Randomize Last Digits wieder auszugleichen, müssten die in den Interviews erhaltenen Datensätze im nachhinein jeweils mit der Inversen der Blockbesetzung, d.h. der Inversen der Zahl eingetragener Nummern in diesem Block, gewichtet werden, was äußerst aufwendig wäre. So würde ein Datensatz, der aus einem Interview mit einem Teilnehmer stammt, in dessen Block noch 40 weitere eingetragene Nummer stehen, mit dem Faktor 1/40 multipliziert werden. Ein anderer Teilnehmer, in dessen Block nur zwei eingetragene Nummern vorkommen, erhielte dagegen den Faktor $\frac{1}{2}$.

Am ZUMA wurde Mitte der 90er Jahre ein Ansatz entwickelt, der eingetragenen und nicht eingetragenen Telefonnummern die gleiche Chance gibt, in die Stichprobe zu gelangen (vgl. Häder/Gabler 1998). Dieser soll im folgenden kurz vorgestellt werden:

- Für jeden Ortsnetzbereich ist es möglich, genau die 100er Blocks festzustellen, in denen wenigstens eine eingetragene Telefonnummer vorkommt.
- Zwischen diesen 100er Blocks existieren häufig große Lücken. Diese sind der wesentliche Grund für die nur geringe Effizienz bei der Anwendung von reinen Zufalls-Ziffern-Anwahlen (Random Digit Dialling). Wir gehen davon aus, dass in diesen Lücken keine nicht eingetragenen Nummern vorkommen.
- Für jeden der mit mindestens einer eingetragenen Nummer besetzten 100er Blocks werden anschließend alle möglichen Ziffernfolgen generiert (bei K belegten Blocks also insgesamt $K \cdot 100$ Ziffernfolgen).
- Aus dieser Menge wird schließlich (z.B. uneingeschränkt zufällig oder geschichtet) eine vorher bestimmte Zahl von Ziffernfolgen gezogen.

1.3.3 ZUMA-Service: Bereitstellung von Telefonstichproben nach dem Gabler-Häder-Design

Dieses Stichprobendesign hat sich in den letzten Jahren als methodischer Standard in der empirischen Sozialforschung in Deutschland etabliert. Da sich aber nicht jeder Wissenschaftler an einer Universität oder anderen öffentlichen Forschungseinrichtung selbst den notwendigen Auswahlrahmen für bundesweite oder auch regional begrenzte Bevölkerungsbefragungen herstellen kann – dies ist eine sehr zeitaufwendige Arbeit – stellt ZUMA als Serviceleistung diesem Personenkreis bei Bedarf Telefonstichproben zur Verfügung. Die entsprechenden ASCII-Dateien sind wie folgt aufgebaut:

Studienkennung	laufende Nummer	Vorwahl	Rufnummer	Eintrag	Blockbesetzung
ZUMA
ZUMA	47	04202	2630	0	71
ZUMA	48	0621	13949	1	51
ZUMA	49	0731	29752	0	3
ZUMA

In der ersten Spalte steht eine vierstellige Studienkennung, in der zweiten Spalte folgt die laufende Nummer des Datensatzes von 1 bis n (Stichprobenumfang). Die dritte Spalte beinhaltet die Vorwahlnummer. In der vierten Spalte folgt die von uns generierte Ziffernfolge, die anzuwählen ist. Sie kann eine eingetragene Telefonnummer repräsentieren (dann hat das Merkmal „Eintrag“ in Spalte 5 die Ausprägung „1“), die Rufnummer eines nicht eingetragenen Teilnehmers oder eine nicht vergebene Rufnummer sein. In den letzten beiden Fällen steht beim Merkmal „Eintrag“ eine 0. In der sechsten Spalte folgt schließlich eine technische Information: Sie beinhaltet die Zahl eingetragener Rufnummern im Block der generierten Ziffernfolge. Die Angaben aus den Spalten 5 und 6 können für die Kontrolle der Güte der Bruttostichprobe herangezogen werden. Da die mittlere Blockdichte eingetragener Rufnummern in Deutschland in unserem gegenwärtig verwendeten Auswahlrahmen 43 beträgt, sollten bei bundesweiten Stichproben etwa 43% der Ziffernfolgen die Ausprägung

„1“ in Spalte 5 haben; der Mittelwert in Spalte 6 sollte ebenfalls in der Nähe des Wertes 43 liegen (einen hinreichend großen Stichprobenumfang vorausgesetzt).

