

2.3.3.2. Wahl der Stichprobe

Das Ziel einer Auswahl einer zu untersuchenden Teileinheit - der Stichprobe, oder auch "Sample" genannt - ist, dass diese die gleichen Merkmale im richtigen Verhältnis widerspiegelt - sie repräsentiert das Bild der Grundgesamtheit im Kleinen.

Für die Wahl der Stichprobe gibt es eine Reihe von Verfahren, die in Abb. 31 übersichtsartig zusammengestellt sind. Der Ausgangspunkt für jede Auswahl ist die Definition der Grundgesamtheit (s. o.). Dabei wird festgelegt welche Merkmale in der Studie erfasst werden sollen. Das betrifft den räumlichen Rahmen (ökologische Zone, Stadt/Land usw.) und den zeitlichen Rahmen (Saison, Ernte-/Fastenzeit, Wochenrhythmen, Alltag/Festtag usw.), der auch Überlegungen zu den zeitlichen Beziehungen zwischen Ernährung und ihren Auswirkungen enthält, z. B. lange Latenzzeiten bei der Krebsentstehung. Es muss festgelegt werden welche Einheiten zur Auswahl herangezogen werden.

Die *Erhebungseinheiten* können einzelne Menschen sein, aber auch Haushalte, Familien, Häuser, Straßenzüge, Betriebe, Vereine usw. Die Erhebungseinheit kann mit der *Untersuchungseinheit* identisch sein, muss es aber nicht. So können größere Einheiten zur Auswahl herangezogen werden, von denen dann die einzelnen Menschen untersucht werden. Die Ergebnisse bzw. die Aussagen der Studien können sich wiederum auf gleiche, aber auch unterschiedliche Einheiten beziehen. Die *Aussageeinheiten* von Studien können eine Zusammenfassung der Ergebnisse der kleineren Untersuchungseinheiten sein. So könnten Haushalte als Erhebungseinheit ausgewählt sein, deren einzelne Personen wären Untersuchungseinheiten, deren Ergebnisse dazu herangezogen würden die Beziehung zwischen Ernährung und menschlichen Funktionen in Abhängigkeit von verschiedenen Altersgruppen - den Aussageeinheiten - abzuleiten.

Es gehört zur weiteren Planung, dass ermittelt wird, welcher *Stichproben-Auswahlrahmen* verfügbar ist und genutzt werden kann; das könnten sein: Volkszählungsdateien, Listen von Einwohnermelde-Behörden, Adress- und Telefonbücher, Schulregister usw. Die Qualität der entsprechenden Informationen ist zu überprüfen, dabei sollte auf folgende Gesichtspunkte geachtet werden:

- Vollständigkeit bzw. fehlende Personengruppen,
- Einheitlichkeit - einheitliche Darstellung der Merkmale der Elemente, weder lückenhafte, noch
mehrfache Angaben usw.,
- Aktualität,
- Zugänglichkeit bzw. Erlaubnis der Nutzung der Daten.

Abb.31 Übersicht von verschiedenen Stichproben-Typen (nach Galtung 1970; aus Sichert, Oltersdorf et al. 1984)

Von den vorhandenen Angaben über die Grundgesamtheit hängt die Wahl der Stichprobenauswahl-Methode ab. Nur bei bekannter und definierter Grundgesamtheit und deren Zugänglichkeit kann ein gewünschter Anteil von Teilnehmern quasi per Los, also zufällig bestimmt werden.

Für eine solche *Wahrscheinlichkeits-Stichprobe*, auch Zufalls-Stichprobe (random samples) genannt - muss die Grundgesamtheit wenigstens symbolisch "anwesend" und mischbar sein. Solche "Karten" bzw. Adressen werden dann entweder als Los aus einer Urne oder nummeriert mit Hilfe von Zufallszahlen-Tabellen gezogen (*einfache Zufalls-Stichprobe*). In

Fällen, in denen die vorhandene Datei nicht nach Merkmalen geordnet ist, die auf die Studie Einfluss haben könnten (wie z. B. Alter- und Berufsgruppen), sondern nach anderen Gesichtspunkten (wie z. B. nach dem Alphabet), kann auch jede x.te Karte systematisch ausgewählt werden (*systematische Stichprobe*). Dabei wird der Beginn zufällig bestimmt - z. B. mit einer Zufallszahl wie beispielsweise 17 - und danach wird jede x.te Karte gezogen - z. B. wenn 2% in die Stichprobe kommen sollen, jede 50.te, d. h. für dieses Beispiel Nr. 67, 117, 167 u. s. f.

