

SDM Survey Guidelines

Stichproben in der Praxis

Sabine Häder

Januar 2015, Version 1.0

Zusammenfassung

In diesem Kapitel widmen wir uns der stichprobenpraktischen Seite von Erhebungen. Es wird gezeigt, wie insbesondere für telefonische, aber auch schriftliche, persönlich-mündliche und Online-Befragungen jeweils die Auswahlrahmen und Stichprobenstrategien bestimmt werden. Schließlich werden Stichproben für Cross-Cultural-Surveys behandelt. Hierbei dient der European Social Survey als Beispiel.

Zitierung

Häder, Sabine (2015). Stichproben in der Praxis. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (SDM Survey Guidelines). DOI: 10.15465/sdm-sg_014

1. Worum geht es?

Stichprobenerhebungen für Bevölkerungsumfragen oder auch Befragungen spezieller Populationen können die unterschiedlichsten Gestalten annehmen. Sie sind z.B. denkbar als Telefonumfragen, als persönlich-mündliche Erhebungen, als schriftlich-postalische Umfragen oder als Online-Befragungen. Wesentlich für die Auswahl des Befragungsmodus sind häufig Zeit- und Kostenrestriktionen, aber auch spezifische Eigenschaften der Befragung, wie z.B. der Anteil sensibler Fragen oder die Komplexität des Erhebungsinstruments (siehe Kapitel A2).

Für welchen Befragungsmode die Entscheidung auch gefallen ist, am Anfang der stichprobenpraktischen Überlegungen steht immer die Frage nach einem geeigneten Auswahlrahmen. Weiterhin wird geklärt, ob im Auswahlrahmen ergänzende Merkmale für eine Schichtung zur Verfügung stehen. Schließlich wird das spezielle Auswahlverfahren festgelegt.

Konkrete Anwendungsbeispiele für Stichprobenerhebungen in den unterschiedlichen Befragungsmodi finden sich im folgenden Kapitel.

2. Wie funktionieren Telefonstichproben?

Telefonumfragen waren von 1998 bis 2008 der am häufigsten eingesetzte Befragungsmodus in Deutschland. Ein adäquater Auswahlrahmen für die Ziehung von Festnetznummern war Mitte der 1990er Jahre bei GESIS entwickelt worden (Gabler-Häder-Design). Es hat sich allerdings eine Tendenz angedeutet, die die alleinige Nutzung dieses Frames als unzureichend zur Abdeckung der Gesamtheit der Privathaushalte erscheinen lässt: Ein wachsender Anteil der Haushalte ist lediglich über das Mobiltelefon erreichbar. Diese Haushalte haben bei Festnetzbefragungen keine positive Auswahlchance. Damit kann es zu systematischen Verzerrungen in den Stichproben kommen, da sich Festnetzhaushalte und Mobilfunkhaushalte hinsichtlich für die Sozialforschung relevanter Merkmale unterscheiden. Um die ohnehin in Telefonumfragen schon häufig unterrepräsentierte junge, mobile Bevölkerungsgruppe mit formal niedrigem Bildungsabschluss nicht noch weiter auszuschließen, ergab sich die Notwendigkeit zu Überlegungen hinsichtlich der Integration von Mobilfunkanschlüssen in Telefonstichproben. Ergebnisse zu Forschungen zur Stichprobenstrategie, zur Antwortqualität und zur Teilnahmebereitschaft bei Mobilfunkbefragungen sollen vorgestellt werden.

2.1 Funktioniert das Telefonbuch als Auswahlrahmen für allgemeine Bevölkerungsbefragungen im Festnetz?

Noch bis 1992 war das Telefonbuch in Deutschland ein akzeptabler Auswahlrahmen für Festnetz-Telefonstichproben (in Westdeutschland), da die Telefonteilnehmer bis zu diesem Zeitpunkt mit nur sehr wenigen Ausnahmen verpflichtet waren, ihre Telefonnummer zu veröffentlichen. Inzwischen besteht diese Eintragspflicht nicht mehr und der Anteil sogenannter „Nonpubs“, d.h. nicht eingetragener Telefonteilnehmer, ist in den letzten Jahren stark angestiegen. Inzwischen ist etwa die Hälfte der vergebenen Festnetz-Rufnummern an private Haushalte nicht mehr im Telefonbuch eingetragen.

Tabelle 1: Verhältnis vergebene vs. eingetragene Festnetznummern in Deutschland 2009 (Quelle: Heckel 2011)

| HH 2009 in Mio. | Private Rufnummern-einträge in Mio. | Vergebene private Rufnummern in Mio. | Anteil Einträge an Privat-haushalten | Anteil Einträge an vergebenen Rufnummern |
|-----------------|-------------------------------------|--------------------------------------|--------------------------------------|--|
| 39,331 | 22,658 | 536,221 | 57,6 % | 4,2 % |

Es gilt dabei zu beachten, dass der Anteil der Nonpubs in urbanen Gebieten, insbesondere Großstädten, wesentlich höher sein kann. In ländlichen Gebieten insbesondere in Westdeutschland ist der Anteil der ins Telefonbuch eingetragenen Festnetzanschlüsse dagegen höher.

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 größerer 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 2002) oder die Auswahl von Unternehmen oder Institutionen sowie für die Ziehung von Stichproben für einzelne westdeutsche ländliche Regionen ist es weiterhin wertvoll.

2.2 Wie funktioniert das Gabler-Häder-Design zur Auswahl von Festnetzstichproben?

Dieser Auswahlrahmen wird angewendet, wenn keine vollständige Liste mit Telefonnummern vorliegt.

Um sowohl eingetragene als auch nicht eingetragene Telefonnummern in die Stichprobe einzuschließen wäre es zunächst denkbar, das aus den USA bekannte Random Digit Dialing – die einfache Zufallsziffernanwahl – anzuwenden. Dazu müsste für alle 5,200 (oder auch nur interessierende ausgewählte) Ortsnetzbereiche in Deutschland jeweils die Spannweite zwischen der niedrigsten und der höchsten eingetragenen Telefonnummer bestimmt werden. Diese würde die Menge der als Telefonnummern denkbaren Ziffernfolgen umfassen, aus denen die Stichprobe zu ziehen wäre. Dazu ein Beispiel: In einer Telefon_CD vom Januar 2014 ist für den Ortsnetzbereich „06321“ die niedrigste Nummer eines privaten Haushalts 2001, die höchste 95600100. Insgesamt sind für diesen Ortsnetzbereich gegenwärtig 12868 private Nummern im Telefonbuch veröffentlicht. Wir unterstellen nun, dass es in diesem Vorwahlbereich ebenso viele nichteingetragene Nummern gibt. Wenn man eine Stichprobe aus der gesamten Spannweite zöge, dann würde man nur in 0,027% der Fälle auf eine geschaltete Nummer treffen. Damit dürfte klar sein: Dieser Ansatz ist in Deutschland aufgrund des komplizierten Nummernsystems nicht praktikabel, da er entschieden zu zeit- und kostenintensiv ist.

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 aufwändig wäre. So würde ein Datensatz, der aus einem Interview mit einem Teilnehmer stammt, in dessen Block insgesamt 40 eingetragene Nummern stehen, mit dem Faktor 1/40 multipliziert werden. Ein anderer Teilnehmer, in dessen Block nur zwei eingetragene Nummern vorkommen, erhielte dagegen den Faktor 1/2.