Erfahrungen aus der Umsetzung des Stichprobenplanes haben ergeben, dass als Umfang der Bruttostichprobe das Fünffache der Zahl der zu realisierenden Interviews angesetzt werden sollte. Damit werden sowohl stichprobenneutrale Ausfälle, wie z.B. Ziffernfolgen, die keine Rufnummern darstellen oder die Geschäftsanschlüsse repräsentieren, als auch Verweigerungen und andere nicht stichprobenneutrale Ausfälle (bei einer angenommenen Responserate von ca. 50%) ausgeglichen.

Telefonstichproben können für bundesweite Umfragen, Umfragen in einzelnen Vorwahlbereichen und ausgewählten Gemeinden bereitgestellt werden. Bei letzteren ist allerdings zu berücksichtigen, dass Vorwahlbereichs- und Gemeindegrenzen nicht kongruent sind und die Telefonnummern innerhalb eines Vorwahlbereiches auch nicht nach Gemeinden vergeben werden. Deshalb ist es möglich, dass generierte Nummern außerhalb einer Zielgemeinde liegen. Soll z.B. eine Befragung in der Gemeinde Neustadt an der Weinstraße durchgeführt werden, so müssen die Vorwahlbereiche 06321 und 06327 berücksichtigt werden, da beide Teile von Neustadt überdecken. Beide Bereiche gelten aber auch für andere Gemeinden (06321 auch Kirrweiler und Maikammer; 06327 auch Altdorf, Böbingen und Gommersheim). In einer Frage zu Beginn des Interviews müsste deshalb erkundet werden, ob der betreffende Anschluss in der Gemeinde Neustadt an der Weinstraße liegt, andernfalls wäre das Interview abubrechen.

4. Tipps für den Einsatz von Telefonstichproben

Unabhängig davon, auf welche Weise die konkrete Stichprobe für eine Telefonstudie konstruiert worden ist, können verschiedene allgemeingültige Tipps für deren Umsetzung gegeben werden.

- Telefonstichproben sollten vor Beginn einer Studie „gemischt“ werden, wenn sie zunächst z.B. nach der Größe der Vorwahl- und/oder Teilnehmernummer, nach Regionen oder nach Unternehmensbereichen geordnet vorliegen. Es empfiehlt sich dann, die Ziffernfolgen „päckchenweise“, d.h. in Sätzen zu z.B. 100 Nummern abzuarbeiten. Damit wird vermieden, dass zunächst alle Nummern einmal angerufen werden und viele Interviews schon beim ersten Kontakt zustande kommen. Schwerer erreichbare Zielpersonen hätten dann eine geringere Chance in die Stichprobe zu gelangen und diese wäre damit hinsichtlich des Merkmals „Erreichbarkeit“ verzerrt. Deshalb sollte also ein Ziffernfolgensatz zunächst vollständig abgearbeitet, d.h. bei Nichtkontakt die maximale Zahl der Kontaktversuche absolviert werden, bevor ein neuer Satz begonnen wird. Die „Mischung“ vor Beginn der Erhebung ist insbesondere deshalb wichtig, da mitunter nicht alle „Päckchen“ der Bruttostichprobe für das Erzielen einer vorgegebenen Zahl von Interviews begonnen werden müssen.
- Es ist ratsam, jede Ziffernfolge mindestens zehn Kontaktversuchen an unterschiedlichen Wochentagen und zu verschiedenen Tageszeiten zu unterziehen. In den Richtlinien des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute (nachzulesen unter <http://www.adm-ev.de/homepage.html>) wird davon ausgegangen, dass keine Verletzung der Privatsphäre vorliegt, wenn Kontaktversuche in der Zeit von 9.00 Uhr bis 21.00 Uhr vorgenommen werden.
- Wenn die Zielpersonen nicht von vornherein feststehen, weil sie z.B. auf einer Liste verzeichnet sind, müssen diese innerhalb des kontaktierten Haushaltes zufällig ausgewählt werden. Dafür sind verschiedene Vorgehensweisen üblich (vgl. Lavrakas 1993). Als

