

Das wissenschaftlich- religionspädagogische Lexikon im Internet

(WiReLex)

Jahrgang 2016

Stichprobe/Sampling

Wolfgang Ilg, Mirjam Rutkowski

erstellt: Februar 2021

Permanenter Link zum Artikel:

<http://www.bibelwissenschaft.de/stichwort/200848/>



DEUTSCHE
BIBEL
GESELLSCHAFT

Stichprobe/Sampling

Wolfgang Ilg, Mirjam Rutkowski

In der Religionspädagogik sind empirische Studien unverzichtbar, um ein Bild von der Realität eines Feldes, z.B. im Bereich des Religionsunterrichts, zu gewinnen. An der Frage des Samplings entscheidet sich, wie aussagekräftig eine Studie ist – dennoch bleibt dieses Thema oft wenig beachtet. Sowohl für die eigenständige Erstellung einer Studie als auch für die kritische Wahrnehmung vorliegender Untersuchungen sollen im Folgenden Grundlagen der *Stichprobenauswahl* (engl. *Sampling*) beschrieben und anhand exemplarischer Studien erläutert werden.

1. Warum Stichproben?

Empirische Forschung (→ [Empirie](#)) zielt darauf, wissenschaftlich abgesicherte Aussagen auf Grundlage nachvollziehbarer Forschungsmethoden zu gewinnen. Sowohl im quantitativen als auch im qualitativen Forschungsparadigma wird hierbei mit Stichproben gearbeitet (→ [Forschungsmethoden, religionspädagogische](#)).

Die Ziehung von Stichproben, das Sampling, ist lediglich in einem Fall verzichtbar: Solange die zu untersuchenden Gruppen klein genug sind, kann eine sogenannte *Vollerhebung* durchgeführt werden: Jede Person, die Teil dieser Gruppe ist, wird zu einem bestimmten Thema befragt. Ein Beispiel dafür sind die lokalen Erhebungen zur Konfirmandenarbeit, wie sie im Forschungsprojekt i-konf (www.konfirmandenarbeit.eu) umgesetzt werden: Eine Kirchengemeinde gibt an jede Konfirmandin und jeden Konfirmanden einen (digitalen) Fragebogen aus. Die Ergebnisse spiegeln die Stimmung der gesamten Konfi-Gruppe dieser Kirchengemeinde wider – vorausgesetzt, dass auch wirklich alle Konfirmandinnen und Konfirmanden antworten.

Will hingegen eine Landeskirche eine Rückmeldung aller Konfirmandinnen und Konfirmanden erheben, ist eine Vollerhebung aus forschungsökonomischen Gründen zu aufwendig. Für dieses Problem bietet die *Inferenzstatistik* (auch: *schließende Statistik*) Methoden an, mit denen von einer kleineren ausgewählten Gruppe, der Stichprobe, auf die Grundgesamtheit geschlossen werden kann. Die *Grundgesamtheit* bezeichnet diejenigen Einheiten, „auf die sich die Untersuchungshypothesen beziehen, wobei die Einheiten real existieren müssen“ (Gehring/Weins, 2009, 195). Die Grundgesamtheit im obigen Beispiel besteht

also aus allen Konfirmandinnen und Konfirmanden einer Landeskirche.

Stichproben finden immer dann Anwendung, wenn eine Vollerhebung der Grundgesamtheit nicht möglich oder zu kostspielig ist (Atteslander, 2010, 273). Mithilfe der Stichproben sollen gültige Aussagen zu bestimmten Eigenschaften oder Ausprägungen der Grundgesamtheit getroffen werden, wie beispielsweise zur Lesefähigkeit 15-Jähriger (PISA, Reiss/Weis/Klieme/Köller, 2019). Statt alle 15-Jährigen in Deutschland auf ihre Lesefähigkeit hin zu untersuchen, wird eine ausreichend große Gruppe ausgewählt. Auf Grundlage der Verteilung des interessierenden Merkmals in der Stichprobe, also beispielsweise der Lesekompetenz, soll auf die Verteilung dieses Merkmals in der Grundgesamtheit, nämlich unter allen 15-Jährigen in Deutschland, geschlossen werden. Die Stichprobe ist daher immer eine Teilmenge der Grundgesamtheit. Entsprechend der Ziehung zufälliger Lottozahlen aus dem gesamten Topf aller vorhandenen Zahlen wird das Zustandekommen einer Stichprobe als *Stichprobenziehung* bzw. englisch als *Sampling* (von Sample = Muster, [Stich-]Probe) bezeichnet.

Eine gute Stichprobenauswahl liegt dann vor, wenn die Verteilung des zu untersuchenden Merkmals in der Stichprobe der Verteilung in der Grundgesamtheit möglichst genau entspricht. Dies setzt voraus, dass die Grundgesamtheit bekannt ist, ebenso die Verteilung mindestens einiger Merkmale, also beispielsweise Alter, Schulart und Konfession der zu befragenden Jugendlichen.