Es ist sicherzustellen, dass die Merkmale, die für wesentlich erachtet werden, auch tatsächlich im der Grundgesamtheit entsprechenden Maße in einer Stichprobe vertreten sind, und dass der Zugang zu den entsprechenden Merkmalen vorhanden ist. In solchen Fällen können die Grundgesamtheiten nach den Merkmalsgruppen sortiert werden. So wird z. B. nach Altersgruppen, Haushaltsgrößen, Berufsgruppen, usw. geschichtet. Diese Schichtung - auch Stratifizierung genannt - kann nach vielen Merkmalen erfolgen und die daraus abzuleitende Stichprobe kommt der einer Quotenstichprobe (s. u.) recht nahe. Dabei gilt zu bedenken, dass allein drei Merkmale in nur dichotomer Ausprägung - wie männlich/weiblich, alt/jung, Stadt/Land - bereits eine acht-zellige Gruppierung ergibt.

Der Vorteil der Schichtung ist, dass damit bestimmte Merkmale in einer Stichprobe zu einem bestimmten Anteil vorhanden sind; dies betrifft besonders die seltenen Merkmale, d. h. solche die weniger als 5% Anteil in der Grundgesamtheit haben. Die Stichprobenauswahl erfolgt in der jeweiligen Schicht nun nach dem Zufallsprinzip bzw. systematisch; dabei kann der Anteil des jeweiligen Merkmals, der der Grundgesamtheit *entsprechen* (*proportionierte, geschichtete Stichprobe* bzw. "equal unit stratified sample"). Bestimmte Merkmale, die vielleicht für die Studie besonders interessant sind, können auch bevorzugt bzw. überproportional ausgewählt werden (*disproportionierte, geschichtete Stichprobe* bzw. "unequal unit stratified sample").

Bei kleinen Grundgesamtheiten - wie z. B. der Wohnbevölkerung einer Gemeinde - kann die Stichprobe direkt ausgewählt werden (*einstufige Stichprobe* bzw. single stage sampling). Bei größeren Grundgesamtheiten - wie Großstädte oder gar Staatsbevölkerungen - wird die Stichprobe nicht direkt gezogen, sondern in Stufen. Bei der *mehrstufigen Stichprobe* (bzw. multistage sampling) wird die Grundgesamtheit in Untereinheiten zerlegt, das sind in der Regel Verwaltungseinheiten - wie Bundesländer, Regierungsbezirke, Landkreise, Städte/Gemeinden, Ortsteile, Straßen oder Wohnblöcke - und man erhält so Gebiets- bzw. Flächen-Stichproben. Es können aber auch andere zugängliche Untereinheiten definiert werden - z. B. durch eigene Festlegung der Grenzen auf einer Landkarte bzw. einem Stadtplan (Klumpen- bzw. Cluster-Stichproben).

In einer ersten Stufe wird aus den größeren Untereinheiten ausgewählt, z. B. den Bundesländern oder für HANES (s. S. 122) in den USA wurden 1900 primäre Sample Units (PSU) definiert; jeder PSU entspricht einer Stadt bzw. einem Kreis, von denen 65 PSU ausgewählt wurden. In der folgenden Auswahlstufe wird die nächste Untereinheit herangezogen - z. B. bei HANES waren dies Häuser (housing units); und letztlich werden dann daraus die jeweiligen Erhebungseinheiten ausgewählt. Die Auswahl auf jeder Stufe kann nach einem der bereits geschilderten Verfahren - also per Zufall oder systematisch; geschichtet oder ungeschichtet - erfolgen, wobei die möglichen unterschiedlichen Größen der Untereinheiten berücksichtigt werden müssen. Größere Gebiete bzw. "Klumpen" erhalten entsprechend höhere Auswahlwahrscheinlichkeiten als kleine.