praktikabel hat sich insbesondere die „Last-Birthday-Methode“ erwiesen. Bei diesem Auswahlmechanismus wird die Kontaktperson im Haushalt um Auskunft darüber gebeten, wer von den zur Grundgesamtheit gehörenden Personen zuletzt Geburtstag hatte. In einer allgemeinen Bevölkerungsbefragung könnte die entsprechende Frage z.B. lauten: „Wer von den Personen, die in ihrem Haushalt leben und 18 Jahre oder älter ist, hatte als letzter Geburtstag?“ Mit der genannten Person würde dann das Interview geführt bzw. zunächst ein Termin für das Interview vereinbart. (Diese Auswahl ist natürlich nicht nötig, wenn Gegenstand der Studie Informationen über den Haushalt sind, die jedes Haushaltsmitglied gleichermaßen geben kann.)

- Wenn die zu befragende Zielperson über ein Auswahlverfahren im Haushalt ermittelt werden muss, sind die Daten für designbasierte Schätzungen einer Gewichtung zu unterziehen. Dazu muss der Datensatz mit einem Faktor versehen werden, der der Inversen der Auswahlwahrscheinlichkeit im Haushalt entspricht (bei zwei Personen im Haushalt, die zur Grundgesamtheit gehören, ist dieser Faktor z.B. 0,5). Dieses Verfahren ist allerdings bei nicht vollständig umgesetzten Stichprobenplänen umstritten (vgl. Rothe 1994).
- Bei jeder Studie ist die sorgfältige Dokumentation der Ausschöpfung von Bedeutung für die Einschätzung der Qualität der Stichprobe. Ein für alle Umfragen verwendbares Schema für die Berechnung der Ausschöpfungsquote gibt es allerdings nicht, da die Wahl sinnvoller Kategorien von der jeweiligen Definition der Grundgesamtheit abhängt. Ein Beispiel für eine bundesweite Befragung der deutschsprachigen Bevölkerung ab 18 Jahren soll aber zeigen, welche Kategorien von Ausfällen vermerkt werden könnten (vgl. Tabelle 1):
 - Die Kategorie „Kein Anschluss unter dieser Nummer“ wird vorrangig in Studien auftreten, bei denen generierte Ziffernfolgen verwendet werden, bei listenbasierten Umfragen ist nur in Ausnahmefällen (wenn die Telefonnummer nicht mehr aktuell ist) mit dieser Ansage zu rechnen.
 - Problematisch ist insbesondere die Zuordnung der Kategorie „nach zehn Kontaktversuchen mit Freizeichen kein Kontakt“, da bei zufallsgenerierten Stichproben nicht eindeutig unterschieden werden kann, ob es sich um nicht vergebene Rufnummern oder nicht erreichte Haushalte handelt (nicht in jedem Fall erfolgt eine automatische Ansage, wenn eine Ziffernfolge angewählt wird, die keine Rufnummer ist): Nicht vergebene Rufnummern sind den neutralen Ausfällen zuzuordnen, nicht erreichte Teilnehmer dagegen nicht. Letztlich ist dieses Problem nur durch eine ausreichend lange Feldzeit und viele Kontaktversuche lösbar (im hier dargestellten Beispiel dauerte die Feldzeit nur sieben Tage, was offensichtlich zu kurz war).
 - Auch die Kategorie „Geschäftsadresse“ wird bei listenbasierten Auswahlen kaum auftreten (wenn nicht Geschäfte befragt werden sollen, aber dann wären die entsprechenden Nummern natürlich nicht als neutrale Ausfälle zu behandeln).

Diese Beispiele mögen genügen, um *Tabelle 1: Ausfallstatistik des bundesweiten ZUMA-Stichprobenexperiments (April 1999)*

Ausfallgründe	absolut	in %
Bruttostichprobe I	1.500	100
Stichprobenneutrale Ausfälle		
• Ansage "Kein Anschluss unter dieser Nummer"	399	26.6
• Ansage "Rufnummer geändert"	4	0.3
• nach zehn Kontaktversuchen mit Freizeichen kein Kontakt	218	14.5
• bei zehn Kontaktversuchen besetzt	33	2.2
• ISDN-Modem-Karte	6	0.4
• Geschäftsadresse	78	5.2
• Anrufbeantworter	73	4.9
Σ	811	54.1
Bruttostichprobe II	689	100.0
Kontakt mit Teilnehmer, aber...		
• Teilnehmer hat keine Zeit, kein Interesse	78	11.3
• Teilnehmer hat Sprachprobleme	11	1.6
• Teilnehmer legt wortlos auf	34	4.9
• Teilnehmer verweigert Zugang zur Zielperson	3	0.4
• anderer Ausfallgrund bei Teilnehmer	15	2.1
Kontakt mit Zielperson, aber...		
• Zielperson hat keine Zeit	17	2.5
• Zielperson hat kein Interesse	150	21.8
• Zielperson zu alt/ krank	28	4.0
• Zielperson ist in der Feldzeit nicht da	24	3.5
• anderer Ausfallgrund bei Zielperson	35	5.1
• Interview abgebrochen	8	1.1
Realisiert	286	41.5