Bei GESIS 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 in unserem Modell davon aus, dass in diesen Lücken keine nicht eingetragenen Nummern vorkommen. Dieses Modell war in der Vergangenheit hinreichend realitätskonform. In der letzten Zeit hat sich allerdings herausgestellt, dass insbesondere durch die Marktpräsenz zusätzlicher Provider neue Blöcke aktiviert werden, in denen keine Nummer im Telefonbuch eingetragen ist. Diese fehlen dann in unserem Auswahlrahmen. Hier sind verstärkt Bemühungen notwendig, Informationen von der Bundesnetzagentur einzubeziehen (Sand 2014a, S.13, Sand 2014b).

Für jeden der mit mindestens einer eingetragenen Nummer besetzten 100er Blocks werden 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.

Dieses Vorgehen ist mittlerweile in Deutschland der Standard für Telefonstichproben. Vom Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute ist es insofern weiterentwickelt worden, als auch regionale Schichtungen möglich sind (Heckel 2002, von der Heyde 2002).

2.3 Wie werden Mobilfunknummern bei Telefonstichproben behandelt?

Zwar können Befragungen über das Festnetz in Hinsicht auf die Stichprobenziehung und die Datenqualität als weitgehend methodisch erforscht angesehen werden. Diese sind aber zur Abdeckung der Grundgesamtheit der deutschen Bevölkerung allein nicht mehr ausreichend.

Mobiltelefone sind seit einiger Zeit europaweit verbreitet. 2010 verfügten 87% der Befragten des Eurobarometer über ein Handy (Bericht des Eurobarometer Spezial 356: 8). Auffällig ist dabei der Trend, dass Mobiltelefone in immer größeren Bevölkerungskreisen zum einzigen Telefonmittel werden, d.h. auf einen Festnetzanschluss im Haushalt verzichtet wird. Im Jahr 2010 waren 26% der Befragten in Europa „Mobile-onlys“. Da sich aber „Mobile-onlys“ und „Gemischtnutzer“ hinsichtlich soziodemographischer Merkmale unterscheiden, bedeutet der Verzicht auf die Gruppe der ausschließlichen Handynutzer einen Undercoverage, der zu einer systematischen Verzerrung der Schätzer führen kann.

In Deutschland ist die Marktpenetration mit Mobilfunkanschlüssen von 96% im Jahr 2005 auf 137% im 3. Quartal 2011 gestiegen, d.h. eine Vielzahl von Bürgern verfügten sogar über mehrere Handys bzw.

SIM-Karten. 83% der deutschen Bevölkerung hatten zu dieser Zeit (mindestens) einen Mobilfunkanschluss, 12,4% gehörten zur Gruppe der „Mobile-onlys“ (ADM 2012). Damit liegt Deutschland noch weit unter dem europäischen Mittel, was vermutlich dem Umstand geschuldet ist, dass für einen komfortablen Internetzugang immer noch Festnetzanschlüsse bevorzugt werden. Aber auch in Deutschland zeigen sich die Abweichungen von der übrigen Bevölkerung hinsichtlich soziodemographischer Merkmale: Sie sind häufiger jünger, eher männlich, stammen überproportional häufig aus Ostdeutschland und haben eher niedrigere Bildungsabschlüsse. So haben nur 16,8% der befragten „Mobile-onlys“ des TNS-Infratest F2F-Busses 2010 (n=30.000) Abitur, im Bundesdurchschnitt sind es 24,6%. 55,4% der Befragten mit ausschließlich Mobiltelefonen in diesem Surveyprogramm waren männlich, bundesweit sind es nur 48,4%.

Da mit einem weiteren Ansteigen der Zahl der „Mobile-onlys“ zu rechnen ist – das lehren internationale Erfahrungen – könnte man schlussfolgern, dass Telefonumfragen künftig nur über Mobilfunk erhoben werden sollten. Hier muss jedoch berücksichtigt werden, dass 18,6 % der Bevölkerung lediglich über das Festnetz erreichbar sind (ADM 2012). Dabei handelt es sich häufiger um ältere, auf dem Land in Westdeutschland lebende Personen. Damit bleibt als einzige Möglichkeit, um einen beachtlichen Undercoverage zu vermeiden, beide Telefonmodi parallel einzusetzen, d.h. einen Dual-Frame-Ansatz zu implementieren.

Von den Mobilfunknummern sind sogar nur etwa 20% im Telefonbuch eingetragen – und dies sind überwiegend Geschäftsanschlüsse. Hier wird der Auswahlrahmen so konstruiert, dass für alle gültigen Providereinwahlen (also z.B. 0171, 0165, 0177) alle sinnvollen Nummernfolgen generiert werden. Diese werden zusätzlich mit aktuellen Informationen aus dem Internet abgeglichen. Auf diese Weise kommt ein Auswahlrahmen mit ca. 212,19 Millionen Ziffernfolgen, d.h. tatsächlichen und potenziellen Anschlüssen, zustande.

2.4 Wie funktionieren Dual-Frame-Stichproben für Telefonumfragen?

Es wurde bereits argumentiert, dass bei Verwendung eines der beiden Auswahlrahmen (Festnetz oder Mobilfunk) nicht alle Elemente der Grundgesamtheit erfasst werden. Wenn ein zweiter Auswahlrahmen verfügbar ist, der u.a. einen Großteil der fehlenden Elemente (z.B. Mobile-onlys) enthält, sodass beide Frames zusammen (fast) die gesamte Population überdecken, sollten die Stichproben deshalb unabhängig aus beiden Auswahlgrundlagen gezogen werden. Die Schwierigkeit ist dabei lediglich, dass beide Frames einen Overlap haben. Dann stellt sich die Frage, wie mit diesem Overlap umzugehen ist, um annähernd unverzerrte Schätzer zu erhalten. Eine relativ elegante und einfache Möglichkeit ist, die Inklusionswahrscheinlichkeiten der Elemente der Grundgesamtheit zu ermitteln und den Horvitz-Thompson (HT)-Schätzer für die Schätzung des Totals einer interessierenden Variable einzusetzen.

Für eine allgemeine Formel der Inklusionswahrscheinlichkeiten (vgl. Gabler /Ayhan 2007) müssen zunächst die relevanten Parameter definiert werden (vgl. Übersicht 1):

Übersicht 1. Parameter, die für das Dual-Frame-Modell benötigt werden

| Festnetz | | Mobilfunk | |
|----------|--|-----------|--|
| M^F | Größe des Auswahlrahmens Festnetz | M^C | Größe des Auswahlrahmens Mobilfunk |
| m^F | Stichprobenumfang Festnetz | m^C | Stichprobenumfang Mobilfunk |
| k_i^F | Anzahl der Festnetznummern, unter denen Person i erreichbar ist | k_i^C | Anzahl der Mobilfunknummern, unter denen Person i erreichbar ist |
| z_i | Anzahl der Personen, die zur Grundgesamtheit gehören, im Haushalt der Person i | | |

Zur Vereinfachung der dann abzuleitenden Formeln soll noch folgende Grundannahme getroffen werden:

Die Wahrscheinlichkeit, dass zwei (nicht notwendig verschiedene) Mitglieder desselben Haushalts über verschiedene Wege ausgewählt werden, ist vernachlässigbar.