2. Qualität einer Stichprobe

Damit eine Stichprobe es zulässt, Schlüsse über Merkmalsverteilungen zu ziehen, muss sie einige Qualitätskriterien erfüllen.

2.1. Populationskonzepte

In der Praxis der Stichprobenziehung muss zwischen Grundgesamtheit und Auswahlgesamtheit unterschieden werden. Für PISA 2012 galten beispielsweise alle 15-jährigen Schülerinnen und Schüler als Grundgesamtheit, als Auswahlkriterium galt das Geburtsjahr. Dennoch konnte die Auswahl für die Stichprobe nicht aus der tatsächlichen Grundgesamtheit erfolgen, da nicht alle 15-jährigen ermittelt werden konnten (Tachtsoglou/König, 2017, 260-261). Die Stichprobe wurde daher aus der *Auswahlgesamtheit* gezogen. Damit ist die Population gemeint, die eine Chance hat, in die Stichprobe zu gelangen.

Eine Differenz zwischen Grundgesamtheit und Auswahlgesamtheit kann zu zwei unterschiedlichen verzerrenden Effekten führen: Wenn Personen der Grundgesamtheit nicht in der Auswahlgesamtheit vorkommen (beispielsweise, weil sie durch Datenschutzbestimmungen nicht erreichbar sind), spricht man von

einer Untererfassung (engl. *Undercoverage*). Befinden sich allerdings Personen in der Auswahlgesamtheit, die nicht in der Grundgesamtheit zu finden sind (beispielsweise Jugendliche, deren Alter fälschlich mit 15 angegeben wurde), spricht man von einer Übererfassung (engl. *Overcoverage*) (Diekmann, 2013, 377-378). Die Qualität einer Stichprobe ist also dann hoch, wenn sie wenig von Over- bzw. Undercoverage betroffen ist.

2.2. Rücklaufquote/Response Rate

Ein weiterer Indikator für die Qualität einer Stichprobe ist die *Rücklaufquote* (auch *Ausschöpfungsrate* oder *Response Rate*). Diese beschreibt den prozentualen Anteil von Personen der Auswahlgesamtheit, die tatsächlich an einer Befragung teilnehmen. Sofern die Rücklaufquote bei nahezu 100% liegt, ist das ideal. Allerdings sind solche Quoten eher die Ausnahme und nur mithilfe einer engen Begleitung einer Studie zu erreichen (beispielsweise mehrfache Ansprache, Kombination mehrerer Befragungsmodi, Unterstützung der Befragung durch Vorgesetzte oder Gewährsleute, Bereitstellung von Belohnungen, sogenannter *incentives*). Häufig bleiben Befragungen, insbesondere, wenn diese online durchgeführt werden, bei einem Rücklauf von unter 20% der Auswahlgesamtheit.

Das Ausbleiben einer erfolgreichen Befragung (engl. *Non-Response*) kann unterschiedliche Gründe haben. Es wird dabei zwischen zufällig auftretendem Ausbleiben und systematischem Ausbleiben unterschieden. Zufälliges Ausbleiben könnte beispielsweise Nicht-Erreichbarkeit oder Verweigern sein. Im Unterschied dazu zeichnen sich systematische Ausfälle dadurch aus, dass der Grund für den Ausfall mit der Untersuchungsmethodik in systematischem Zusammenhang steht (Schnell/Hill/Esser, 2013, 458). Beispielsweise würden bei einer Telefonbefragung mithilfe des Telefonbuchs diejenigen Personen systematisch ausgeschlossen, deren Nummern nicht im Telefonbuch stehen. Die Befragungsergebnisse sind in einem solchen Fall nicht auf die Auswahlgesamtheit verallgemeinerbar (Döring/Bortz, 2016, 295).

In der Regel ist es leichter, eine hohe Rücklaufquote zu erhalten, wenn die Befragung in einem institutionalisierten Umfeld durchgeführt werden kann, sodass die Teilnehmenden in einem bestimmten Organisationsrahmen die Umfrage ausfüllen (müssen), beispielsweise im Schulunterricht. Ein Beispiel dafür ist die Studie „Jugend – Glaube – Religion“ (Schweitzer/Wissner/Bohner/Nowack/Gronover/Boschki, 2018, 15-16), bei der durch die Erhebung im Klassenverband eine Rücklaufquote von 70% erreicht werden konnte. Ein Beispiel für eine hohe Response Rate durch intensive Begleitaktivitäten stellt das Forschungsprojekt „Jugend zählt“ dar, bei dem ein Rücklauf von 85% erreicht werden konnte. Damit die gut 2400 angeschriebenen Kirchengemeinden und Jugendverbände ihre Daten zu Jugendangeboten wie gewünscht online erfassten, wurden hierbei von einem Servicebüro 850 Telefonate und 750 individuelle E-Mails abgewickelt (Ilg/Heinzmann/Cares, 2014,

27;337-339).