Diese Beispiele mögen genügen, um zu zeigen, dass die Ausschöpfungsberechnung an die konkrete Studie angepasst werden muss. Wichtig ist letztlich, dass eine sinnvolle, mit anderen Studien zum selben Thema vergleichbare Responstrate berechnet werden kann.

5. Ermittlung des notwendigen Stichprobenumfangs

Es ist recht schwierig, die Frage nach dem notwendigen Stichprobenumfang für eine Studie zu behandeln, weil dieser häufig durch beschränkte Budgets für die Erhebung und nicht durch statistische Formeln festgelegt wird.

Statistisch formal stellt sich die Situation bei uneingeschränkter Zufallsauswahl (Simple Random Sampling), wie sie z.B. mit dem Gabler-Häder-Design möglich ist, folgendermaßen dar: Als Parameter in die Berechnung des Stichprobenumfangs gehen der tolerierbare Stichprobenfehler um den Punktschätzer, die Irrtumswahrscheinlichkeit, der Umfang der Grundgesamtheit und Vorkenntnisse über die Streuung der interessierenden Variablen ein. Soll z.B. der Anteilswert in einer „großen“ Gesamtheit (d.h. einer Gesamtheit mit mehr als 100.000 Elementen) bei einer Irrtumswahrscheinlichkeit von 5% um nicht mehr als ± 3 Prozentpunkte vom Punktschätzer abweichen und ist nichts über die Streuung der interessierenden Merkmale bekannt, ist die folgende Berechnungsvorschrift anwendbar:

Es sei

n_{srs}	Stichprobenumfang unter SRS
N	Umfang der Grundgesamtheit
$z_{\alpha/2}$	tabellierter Wert aus der Standardnormalverteilung, für $\alpha = 0.05$ ist $z_{\alpha/2} = 1.96$
p	Anteil des interessierenden Merkmals in der Stichprobe, entweder aus Voruntersuchung bekannt oder im ungünstigsten Fall $p=0.5$
e	zulässiger absoluter Stichprobenfehler, $2e$ entspricht Länge des Konfidenzintervalls, im obigen Beispiel ist $e=0.03$

$$n_{srs} \geq \left(\frac{z_{\alpha/2}}{e} \right)^2 \cdot p (1-p)$$

$$n_{srs} \geq \left(\frac{1.96}{0.03} \right)^2 \cdot 0.5 (1-0.5)$$

$$n_{srs} \geq 1067.1$$

Im angegebenen Beispiel müssten also mindestens 1068 Personen befragt werden. Dieser Stichprobenumfang ist für viele Bevölkerungsumfragen typisch. Zu beachten ist allerdings, dass bei der Auswertung von Subpopulationen, z.B. nur der weiblichen oder nur der 18 bis 25jährigen Befragten, die Konfidenzintervalle i.d.R. deutlich größer sind.

Bei kleineren Grundgesamtheiten (also etwa Gesamtheiten mit 100.000 oder weniger Elementen) muss der Korrekturfaktor $1 - \frac{n}{N}$ für das Ziehen ohne Zurücklegen wie folgt berücksichtigt werden:

$$n_{srs} \geq \frac{N \cdot z_{\alpha/2}^2 \cdot p (1-p)}{z_{\alpha/2}^2 \cdot p (1-p) + N \cdot e^2}$$

In der folgenden Tabelle sind notwendige Stichprobenumfänge für verschiedene „kleine“ Gesamtheiten, z.B. die Mitarbeiter einer Firma oder die Besucher eines Fußballspiels, dargestellt. Es zeigt sich, dass bei relativ kleinen Gesamtheiten, z.B. $N=300$, sehr viele Elemente in die Befragung einbezogen werden müssen, um hinreichend genaue Ergebnisse zu