Diese Annahme kann lediglich bei kleinen regionalen Stichproben problematisch sein.

Wegen der Grundannahme vernachlässigen wir Stichproben, die einen Haushalt mit verschiedenen Festnetznummern enthalten. Weiter ist die Wahrscheinlichkeit, eine Person sowohl aus dem Festnetz- als auch dem Mobilfunkrahmen auszuwählen verschwindend klein, so dass wir diese ebenfalls vernachlässigen können.

$$\pi_i^{F \cap C} = \pi_i^F \pi_i^C \approx 0$$

Für die Inklusionswahrscheinlichkeit der Person i erhalten wir daher in guter Näherung

$$\pi_i \approx k_i^F \frac{m^F}{M^F} \cdot \frac{1}{z_i} + k_i^C \frac{m^C}{M^C}$$

Die Verwendung dieses Ansatzes hat sich in Deutschland durchgesetzt. Hierzu wurden in den von der Deutschen Forschungsgemeinschaft geförderten Studien CELLA1 und CELLA2 (GESIS und TU Dresden) Erfahrungen gesammelt.

Die Umfänge der Auswahlrahmen betragen gegenwärtig:

$$M^F = 158,88 \text{ Millionen}$$

$$M^C = 212,19 \text{ Millionen}$$

Für die Parameter m^C und m^F , sollte die für einen Survey tatsächlich gezogene Zahl von Nummern pro Mode eingesetzt werden, unabhängig davon, ob diese dann auch „abtelefoniert“ worden sind oder nicht. Weiterhin ist darauf zu achten, dass gleichartige Stichproben (jeweils mit oder ohne nicht geschaltete Anschlüsse) eingesetzt werden. Bei Festnetzstichproben können nicht geschaltete Anschlüsse über Predialer aussortiert werden. HLR Lookup (Sand 2014a) kann eingesetzt werden, um den Status von Mobilfunknummern zu bestimmen (Struminskaya 2009). Die anderen Parameter müssen im Interview erfragt werden (Gabler et al. 2013).

In der deutschen Markt- und Sozialforschung wird gerade diskutiert, welches Verhältnis der Stichprobenumfänge Festnetz:Mobilfunk empfehlenswert ist. Simulationen, die der ADM angestellt hat, deuten darauf hin, dass ein Verhältnis von 70:30 bei bundesweiten Studien mit hinreichend großem

Gesamtstichprobenumfang zu stabilen Schätzern für die Mobile-onlys führt und damit insgesamt ausreichend sein dürfte (vgl. ADM 2012).

2.5 Wie funktionieren Telefonstichproben, wenn ein Auswahlrahmen vorhanden ist?

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 gezogen und die entsprechenden Kunden angerufen.

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).

2.6 Welche Hinweise sollten beim Einsatz von Telefonstichproben beachtet werden?

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-, Provider- 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 acht 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 bei Festnetzbefragungen 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. Diese Auswahl ist natürlich nicht nötig, wenn Gegenstand der Studie Informationen über den Haushalt sind, die jedes Haushaltsmitglied gleichermaßen geben kann.
- Bei Telefonumfragen werden die Zielpersonen nicht mit gleichen Wahrscheinlichkeiten ausgewählt. Deshalb muss bei der Auswertung der Daten eine Designgewichtung vorgenommen werden. Dies gilt für Festnetzstichproben, Mobilfunkstichproben und Dual-Frame-Stichproben.
- Bei jeder Studie ist die sorgfältige Dokumentation der Ausschöpfung von Bedeutung für die Einschätzung der Qualität der Stichprobe. Diese sollte entsprechend der AAPOR-Standards (http://www.aapor.org/Response_Rates_An_Overview1.htm#.U0PvGhBI85Q) unter Zuhilfenahme des AAPOR Outcome Rate Calculators erfolgen.

3. Wie zieht man Stichproben für schriftliche Befragungen?

Voraussetzung für die Ziehung einer Stichprobe für eine schriftliche Befragung ist eine Liste mit Namen und Adressen der Personen, die zur Auswahlgesamtheit gehören. Aus dieser Liste kann dann eine vorher bestimmte Zahl n von Personen ausgewählt werden. Bei einer Liste, die rechnergestützt vorliegt, wird man das technisch realisieren, indem man jedem Datensatz (jeder Person) eine Zufallszahl zuordnet, die Datensätze dann nach der Größe dieser Zufallszahlen sortiert und die ersten n Datensätze entnimmt. Liegt die Liste dagegen nur in gedruckter Form vor, empfiehlt sich die Anwendung der systematischen Zufallsauswahl. Dabei wird das erste zu ziehende Stichprobenelement zufällig ermittelt als Zahl zwischen 1 und der Schrittweite. Die Schrittweite ergibt sich als Quotient aus dem Umfang der Auswahlgesamtheit und dem Stichprobenumfang. Ausgehend vom zufällig ermittelten Startpunkt wird dann das nächste Element jeweils durch Addition der Schrittweite ermittelt.

Nach diesem Prinzip wurde beispielsweise für ein Projekt, in dem die Geistesgeschichte des Tropischen Regenwalds erkundet werden sollte, die Stichprobe gezogen. 200 Erdkundelehrer im Oberschulamtsbezirk Stuttgart wurden aus dem Adressbuch des Philologenverbandes zufällig ausgewählt und angeschrieben.

In einer anderen Untersuchung sollte die Lärmbelastigung bei der EXPO 2000 für die Anwohner ermittelt werden. Hier wurde zunächst festgestellt, wie viele Anwohner im angezielten Alter in den

einzelnen, das Expo-Gelände begrenzenden Straßen wohnen. Dann fand eine Auswahl von zu Befragenden proportional zur Zahl der Straßenanwohner im entsprechenden Alter aus dem Einwohnermelderegister der Gemeinde Hannover statt. Die ausgewählten Personen wurden schriftlich befragt.

Schwieriger ist die Situation, wenn keine Liste vorhanden ist, auf der die Elemente der Auswahlgesamtheit notiert sind. Dann muss zu Hilfslösungen gegriffen werden, die den Nachteil haben, dass die Inklusionswahrscheinlichkeiten nicht genau bestimmt werden können bzw. die Response-Raten sehr gering sind.

Eine Möglichkeit ist es, ein Telefon-Screening zur Personenauswahl vorzunehmen. Dazu werden Telefonnummern generiert (z.B. nach dem Gabler-Häder-Design) und bei den entsprechenden Haushalten angerufen. Innerhalb des Haushaltes muss dann eine Zielperson ermittelt (üblicherweise nach der Last-Birthday-Methode) und deren Adresse erfragt werden. Sie bekommt den Fragebogen zugeschickt. Leider sind die mit dieser Methode erzielten Response-Raten unterdurchschnittlich gering. Zudem ist das Verfahren relativ kosten- und zeitaufwändig. Eine andere Möglichkeit besteht in einem Random-Walk mit Verteilung der Fragebögen. Dabei steckt ein Interviewer die Fragebögen entsprechend einer Random-Walk-Anweisung in die Briefkästen bzw. gibt sie im Haushalt ab. Problematisch ist es, wenn der Interviewer nicht dazu kommt, die Auswahl der Befragungsperson innerhalb des Haushaltes vorzunehmen. Dann muss sich eine entsprechende Anweisung auf dem Fragebogen befinden und davon ausgegangen werden, dass die Zielperson im Zielhaushalt selbständig bestimmt wird. Dieses Verfahren ist nur bei Befragungen in kleineren Regionen empfehlenswert, da es sonst zu aufwändig wäre. Als problematisch muss darüber hinaus angemerkt werden, dass die Inklusionswahrscheinlichkeiten der zu Befragenden bei diesem Design nicht exakt bestimmt werden können.