In vielen Fällen von ernährungs-epidemiologischen Studien ist die Grundgesamtheit unbekannt, dann kann keine Zufallsauswahl erfolgen, sondern es muss nach den

entsprechenden Auswahlmerkmalen gesucht werden, z. B. durch die Interviewer (*Quotenstichprobe bzw. zweckbezogene Stichprobe*).

Bei diesen Auswahlverfahren steht das Untersuchungsziel im Vordergrund; die Studienteilnehmer werden nach Hypothesen-geleitenden Gesichtspunkten - also bewusst ausgewählt (*bewusste Auswahl, purposive sample*). Sollen nur ganz bestimmte Personengruppen - im Extremfall interessiert dabei nur 1 Merkmal, z. B. Kindergartenkinder - untersucht werden, dann wird eine *homogene Stichprobe* gewählt. Dies stellt jedoch eine definitorische Willkür dar, bei differenzierterer Betrachtung lassen sich Unterschiede in der homogenen Gruppe finden (z. B. männlich/weibliche, dicke/dünne, usw. Kindergarten-Kinder). So entsteht eine *heterogene Stichprobe*. Entsprechend der Studienziele können möglichst unterschiedliche Ausprägungen der gewählten Merkmale in die Stichprobe aufgenommen werden - z. B. Fall/Kontrolle; Exponierte/Nicht-Exponierte; Normal-/Übergewichtige; Vegetarier/Misch-Köstler usw. - dann sind dies *Stichproben von Extremfällen bzw. von abweichenden Fällen*.

Die Auswahl solcher Quoten-Stichproben kann nach verschiedenen Gesichtspunkten erfolgen. Da es hier keine ausreichenden Angaben über die jeweiligen Grundgesamtheiten gibt, kann keine Auswahl nach Wahrscheinlichkeitsmethoden erfolgen. Es wird versucht, dass über den Interviewer das Zufallsprinzip zum Tragen kommt, in dem dieser durch ein geschicktes Untersuchermanagement entsprechend gelenkt wird (*"Random Road-Verfahren"*). Dieses Auswahlverfahren erfordert eine aufwendige Organisation und somit einen hohen Mittel- und Zeiteinsatz. Der Aufbau eines entsprechenden Interviewer-Netzes, das sich über den ganzen Untersuchungsraum - z. B. in einem ganzen Land, wie die Bundesrepublik - erstreckt, lohnt sich nur, wenn dies auch für längere Zeit benutzt werden kann. So ist diese Quoten-Stichproben-Auswahlmethode eine gute Alternative für Institutionen, deren Daueraufgabe es ist, Erhebungen durchzuführen, wie eben Meinungsforschungsinstitute, weniger jedoch für die Durchführung einer einzelnen Studie.

Für begrenzte ernährungsepidemiologische Studien, vor allem solche die quasi aus dem Stehgreif von einzelnen Wissenschaftlern geplant werden - und dies ist eine typische universitäre Forschungssituation, z. B. für Promotionsarbeiten - würde dieser große Aufwand in keinem Verhältnis zu den geplanten Forschungsnutzen stehen. Hier müssen einfachere, bequemere zweckbezogene Stichproben-Auswahlverfahren herangezogen werden (*"convenience sampling"*).

Das Forschungsziel kann z. B. darin bestehen, dass prinzipiell erkundet werden soll, wie sich bestimmte Merkmale verteilen. Ein Beispiel für solche explorative Studien wäre die Testung der Akzeptanz eines neuen Lebensmittels bzw. von physiologischen Reaktionen darauf. Für solche Studien reichen beliebige Auswahlen von "normalen" Durchschnitts-Menschen. Solche gezielt für einen bestimmten Untersuchungszweck ausgewählten Personen (*zweckbezogene Stichprobe*) können aus dem näheren Bekanntenkreis des Untersuchers stammen, sie können an Arbeitsstätten - z. B. Studenten an Universitäten - erfasst werden und es können zufällige Straßenpassanten oder Kaufhauskunden sein.