Tabelle 2: Minimaler Stichprobenumfang n für gegebenen absoluten Stichprobenfehler e bei Irrtumswahrscheinlichkeit $\alpha = 0.05$ für Anteile $p = 0.5$ und $p = 0.8$ (oder $p = 0.2$) (nach Borg 2000, S. 144)

p = 0.5			p = 0.8 oder p = 0.2		
N	e = 0.03	e = 0.05	N	e = 0.03	e = 0.05
200	168	132	200	155	110
300	235	168	300	208	135
400	291	196	400	252	152
500	340	217	500	289	165
750	440	254	750	357	185
1000	516	278	1000	406	197
3000	787	341	3000	556	227
7500	934	365	7500	626	238
10000	964	370	10000	639	240
50000	1045	381	50000	674	245
100000	1056	383	100000	678	245

erhalten. Da in vielen Umfragen mit einem bestimmten Anteil von Nichtteilnehmern gerechnet werden muss, empfehlen sich hier deshalb eher Totalerhebungen.

Zu beachten ist weiterhin, dass bei komplexeren Stichprobenplänen im Fall designbasierter Schätzungen sogenannte Design-Effekte durch Klumpung und Gewichtung (Gabler/Häder 2000) auftreten, die die Konfidenzintervalle zusätzlich vergrößern.

Da leider in vielen Fällen finanzielle Restriktionen keine Stichprobenumfänge erlauben, wie sie insbesondere für designbasierte Schätzungen in Subpopulationen notwendig wären, kann hier nur der Rat gegeben werden, die Ergebnisse mit der gebotenen Vorsicht zu interpretieren. Für eine ausführlichere Darstellung der Problematik vergleiche Särndal/Swensson/Wretmann (1992).

Literatur

- Borg, I. (2000): Führungsinstrument Mitarbeiterbefragung. Theorien, Tools und Praxiserfahrungen. Göttingen: Verlag für Angewandte Psychologie.
- Friedrichs, J. (2000): Effekte des Versands des Fragebogens auf die Antwortqualität bei einer telefonischen Befragung. In: V. Hüfken (Hrsg.), Methoden in Telefonumfragen. Opladen: Westdeutscher Verlag.
- Gabler, S./ Häder, S., 2000: Über Design-Effekte. In: P. Mohler/P. Lüttinger (Hrsg.), Querschnitt. Festschrift zum 65. Geburtstag von Max Kaase. Mannheim: ZUMA.
- Häder, S./ Gabler, S. (1998): Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. In: S. Gabler, S./Häder, S./Hoffmeyer-Zlotnik, J.H.P. (Hrsg.), Telefonstichproben in Deutschland. Opladen: Westdeutscher Verlag.
- Hüfken, V. (2000): Kontaktierung bei Telefonumfragen. Auswirkungen auf das Kooperations- und Antwortverhalten. In: V. Hüfken (Hrsg.), Methoden in Telefonumfragen. Opladen: Westdeutscher Verlag.
- Humpert, A./ Schneiderheinze, K. (2000): Stichprobenziehung für telefonische Zuwandererumfragen – Einsatzmöglichkeiten der Namenforschung (Onomastik). In: ZUMA-Nachrichten 47.
- von dem Knesebeck, O./ Lüschen, G. (2000): Zum Problem von Verweigerungen in telefonischen Befragungen alter Menschen. In: V. Hüfken (Hrsg.), Methoden in Telefonumfragen. Opladen: Westdeutscher Verlag.
- Lavrakas, P. (1993): Telephone Survey Methods. Newbury Park: Sage.
- Porst, R. (2000): Praxis der Umfrageforschung. Wiesbaden: Teubner.
- Rothe, G. (1994) Wie (un)wichtig sind Gewichtung? Eine Untersuchung am ALLBUS 1986. In: S. Gabler/ D. Krebs/ J. Hoffmeyer-Zlotnik (hrsg.), Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag.
- Särndal, C.-E./ Swensson, B./Wretmann, J. (1992): Model Assisted Survey Sampling. New York: Springer.
- Statistisches Bundesamt (1999): Mitteilung für die Presse vom 20. Dezember 1999. <http://www.statistik-bund.de/presse/deutsch/pm/p9435024.htm>.