4. Wie zieht man Stichproben für persönlich-mündliche Befragungen?

Für persönlich-mündliche Bevölkerungsbefragungen, die auf Zufallsstichproben beruhen, werden im Wesentlichen zwei Stichprobendesigns eingesetzt, die im Folgenden vorgestellt werden sollen:

4.1 Das ADM-Stichproben-System

Die Arbeitsgemeinschaft Deutscher Markt- und Sozialforschungsinstitute e.V. (ADM) verwendet für persönlich-mündliche Bevölkerungsbefragungen der Markt-, Media- und Sozialforschung in der Regel ein speziell entwickeltes Design, das sogenannte ADM-Design. Es handelt sich dabei um eine dreistufige geschichtete Zufallsauswahl.

Erster Schritt: Auswahl der Gebiete

Das gesamte bewohnte Gebiet der Bundesrepublik lässt sich auf der Basis der amtlichen Statistik in etwa 50.000 bis 60.000 Flächen aufteilen. Hierbei handelt es sich um die feinste von der amtlichen Statistik mit Daten unterstützte Gliederung. Diese Gebiete lassen sich bei Gemeinden mit über 10.000 Einwohnern zudem mithilfe von digitalisierten Straßenkarten abgrenzen. Aus diesen Flächen wird vom ADM nach dem Zufallsprinzip eine Stichprobe gezogen, wobei die dabei ermittelten Elemente als Samplepoints bezeichnet werden. Während von 1997 bis zum Jahr 2003 die etwa 80.000 Stimmbezirke der Bundesrepublik die Grundlage für die Festlegung der Samplepoints bildeten, erfolgt seit 2003 eine verfeinerte Aufteilung der Samplepoints in Straßenabschnitte.

Für die Bildung der Samplepoints verwendet man die zur Verfügung stehenden kleinsten administrativen Flächeneinheiten bis auf Baublockebene. Dazu können inzwischen amtliche innenstädtische Gliederungen und Sachdaten systematisch herangezogen werden. Zu diesen Daten zählen Stadtteildaten, Einträge in Telefonverzeichnisse und so weiter.

Einer bewährten Konvention folgend werden auch für das ADM Design ab 2003 nun für 1.000 zu realisierende Interviews im westlichen Teil Deutschlands 210 und im östlichen (seit der Gebietsreform 1996) 48 solcher Samplepoints ausgesucht. Diese sind so angelegt, dass sie durchschnittlich 600 bis 700 Haushalte umfassen.

Um in den Samplepoints Befragungen durchführen zu können, ist ein sogenanntes Netz erforderlich, also eine bestimmte Anzahl an Interviewern, die auf diese Gebiete verteilt sind.

Unserem oben entwickelten Modell folgend liegt hier ein Design mit Schichtung vor. Als Schichtungskriterien dienen beim ADM-Design die Bundesländer, die Regierungsbezirke und die Kreise in Kombination mit einer Gemeindetypisierung. „Die Zellen im Stichprobensystem werden aus den Landkreisen/kreisfreien Städten und den Gemeindegrößenklassen gebildet. Die Gemeindegrößenklassen werden hierbei in erweiterter Form in einer 10er-Einteilung herangezogen ... Die Gemeinden, die nicht zu einer Stadtregion gehören, werden innerhalb einer Schicht (Landkreise) entsprechend ihrer tatsächlichen (politischen) Größe zugeordnet.“ (ADM 1999:86f.)

Zweiter Schritt: Die Auswahl des Zielhaushaltes

Das Ziel des zweiten Auswahltrittes besteht darin, in den Samplepoints die Zielhaushalte zu bestimmen, in denen dann später eine Befragung stattfinden soll. Dazu erfolgt, ausgehend von einer bestimmten Startadresse eine zufällige Begehung (auch als Random Walk oder Random Route bezeichnet) des Samplepoints. Für dieses Random Walk erstellen die Erhebungsinstitute Begehungsvorschriften, denen die Interviewer zu folgen haben.“ (Häder 2014)

Dritter Schritt: Die Auswahl der Zielperson

Innerhalb der im zweiten Schritt ermittelten Zielhaushalte gilt es nun im dritten Schritt, die eigentliche Zielperson zu finden. Hierfür werden in der Umfragepraxis zwei unterschiedliche Methoden eingesetzt. Bei der ersten Variante wird zunächst eine Auflistung aller Haushaltsmitglieder, die zur Grundgesamtheit gehören und damit für eine Befragung prinzipiell infrage kommen, vorgenommen. Diese hat so zu erfolgen, dass die Personen geordnet nach ihrem Alter in eine Liste eingetragen werden. Auf dem Adressauflistungsblatt sind Zufallszahlen aufgedruckt, nach denen danach die im Haushalt zu befragende Zielperson ermittelt wird. Bei der zweiten Variante wird jene Person um die Teilnahme an der Befragung gebeten werden, die als letzte (oder nächste) Geburtstag hat(te).

Theoretisch hat beim ADM-Design jeder Haushalt die gleiche Chance, ausgewählt zu werden. Die Auswahlchance des Haushaltmitglieds hingegen ist abhängig von der Zahl der zur Grundgesamtheit gehörenden Personen im Haushalt. Deshalb ist bei Analysen der Daten eine Designgewichtung notwendig.

In der Praxis tritt bei der Anwendung des ADM-Designs eine Reihe von Problemen auf, die hier zumindest benannt, wenn auch nicht ausführlich diskutiert sein soll:

- Der subjektive Entscheidungsspielraum der Interviewer bei der Auswahl der zu befragenden Haushalte ist relativ groß, insbesondere beim Random-Route-Verfahren. Etwas eingeschränkt ist er bei der Anwendung des Adressvorlaufverfahrens (vgl. Häder/Häder 1997)

- Die Möglichkeiten des Instituts, die Arbeit der Interviewer bei der Auswahl der Haushalte/Personen zu kontrollieren, sind – etwa im Vergleich zur Einwohnermeldeamtsstichprobe – sehr eingeschränkt.
- Zunächst liegt keine Liste von Haushalten vor, sondern diese muss erst durch eine Begehung im Wahlkreis erzeugt werden. Dies impliziert wiederum eine relativ starke räumliche Klumpung, wie sie z.B. bei der Einwohnermeldeamtsstichprobe nicht auftritt.
- Die Einhaltung der Definition der Grundgesamtheit ist mitunter schwierig. So ist es z.B. nicht ohne weiteres zu entscheiden, wie bei einer Befragung, deren Grundgesamtheit die deutschen Staatsbürger sind, mit ausländisch klingenden Namen umzugehen ist.
- Die Designgewichtung bei durch Nonresponse unvollständig umgesetzten Stichprobenplänen kann u.U. zu einer Verzerrung des Schätzers führen (Rothe 1994, Rothe/Wiedenbeck 1994).
- Durch die Startadressenregelung bei der Begehung haben in längeren Straßen Haushalte in hinteren Abschnitten keine Chance, ausgewählt zu werden.