In all diesen Fällen muss der Untersucher überlegen, wo entsprechende Studienteilnehmer aufzufinden sind. Erscheint eine Suche aufwendig, dann kann durch entsprechende Öffentlichkeitsarbeit - z. B. durch Anzeigen und Berichte in lokalen Zeitungen bis hin zu überregionalen Massenmedien - zur freiwilligen Meldung aufgerufen werden (*"Freiwilligen-Stichprobe"*). So wurde in Frauenzeitschriften zur Teilnahme an Bulimie-Studien und in Reformhaus-Kundenzeitschriften für Vegetarier-Studien aufgerufen.

In den Fällen, bei denen Studienteilnehmer mit ganz speziellen, selten Merkmalen notwendig sind, muss gelegentlich die sogenannte *”Schneeball- oder Ketten-Stichprobe”* herangezogen werden. Hier wird nach den wenigen Personen mit den gewünschten Merkmalen - wie z. B. im Rahmen von Krebs-epidemiologischen Studien alte Menschen, die seit ihrer Jugend ihren Wohnort nicht gewechselt haben oder Ehepaare, die entweder beide einheimisch sind oder einen *”fremden”* Partner haben, um auf diesem Weg die Ernährungsverhaltens-Konstanz abzuschätzen - gezielt gesucht, und jede passende Person wird nach weiteren entsprechenden Merkmalsträgern gefragt.

Alle einfachen Convenience-Stichprobenverfahren haben große Nachteile, vor allem bezüglich der Stichprobenfehler. Sie haben zwangsläufig größere *”Schiefen”*.

Eine wohl geplante, kontrollierte zweckbezogene Quotenstichproben-Auswahl, die alle wesentlichen Merkmale einer Grundgesamtheit berücksichtigt, kann eine *repräsentative (Quoten-)Stichprobe* erzeugen, die sich kaum von der einer Wahrscheinlichkeitsstichprobe unterscheidet. Diese Behauptung wird heftig diskutiert und dieser Methodenstreit hat teilweise Züge eines *”Glaubenskrieges”*. Die Quotenstichprobe hat unter der Voraussetzung, dass die wesentlichen Merkmale einer Bevölkerung gut bekannt sind, einige wichtige Vorteile gegenüber der Wahrscheinlichkeitsstichprobe:

- Es ist damit möglich auch ohne eine Erfassung der Grundgesamtheit, also ohne den direkten physischen Zugang zu einer Adressenkartei und den damit im Zusammenhang stehenden rechtlichen Problemen - wie z. B. Datenschutzbelange - eine Stichprobe zu erzeugen. So ist die Quotenstichprobe zeit- und kostensparend.
- Bei der Quotenstichprobe gibt es im Prinzip keine Stichprobenausfälle, denn Erhebungseinheiten, die die Teilnahme verweigern, werden durch entsprechende Einheiten mit den selben Merkmalskombinationen ersetzt.

Den Vorteilen dieser Quotenstichproben, die im Englischen auch *”non-probability samples”* genannt werden, stehen einige Nachteile und Schwächen gegenüber:

- Bei einer Quotenstichprobe können nicht die eigentlichen klassischen statistischen Methoden angewandt werden, mit denen die Ergebnisse der Stichprobe auf die Grundgesamtheit übertragen werden. Diese und die Berechnungen der Stichprobenfehler sind nur bei den Wahrscheinlichkeits-Stichproben - englisch *”probability samples”* - möglich.
- Die repräsentative Verteilung innerhalb der Quotenstichprobe ist zwar bezüglich der herangezogenen quotierten Merkmale möglich, doch diese stellen nur eine begrenzte Anzahl von Merkmalen dar.

Die Merkmale sollen leicht erkennbar sein, denn die Interviewer müssen sie vor Beginn der eigentlichen Erhebung erst ermitteln; sie müssen entscheiden, ob die angesprochene Person in die Quote passt oder nicht. Merkmale, die nur schwer zu ermitteln sind, wie z. B. ein bestimmtes Ernährungsbewusstsein, können kein Quotierungskriterium sein.

Die Interviewer-Arbeit wird erschwert, wenn versucht wird zu raffinierte bzw. zu vielfältige Merkmalskombinationen (*”over matching”*) zu erfassen. Es bleibt umstritten, ob Interviewer selbst wenn die Institutsorganisation optimal ist, wirklich alle Merkmale quasi zufällig in die Stichprobe bekommen.