Trotz der aufgeführten Probleme wird das ADM-Design in der Markt-, Medien- und Sozialforschung häufig eingesetzt, da es wesentlich kosten- und zeitgünstiger ist als die qualitativ höherwertige Einwohnermeldeamtsstichprobe (vgl. Häder/Häder 1997). Beispiele für prominente Untersuchungen, die nach dem ADM-Design erhoben wurden, sind z.B.: ALLBUS bis 1994 und 1998 sowie Rekrutierung des German Internet Panel.

4.2 Einwohnermeldeamtsstichproben

Das derzeitig qualitativ hochwertigste Design für bundesweite Bevölkerungsumfragen in Deutschland („best practice“) ist die Einwohnermeldeamtsstichprobe (auch „Registerstichprobe“). Hierbei handelt es sich um ein geschichtetes zweistufiges Zufallsverfahren.

Da in Deutschland kein zentrales Einwohnerregister besteht, sondern die Register auf Gemeindeebene organisiert sind, muss in einem ersten Schritt eine Auswahl von Gemeinden erfolgen. Simulationen haben ergeben, dass bei einer Auswahl von ca. 150 Gemeinden relativ stabile Verteilungen für viele interessierende Variablen erreicht werden. Das heißt, dass die Zahl zu ziehender Gemeinden (Primary Sampling Units=PSUs) nicht wesentlich kleiner sein sollte. Damit ist aber auch schon gesagt, dass es für die Ziehung von Einwohnermeldeamtsstichproben keinen festen Masterplan gibt. Allgemein gilt, dass je höher die Zahl der PSUs, desto größer die Präzision der Schätzer ist. Dies ist aber gerade ein wesentliche Kosten- und Zeitfaktor (vgl. Albers 1997).

Hier soll als Beispiel für eine Einwohnermeldeamtsstichprobe die Stichprobenkonstruktion für den European Social Survey 2003 vorgestellt werden.

Grundgesamtheit dieser Untersuchung waren alle Personen im Alter von 15 Jahren und älter, die in Deutschland in Privathaushalten leben. Für Ost- und Westdeutschland wurde ein disproportionaler Stichprobenansatz derart gewählt, dass in Ostdeutschland 1000 Interviews und in Westdeutschland 2.000 Interviews durchgeführt werden sollten. Die Stichproben wurden für beide Regionen unabhängig generiert.

Zunächst wurden die Gemeinden nach Kreisen und Gemeindegrößenklassen (nach BIK) geschichtet. Dabei ergaben sich 1.085 nichtleere Schichtzellen in West- und 435 Schichtzellen in Ostdeutschland.

In der ersten Auswahlstufe wurden aus diesen Schichten 100 Gemeinden in West- und 50 Gemeinden in Ostdeutschland proportional zur Bevölkerungszahl 15 Jahre und älter ausgewählt. Die Zahl der Gemeinden, die letztlich aus jeder Schicht entnommen wurde, wurde durch kontrolliertes Runden nach

Cox (1987) festgelegt. Die Zahl der Sample Points in Westdeutschland beträgt 108, in Ostdeutschland 55 (einige große Gemeinden sind mit mehr als einem Sample Point in der Stichprobe vertreten).

In der zweiten Auswahlstufe wird in den entsprechenden Einwohnermeldeämtern für jeden Sample Point per systematischer Zufallsauswahl eine feste Zahl von Personen gezogen. Die Festlegung des Umfangs der Bruttostichprobe erfolgte in Abhängigkeit von der zu erwartenden Zahl stichprobenneutraler Ausfälle und der erwarteten Responserate.

5. Wie werden Stichproben für Online-Erhebungen rekrutiert?

Grundsätzlich gilt es anzumerken, dass Online-Befragungen sich naturgemäß jeweils nur an diejenigen wenden können, die Zugang zum jeweiligen Medium haben. Da aber z.B. die Internet-Nutzer nicht als zufällige Substichprobe der Gesamtbevölkerung angesehen werden können, sind repräsentative Bevölkerungsbefragungen auf diese Weise (noch) nicht möglich.

E-Mail-Befragungen

Wenn eine Liste mit E-Mail-Adressen vorhanden ist, kann diese für eine Zufallsauswahl genutzt werden. Diese kann uneingeschränkt zufällig, systematisch oder geschichtet erfolgen. Der Fragebogen ist, sofern es sich nicht um einen außerordentlich kurzen Fragebogen handelt, als Attachment mitzuschicken. Problematisch ist bei der Rücksendung der Fragebögen allerdings, dass der Befragte durch die angegebene E-Mail-Absender-Adresse möglicherweise Bedenken hinsichtlich der Anonymität seiner Angaben haben könnte. Hier muss herausgestellt werden, dass die Angaben aus dem Fragebogen getrennt von den persönlichen Daten des Befragten – etwa seiner E-Mail-Adresse – behandelt werden. Beachtenswert sind außerdem die im Internet üblichen Umgangsformen („Netiquette“). Um diesen gerecht zu werden, ist eine Ankündigungsmail empfehlenswert. Die eigentliche Versendung des Fragebogens sollte nur nach einer Bestätigung des zu Befragenden erfolgen.

Web-Surveys

Bei Stichproben für Web-Surveys ist zunächst zu unterscheiden zwischen Designs, bei denen keine Auswahl durch den Forscher stattfindet und solchen, bei denen zumindest versucht wird, die Auswahl der zu Interviewenden zu beeinflussen. In die erste Kategorie fallen Erhebungen, bei denen der zu Befragende angesprochen wird mittels

- Banner (Hyperlink verweist auf die Homepage der Befragung)
- Newsletter,
- Newsgroups und Mailinglisten (für zielgruppenorientierte Befragungen)

Diese Designs haben den Nachteil vollständiger Selbstselektion durch die potentiellen Zielpersonen. Zugleich ist mit geringen Responseraten zu rechnen.

Eine Auswahl der Zielpersonen durch den Forscher wird mit Hilfe der folgenden Designs zumindest versucht:

- Listenbasierte Auswahl. Der zu Befragende erhält eine E-Mail mit der Bitte um Teilnahme und die Angabe der entsprechenden URL.
- Popup- bzw. Intercept. Jeder n-te Besucher einer Homepage erhält eine Teilnahmeaufforderung. Das Problem bei diesem Design besteht darin, dass nur die Inklusionswahrscheinlichkeiten für den "visit" einer Seite angebbbar sind, nicht jedoch die für den "visitor".

- Rekrutierung über anderes Befragungsmedium (z.B. Telefon-Screening). Auch hier sind die Teilnehmer an der eigentlichen Befragung zunächst nur Internet-Nutzer. Beim German Internet Panel werden Befragungswillige, die über keinen Computer verfügen, mit einem Internet-Zugang ausgestattet. Das GESIS-Panel setzt in diesem Fall auf einen Methoden-Mix und bietet eine schriftliche Befragung an. Der Aufwand für ein Screening ist letztlich nur lohnenswert für Mehrfachbefragungen, d.h. für die Einrichtung von Access-Panels, denn die Ausschöpfung ist in der Regel sehr gering, d.h. es handelt sich um ein äußerst aufwändiges Verfahren.

Die folgende Übersicht zeigt die Einteilung in zufällige und nicht-zufällige Methoden nach (Couper, 2000):

Übersicht 1: Zufällige und nicht-zufällige Methoden der Stichprobenziehung bei Online-Befragungen

| Nonprobability Methods | Probability-Based Methods |
|------------------------------------|--|
| Polls as entertainment | 4. Intercept surveys |
| Unrestricted self-selected surveys | 5. List-based samples |
| Volunteer opt-in panels | 6. Web option in mixed-mode surveys |
| | 7. Pre recruited panels of Internet users |
| | 8. Pre-recruited panels of full population |

Insgesamt ist auf dem Gebiet der Stichprobenziehung für Online-Befragungen noch keine befriedigende Lösung in Bezug auf zufällige Auswahlverfahren gefunden. Das wesentliche Problem ist, dass in keinem der vorgestellten Verfahren Inklusionswahrscheinlichkeiten exakt angegeben werden können. Weiterhin ist es in der Regel nicht möglich, „nachzuhaken“, d.h. über Erinnerungsschreiben die Responserate zu erhöhen. So ist eine hohe Selbstselektivität kennzeichnend für die meisten Online-Befragungen. Diese führt häufig zu systematischen Abweichungen zwischen den Teilnehmern der Studie und Gesamtheit der Online-Nutzer. Als problematisch für eine gegebenenfalls nötige Adjustment-Gewichtung ist das Fehlen von Verteilungen sinnvoller Anpassungsvariablen für die Gesamtheit zu konstatieren.

6. Wie werden Stichproben für Cross-cultural Surveys gezogen?

Für Studien, deren Grundgesamtheit sich über mehrere Staaten oder Nationen erstreckt (engl: cross-cultural surveys, multipopulation surveys), sollten Designs gefunden werden, die den Anspruch einlösen, äquivalente Stichproben für die unterschiedlichen Länder zu generieren. Gleichzeitig müssen die Gegebenheiten in den einzelnen Ländern bzw. Kulturen hinsichtlich der Verfügbarkeit von Auswahlrahmen, Erfahrungen, Traditionen und Kosten berücksichtigt werden.

So ergibt sich zunächst die Frage, wie eine optimale Stichprobenstrategie für einen cross-cultural Survey gestaltet sein sollte. Hierzu existieren in der Profession durchaus konträre Ansichten. Zum Beispiel galt für das Eurobarometer die Maxime, für alle Länder das gleiche Stichprobendesign einzusetzen. In anderen ländervergleichenden Studien wurden neben Zufallsstichproben auch Quotadesigns eingesetzt (frühere ISSP-Umfragen).

Ein allgemeiner Hinweis für die Ziehung von cross-cultural samples findet sich bei Kish (1994, S. 173): "Sample designs may be chosen flexibly and there is no need for similarity of sample designs. Flexibility

of choice is particularly advisable for multinational comparisons, because the sampling resources differ greatly between countries. All this flexibility assumes probability selection methods: known probabilities of selection for all population elements." Häder und Gabler (2003) fügen an: Werden dann noch äquivalente Schätzer verwendet, ergibt sich eine hohe Vergleichbarkeit der Ergebnisse (siehe auch Häder und Lynn 2007). Diese Strategie wird für den European Social Survey, inzwischen aber auch andere europäische Studien, wie z.B. die European Value Study 2008, angewendet.

6.1 Auswahlrahmen

Der wichtigste Punkt bei der Suche nach den bestmöglichen Stichprobendesigns in den ESS-Ländern ist jeweils die Festlegung des Auswahlrahmens. Hierbei gilt es, Listen zu finden, die die definierte Grundgesamtheit möglichst vollständig abdecken. Es sollen also möglichst keine Elemente enthalten sein, die auf der Liste verzeichnet sind, aber nicht zur Gesamtheit gehören (Overcoverage), noch sollen Elemente fehlen (Undercoverage). Die Bestimmung der Grundgesamtheit lautet: "The survey will be representative of all persons aged 15 and over – no upper age limit – resident within private households in each country, regardless of their nationality, citizenship or language."

Folgende Arten von Listen sind für die Stichprobenziehung des ESS im Einsatz:

a) Länder mit verlässlichen Listen der *Einwohner*, verfügbar für die Sozialforschung, z.B. Schweden, Finnland, Norwegen. In diesen Ländern existieren zentrale Einwohnermelderegister, aus denen Stichproben für den ESS gezogen werden können. Zudem sind meist zusätzliche Informationen zu den Individuen, wie z.B. Alter und Nationalität verfügbar. Dies ist sowohl für die Kontrolle der Interviewer als auch für Nonresponseanalysen hilfreich.

b) Länder mit verlässlichen Listen der *Haushalte*, verfügbar für die Sozialforschung, z.B. Zypern. Hier wurde in der Runde 3 eine Liste der Haushalte aus dem Zensus, ergänzt durch die Konsumentenliste des Energieversorgers genutzt.

c) Länder mit verlässlichen Listen der *Adressen*, verfügbar für die Sozialforschung, z.B. Liste der Postadressen "PTT-afgiftenpuntenbestand" in den Niederlanden oder das "Postcode Address File" in Großbritannien.

Schließlich gibt es Teilnehmerländer des ESS, in denen es nicht gelungen ist, geeignete Listen für die Stichprobenziehung zu finden bzw. in denen zwar hervorragende Listen existieren, diese aber für die Sozialforschung nicht zugänglich sind. So besteht in Portugal z.B. ein zentrales Einwohnermelderegister, das Statistische Amt des Landes kann jedoch aufgrund rechtlicher Bestimmungen den Zugriff darauf nicht gewähren. Eine ähnliche Situation besteht in Österreich. In diesen Ländern muss notgedrungen auf Flächenstichproben zurückgegriffen werden, die mit einer Random-Route-Prozedur gekoppelt sind. Das Sampling Expert Panel achtet jedoch darauf, dass dabei das Prinzip befolgt wird, die Adressauflistung und die eigentliche Befragung grundsätzlich zu trennen. In allen Ländern, in denen gegenwärtig noch Random-Route-Techniken verwendet werden, ist es das erklärte Ziel des Sampling Expert Panels, diese sobald wie möglich durch die Auswahl aus Listen zu ersetzen, da es in der statistischen Profession durchaus Zweifel daran gibt, inwieweit Random Routes als „strictly random“ gewertet werden können (Lyberg, evaluation of IALS DATE).

6.2 Stichprobendesign und Sample Design Data File

In Abhängigkeit von den zur Verfügung stehenden Auswahlrahmen werden schließlich die Stichprobendesigns festgelegt. Diese sind beim ESS durchaus sehr unterschiedlich geartet. Der Range reicht von uneingeschränkter Zufallsauswahl (Finnland Runde 1-6) bis zu vierstufigen geschichteten

und geklumpten Designs (Russland Runde 5). Die Designs werden in einem standardisierten Formblatt (Sign off forms) dokumentiert. Hierbei werden für jedes Land die Kategorien

- Target population, Population coverage
- Sampling frame
- Sampling Design
- Inclusion Probabilities (Formel)
- Design effects
- Target response rate
- Sample size
- Special features of the design (e.g. oversampling)

festgehalten.