Die Quotenstichprobe bevorzugt Personengruppen, die konstanteren Lebensstil haben; d. h. solche die nur eine Wohnung haben, nur einen Arbeitsplatz, die immobiler sind (z. B. Kranke) usw. Sie sind mit größerer Wahrscheinlichkeit anzutreffen, als solche, die unsteter leben und mobiler sind. Interviewer haben wie alle Menschen Schwachstellen; inwieweit bleiben sie wirklich auf der vorgeschriebenen, zufälligen Straße - der "random road"? Bestimmte seltene Merkmalskombinations-Träger sind auch für Interviewer "merk-würdig" und wenn bei einem späteren Interview wieder solche Personen gesucht werden, erinnert er sich, wo sie zu finden sind.

Interviewer könnten auch versuchen, nach bestimmten Quoten-Personen zu fragen und damit "Schneeball-System-artig" vorzugehen, statt zufällig. Ebenso besteht die Versuchung, zuerst in bekannten Revieren zu arbeiten, erst mal bei Bekannten einige Quoten abzuarbeiten. Gegen alle möglichen Interviewer-Mogeleien gibt es zwar prinzipiell entsprechende Kontroll-Möglichkeiten; doch auch Interviewer lernen dazu, lernen weitere Kontrollen zu umgehen.

Vordergründig gibt es bei der Quotenstichprobe zwar keine Stichprobenausfälle; doch es muss diskutiert werden, ob nicht doch durch die Verweigerungen von passenden Quoten-Trägern und ihren Ersatz durch andere Personen, die zwar gleiche Merkmale haben, nicht doch Schiefen in die Stichprobe kommen. Beide Personengruppen können sich durch weitere Merkmale, die nicht als Quote herangezogen werden, unterscheiden; ein offensichtlicher Unterschied liegt in ihrer verschiedenen Mitmachbereitschaft.

Die beiden prinzipiellen Arten von Stichproben (Abb. 31) haben also jeweils Vor- und Nachteile. Die zweckbezogene Stichprobe ist dann die einzige Möglichkeit, wenn die Grundgesamtheit entweder nicht bekannt oder nicht zugänglich ist. Wahrscheinlichkeits-Stichproben sind prinzipiell genauer, aber dafür aufwendiger und teurer. Die Entscheidung für eine Wahrscheinlichkeits-Stichprobe wird durch folgende Gesichtspunkte begünstigt:

- Erhebung von spezifischen, sensiblen und unbekanntem Merkmalsbereichen; bei Einthemen-Befragungen, bei Hypothesen-Testungen, wie z. B. für Einstellungen zu gesunder Ernährung, dem Körper-Image usw.;
- bei hoher Heterogenität der Merkmale innerhalb der Grundgesamtheit;
- bei hoher Bedeutung der Genauigkeit und Richtigkeit der Erhebungsdaten; Vermeidung von Selektionseffekten; exakter Berechnung der Stichprobenfehler, usw.;
- bei ungenügender Erfahrung mit Quotenstichproben; wenn noch keine entsprechende Interviewer-Organisation aufgebaut ist.

Die "billigeren" Quoten-Stichproben bieten sich dann an, wenn bekannte, robuste und eher neutrale Themen erfasst werden. Ebenso gilt dies für Mehrthemenumfragen, d. h. solchen mit mehreren, begrenzten Gebieten, die offenbar als nicht zusammengehörig erscheinen, das Forschungsziel sozusagen eher verschleiern. Das können tatsächlich auch solche Erhebungen, die im Auftrag von vielen Kunden durchgeführt werden, und die "Omnibus-Befragungen" genannt werden. Besteht bereits ein gut organisiertes Interviewer-Netz, dann begünstigt dies die Entscheidung für eine Quoten-Stichprobe.