Grundprinzip ist bei der Wahl der Stichprobenpläne, dass es sich jeweils um Zufallsstichproben handeln muss. Das heißt, die Inklusionswahrscheinlichkeiten müssen sämtlich bekannt und dokumentiert sein. Für die Speicherung der Inklusionswahrscheinlichkeiten aller Auswahlstufen und weiterer Variablen zur Beschreibung des Stichprobendesigns wie z.B. Primary Sampling Units, Sample Points und Schichtungsvariablen wird das Sample Design Data File genutzt. Die Angaben aus diesem Datensatz ermöglichen es dem Nutzer, jegliche Einzelheiten der Stichprobenziehung nachzuvollziehen, Berechnungen von Design-Gewichten und Design-Effekten sowie Varianzschätzungen vorzunehmen. Die Sample Design Data Files sind von den Ländern für die Öffentlichkeit zugänglich (<http://ess.nsd.uib.no/>), in denen nicht datenschutzrechtliche Bedenken die Publikation verbieten.

Einen Ausschnitt aus dem Sample Design Data File Frankreichs in Runde 3 bietet der folgende Screen Shot (vgl. Abbildung 2):

| | CNTRY | ROUND | IDNO | PROB1 | PROB2 | PROB3 | PROB4 | PSU | SAMPPOINT | STRATEX1 | STRATIM1 | OUTCOME |
|----|-------|-------|----------|-------------------|-------------------|--------------------|-------|-----|-----------|----------|----------|---------|
| 1 | FR | 3 | 68000426 | 0,000041372282988 | 0,019586507072905 | 0,2500000000000000 | . | 45 | . | 43 | . | 1 |
| 2 | FR | 3 | 68000429 | 0,000041372282988 | 0,019586507072905 | . | . | 45 | . | 43 | . | 2 |
| 3 | FR | 3 | 68001671 | 0,000041372282988 | 0,019586507072905 | 1,0000000000000000 | . | 45 | . | 43 | . | 1 |
| 4 | FR | 3 | 68001860 | 0,000041372282988 | 0,019586507072905 | . | . | 45 | . | 43 | . | 2 |
| 5 | FR | 3 | 68001878 | 0,000041372282988 | 0,019586507072905 | . | . | 45 | . | 43 | . | 2 |
| 6 | FR | 3 | 68002037 | 0,000041372282988 | 0,019586507072905 | . | . | 45 | . | 43 | . | 2 |
| 7 | FR | 3 | 68002733 | 0,000041372282988 | 0,019586507072905 | 0,3333333333333333 | . | 45 | . | 43 | . | 1 |
| 8 | FR | 3 | 68002997 | 0,000041372282988 | 0,019586507072905 | . | . | 45 | . | 43 | . | 2 |
| 9 | FR | 3 | 68003906 | 0,000041372282988 | 0,019586507072905 | 0,3333333333333333 | . | 45 | . | 43 | . | 1 |
| 10 | FR | 3 | 68004503 | 0,000041372282988 | 0,019586507072905 | 0,5000000000000000 | . | 45 | . | 43 | . | 1 |
| 11 | FR | 3 | 68007299 | 0,000041372282988 | 0,019586507072905 | 0,5000000000000000 | . | 45 | . | 43 | . | 1 |
| 12 | FR | 3 | 68008502 | 0,000041372282988 | 0,019586507072905 | . | . | 45 | . | 43 | . | 2 |
| 13 | FR | 3 | 68009867 | 0,000041372282988 | 0,019586507072905 | 0,5000000000000000 | . | 45 | . | 43 | . | 1 |
| 14 | FR | 3 | 68010377 | 0,000041372282988 | 0,019586507072905 | 1,0000000000000000 | . | 45 | . | 43 | . | 1 |
| 15 | FR | 3 | 68010779 | 0,000041372282988 | 0,019586507072905 | . | . | 45 | . | 43 | . | 2 |
| 16 | FR | 3 | 68011601 | 0,000041372282988 | 0,019586507072905 | 0,2500000000000000 | . | 45 | . | 43 | . | 1 |
| 17 | FR | 3 | 68012390 | 0,000041372282988 | 0,019586507072905 | . | . | 45 | . | 43 | . | 2 |
| 18 | FR | 3 | 68012891 | 0,000041372282988 | 0,019586507072905 | . | . | 45 | . | 43 | . | 2 |
| 19 | FR | 3 | 68001389 | 0,000130644217580 | 0,005303476723630 | . | . | 119 | . | 44 | . | 2 |
| 20 | FR | 3 | 68002439 | 0,000130644217580 | 0,005303476723630 | . | . | 119 | . | 44 | . | 2 |
| 21 | FR | 3 | 68002856 | 0,000130644217580 | 0,005303476723630 | 0,5000000000000000 | . | 119 | . | 44 | . | 1 |
| 22 | FR | 3 | 68003168 | 0,000130644217580 | 0,005303476723630 | . | . | 119 | . | 44 | . | 2 |
| 23 | FR | 3 | 68003630 | 0,000130644217580 | 0,005303476723630 | . | . | 119 | . | 44 | . | 2 |
| 24 | FR | 3 | 68003633 | 0,000130644217580 | 0,005303476723630 | 0,3333333333333333 | . | 119 | . | 44 | . | 2 |
| 25 | FR | 3 | 68007302 | 0,000130644217580 | 0,005303476723630 | . | . | 119 | . | 44 | . | 2 |

6.3 Design-Gewichte

Die ungleichen Auswahlwahrscheinlichkeiten, die bei der Stichprobenziehung auftreten, müssen bei der Schätzung wieder ausgeglichen werden. Dazu werden beim ESS vom Sampling Expert Panel Design-Gewichte berechnet. Diese beruhen auf den Angaben zu den Wahrscheinlichkeiten bei den einzelnen Auswahlstufen aus den Sample Design Data Files. So ist beim französischen Design in Abbildung 2 das Produkt von PROB1 und PROB2 konstant, lediglich PROB3 variiert. Hierbei handelt es sich um die Auswahl innerhalb des Haushaltes. Da diese Wahrscheinlichkeit nicht im vornherein bestimmt werden kann, sondern erst beim Kontaktieren des Haushaltes festgestellt wird, sind diese Angaben auch nur für die Nettostichprobe angegeben. Die Auswahlwahrscheinlichkeit ist in diesem Fall umgekehrt proportional zur Zahl der Zielpersonen im Haushalt, also bei einer Person 1, bei 2 Personen 0.5 bei 3 Personen 0.33 usw.

Das Gesamtdesigngewicht ergibt sich als $w_i = 1/(\text{PROB1}_i \cdot \text{PROB2}_i \cdot \text{PROB3}_i)$. Es wird anschließend derart normiert, dass die Summe der Gewichte dem Stichprobenumfang entspricht.

Tabelle 1 zeigt die Wirkung der Designgewichtung für Daten des ESS Runde 3. Für Österreich, Portugal und Russland würde ohne die notwendige Designgewichtung der Anteil der Einpersonenhaushalte deutlich unterschätzt werden. Bei Finnland und Schweden betragen die Designgewichte dagegen 1, da alle Elemente mit gleichen Wahrscheinlichkeiten ausgewählt wurden. Bei der Variable V2 (Anteil der Personen, die wöchentlich 3 und mehr Stunden Fernsehen schauen) ist der Unterschied zwischen gewichteten und ungewichteten Daten nicht so groß, da diese Variable weniger mit der Haushaltsgröße zusammenhängt, die schließlich in die Gewichtung eingeflossen ist.

Tabelle 1: Gewichtete und ungewichtete Schätzer für die Variablen Anteil der Einpersonenhaushalte (%) und Fernsehdauer 3 Stunden und mehr pro Woche (%) für fünf Länder

| | Austria | Portugal | Russia | Finland | Sweden |
|-----------------|---------|----------|--------|---------|--------|
| V1, ungewichtet | 8,8 | 7,6 | 10,8 | 23,4 | 21,1 |
| V1, gewichtet | 21,2 | 17,2 | 23,7 | 23,4 | 21,1 |
| V2, ungewichtet | 13,7 | 22,2 | 23,2 | 14,5 | 9,8 |
| V2, gewichtet | 12,5 | 21,5 | 22,8 | 14,5 | 9,8 |

V1 = Einpersonenhaushalte (%)

V2 = 3 Stunden+ TV pro Woche (%)

7. Literaturverzeichnis

- Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V / A.G.M.A. e.V.. (1999): Stichprobenverfahren in der Umfrageforschung. Opladen: Leske + Budrich.
- Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V./ (2012). ADM-Forschungsprojekt „Dual-Frame-Ansätze“. Forschungsbericht. Zuletzt aufgerufen am 15.4.2014.
<https://www.adm-ev.de/forschungsprojekte/>
- Albers, I. (1997): Einwohnermelderegister-Stichproben in der Praxis. Ein Erfahrungsbericht. In S. Gabler & J. H. P. Hoffmeyer-Zlotnik (Hrsg.), Stichproben in der Umfragepraxis. Opladen: Westdeutscher Verlag.
- Couper, M. P. (2000). Web surveys: a review of issues and approaches. *Public Opinion Quarterly*, 64(4), 464-494.
- Cox, L. H. (1987): A Constructive Procedure for Unbiased Controlled Rounding. *JASA*, Vol. 82, Issue 398: 520-524.
- Eurobarometer Spezial 356. (2011). Roaming im 2010. Zuletzt aufgerufen am 15.4.2014.
http://ec.europa.eu/public_opinion/archives/ebs/ebs_356_de.pdf
- Friedrichs, J. (2000): Effekte des Versandes des Fragebogens auf die Antwortqualität bei einer telefonischen Befragung. In: V. Hüfken (Hrsg.), Methoden in Telefonumfragen. Opladen: Westdeutscher Verlag.
- Gabler, S. & Ö. Ayhan (2007). Gewichtung bei Erhebungen im Festnetz und über Mobilfunk: Ein Dual Frame Ansatz. In: S. Gabler & S. Häder (Hrsg.), Mobilfunktelefonie – Eine Herausforderung für die Umfrageforschung. Mannheim: GESIS, S. 39-45.
- Gabler, S., S. Häder, I. Lehnhoff & E. Mardian (2012). Weighting for Unequal Inclusion Probabilities and Nonresponse in Dual Frame Telephone Surveys. S. 147-167. In S. Häder, M. Häder & M. Kühne (Hrsg.): Telephone Surveys in Europe. Research and Practice. Heidelberg: Springer.
- Häder, M. (2014). Empirische Sozialforschung. Eine Einführung. 3. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Häder, M. & Häder, S. (1997). Adreßvorlaufverfahren: Möglichkeiten und Grenzen. In S. Gabler & J. H. P. Hoffmeyer-Zlotnik (Hrsg.), Stichproben in der Umfragepraxis. Opladen: Westdeutscher Verlag.
- Häder, S. & S. Gabler (1998). Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. In S. Gabler, S. Häder & J.H.P. Hoffmeyer-Zlotnik (Hrsg.), Telefonstichproben in Deutschland. Opladen: Westdeutscher Verlag.
- Häder, S. & S. Gabler (2003). Sampling and estimation. In J. Harkness, F. van de Vijver & P. Mohler (Hrsg.), Cross cultural survey methods. New York: Wiley.
- Häder, S. & P. Lynn (2007). How representative can a multi-nation survey be?. In: R. Jowell et al (Hrsg.), Measuring attitudes cross-nationally. London: Sage Publications.
- Heckel, C. (2002). Erstellung der ADM-Telefonauswahlgrundlage. In S. Gabler & S. Häder (Hrsg.), Telefonstichproben. Methodische Innovationen und Anwendungen in Deutschland. Münster: Waxmann Verlag.

- Heckel, C. (2011). Auswahlrahmen Dual Frame. Vortrag im Rahmen des Treffens der Arbeitsgruppe MOBILSAMPLE am 21. Oktober 2011 an der TU Dresden.
- Hüfken, V. (2000): Kontaktierung bei Telefonumfragen. Auswirkungen auf das Kooperations- und Antwortverhalten. In: V. Hüfken (Hrsg.), Methoden in Telefonumfragen. Opladen: Westdeutscher Verlag.
- Humpert, A. & K. Schneiderheinze (2002). Stichprobenziehung für telefonische Zuwandererumfragen. In: S. Gabler & S. Häder (Hrsg.), Telefonstichproben. Methodische Innovationen und Anwendungen in Deutschland. Münster: Waxmann Verlag.
- Kish, L. (1994). Multipopulation Survey Designs: Five Types with Seven Shared Aspects. International Statistical Review 62, 167-186.
- Lavrakas, P.J. (1993): Telephone Survey Methods. Newbury Park: Sage.
- Rothe, G. (1994): Wie (un)wichtig sind Gewichtungen? In: S. Gabler, J.H.P. Hoffmeyer-Zlotnik & D. Krebs (Hrsg.), Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag.
- Rothe, G. & M. Wiedenbeck (1994). Stichprobengewichtung: Ist Repräsentativität machbar? In: S. Gabler, J.H.P. Hoffmeyer-Zlotnik & D. Krebs (Hrsg.), Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag
- Sand, M. (2014a). Dual-Frame-Telefonstichproben. Entwicklung. Handhabung und Gewichtung. GESIS Technical Reports 2014/02.
- Sand, M. (2014b). "Überarbeitung des GESIS Auswahlrahmens für Telefonstichproben - Führt die Anreicherung durch die Angaben der Bundesnetzagentur zu einer Verbesserung der Auswahlgrundlage?" In: M. Häder & R. Neumann. Dresdner Beiträge zur Soziologie. Band 5. In Druck.
- Struminskaya, B., L. Kaczmirek, I. Schaurer, W. Bandilla, S. Gabler & S. Häder (2011) Identifying non-working numbers in cell phone RDD samples via HLR-Lookup technology: reduced survey costs and higher precision in response rate calculation. In: Survey Practice.
- von der Heyde, C. (2002). Das ADM-Telefonstichprobenmodell. In: S. Gabler & S. Häder (Hrsg.), Telefonstichproben. Methodische Innovationen und Anwendungen in Deutschland. Münster: Waxmann Verlag.