Bei ernährungsepidemiologischen Studien spielt der Zeitfaktor eine wichtige Rolle; so ist es in longitudinalen Studien nötig an bestimmten Stichproben zeitliche Veränderungen zu untersuchen. Hierfür gibt es im Prinzip zwei Möglichkeiten. Es wird einmal eine Stichprobe gezogen und an dieser Personengruppe werden mehrmals entsprechende Untersuchungen

durchgeführt. Der andere Weg ist der, dass in der gleichen Grundgesamtheit mehrmals gleichartige Stichproben gezogen werden. Für die Bezeichnung solcher Stichproben herrscht ein Begriffe-Wirrwarr (s. S. 85). Im medizinisch-epidemiologischen Bereich gibt es für die zeitliche Verfolgung von Daten an einer Stichprobe den Begriff der "Kohorte", der Name "Panel" ist in diesem Zusammenhang ungebräuchlich. Dieser ist jedoch gängig in der sozial-empirischen Forschung.

Im Rahmen der Ernährungsepidemiologie können Kohorten und Panels fast synonym benutzt werden; allerdings sollte die übliche Bezeichnung Panel sein und der Begriff Kohorte sollte nur dann benutzt werden, wenn dabei Altersgruppen-strukturierte Stichproben benutzt werden. Der Begriff "Kohorte" wird in der allgemeinen Epidemiologie dann verwendet, wenn Gruppen gleicher Geburtsjahrgänge über einen bestimmten Zeitraum hinweg wiederholt untersucht werden, um Zusammenhänge zwischen den Vorgängen des biologischen Alterns und dem Krankheitsgeschehen zu untersuchen. Dabei werden die Alterungs- bzw. Generationseffekte bestimmt und nicht die Zeitwirkungen der sich verändernden Umwelt.

Kohorten bzw. Panels haben Vor- und Nachteile. Die Vorteile liegen darin, dass von identischen Erhebungseinheiten mehrmals Datensätze vorliegen und diese untereinander besser zu vergleichen sind. Bei Daten von einer Person zu verschiedenen Zeitpunkten können die relativen Veränderungen besser bestimmt werden; die Daten dienen praktisch untereinander der eigenen Kontrolle. Untersucher und Studienteilnehmer werden mit der Zeit besser miteinander bekannt, das erleichtert die Organisation der Studie. Das bedeutet gleichzeitig aber auch einen Nachteil, denn die zu untersuchenden Personen werden durch die Verlaufsstudie stärker beeinflusst (s. S. 85). Sie lernen und reagieren (HAWTHORNE-Effekt); sie unterscheiden sich mit der Zeit von der Grundgesamtheit aus der sie stammen.

Durch diesen "Panel-Effekt" verlieren sie an ihrer Repräsentativität. Wiederholte Erhebungen wirken ermüdend und dämpfen die Bereitschaft an weiterer Mitarbeit. Das Panel wird bei jeder neuen Erhebungsrunde kleiner und bedarf so besonderer Panelpflege; dadurch kann die "Panel-Mortalität" zwar geringer gehalten werden, doch dies bedeutet eine verstärkte Intervention bzw. einen erhöhten Panel-Effekt. Schließlich haben Panels Probleme bezüglich der Anonymität einer Untersuchung. Die Studienteilnehmer erhalten die Vergewisserung der Untersucher, dass ihre Daten anonym verarbeitet werden; doch sie wissen auch, dass ihre Teilnehmer-Nummer jederzeit mit ihrer Personen verbunden wird und dies aus organisatorischen Gründen auch notwendig ist, denn wie sonst sollten, sie zur nächsten Untersuchungsrunde wieder angesprochen werden.

Die Alternative zu einer Panel-Stichprobe besteht darin, dass die gleichen Erhebungsinstrumente an gleichartig gezogenen Stichproben, die sich dann aber aus verschiedenen Studienteilnehmern zusammensetzt, in zeitlicher Abfolge wiederholt eingesetzt werden. Bei diesen *Folge-Studien* ist der organisatorische Aufwand größer und die Datenvergleiche sind nicht so gut. Es können nur Gruppendaten und nicht individuelle Daten miteinander verglichen werden. Es können auch beide Arten von Studien miteinander kombiniert werden, indem z. B. bei jeder Wiederholung ein Anteil von neuen - aus der gleichen Grundgesamtheit ausgewählten - Studienteilnehmern neben den "alten" mit untersucht werden, von denen ein entsprechender Anteil jedoch ausscheidet ("Panel-Rotation").