Umfragen, die auf rein freiwilliger Basis geplant sind, weisen hingegen häufiger niedrigere Rücklaufquoten auf, weil die zu Befragenden schlechter erreicht werden können oder sich die Zeit für die Befragung nicht nehmen. Ein Beispiel dafür ist eine Erhebung beim Jugendkirchentag in Darmstadt 2014, bei der von ca. 4000 Teilnehmenden lediglich 287 (= 7%) befragt werden konnten, weil die Jugendlichen während der Veranstaltung nur mit einem hohen Personalaufwand zu erreichen waren (Gennerich, 2018, 63-64). Sofern die Ausfallwahrscheinlichkeit unsystematisch verteilt ist, weil bei der hier untersuchten Veranstaltung ein zufälliger Teil der Anwesenden befragt wird, kann auch mit geringen Rücklaufquoten gearbeitet werden. Problematisch sind dagegen systematische Ausfälle, die dazu führen, dass Personen mit einer bestimmten Haltung eher an einer Befragung teilnehmen als andere. So zeigen Kreitzscheck und Haensch (2019) auf, dass bei der V. Kirchenmitgliedschaftsuntersuchung (Bedford-Strohm/Jung, 2015) nur jede/r dritte Befragte zu einem Interview bereit war – und dass mit hoher Wahrscheinlichkeit ein *Non-Response Bias* vorliegt, also kirchenkritische Personen sehr viel seltener geantwortet haben als Menschen mit hoher Kirchenbindung. Solche methodischen Probleme können die Ergebnisse der V. KMU massiv verzerren, werden aber in der kirchlichen Rezeption kaum diskutiert (dazu auch Ilg, 2017).

2.3. Repräsentativität

Soll in einer Untersuchung ein Abbild der Grundgesamtheit erhoben werden, ist die *Repräsentativität* der Stichprobe entscheidend. Obwohl die meisten religionspädagogischen Studien diesen Anspruch erheben, werden die Voraussetzungen dafür oftmals proklamiert, ohne dass sie belegt werden. Dies liegt auch daran, dass das Konzept der Repräsentativität eher ein populärwissenschaftliches Konstrukt als eine statistisch klar zu definierende Größe darstellt (Häder/Häder, 2019, 337).

Grundsätzlich ist Repräsentativität dann erreicht, wenn die Stichprobe ein exaktes „Miniaturlbild der Population“ darstellt (Döring/Bortz, 2016, 298). Ist die Stichprobe ein solches verkleinertes Abbild der Grundgesamtheit über nahezu alle Merkmale hinweg, wird von einer *global repräsentativen* Stichprobe gesprochen. Diese Anforderung kann nur mit sehr komplexen und aufwändigen Stichprobenverfahren erreicht werden. Eine Annäherung daran wird beispielsweise von professionellen Meinungsforschungsinstituten angeboten, die regelmäßig mit sogenannten *Omnibus-Befragungen* eine repräsentative Auswahl von Personen zu einer Vielzahl von Themen (daher das Stichwort „Omnibus“) befragen.

Entspricht eine Stichprobe hingegen nur in ausgewählten Merkmalen der

Grundgesamtheit, ist sie *spezifisch repräsentativ*. Dieser zweite Fall ist für die religionspädagogische Forschung sehr viel üblicher. Die jeweiligen Merkmale hängen von der Forschungsfrage ab. So könnte eine Studie beispielsweise die Repräsentativität der befragten Lehrkräfte im Blick auf Alter, Schulart und Konfession sicherstellen, andere Variablen wie Familienstand oder politische Präferenzen aber außer Acht lassen, weil diese für die untersuchte Thematik als nicht relevant angesehen werden.

Für die Repräsentativität einer Stichprobe ist auch eine gewisse *Stichprobengröße* entscheidend. Allerdings spielt dieser Stichprobenumfang im Vergleich zur Stichprobenauswahl – anders als von vielen Rezipienten fälschlich angenommen – keine hervorgehobene Rolle; er ist eine notwendige, aber keinesfalls hinreichende Bedingung für Repräsentativität. Selbst sehr große Stichproben können aufgrund einer Verzerrung (engl. *bias*) für falsche Ergebnisse sorgen. Bekannt ist das Beispiel einer Stichprobe von über 2 Millionen Wahlberechtigten in den USA, auf deren Basis 1936 für den demokratischen Spitzenkandidaten Franklin Roosevelt lediglich 43% Zustimmung vorausgesagt wurden – bei der Wahl siegte er jedoch deutlich mit 62% der Stimmen. Grund für die fehlerhafte Prognose war, dass die Stichprobe u.a. über Adressen in Telefonbüchern rekrutiert worden war. Aufgrund der geringen Verbreitung von Telefonapparaten zur damaligen Zeit waren aber hauptsächlich Angehörige der oberen sozialen Schichten in Telefonbüchern aufgeführt – und diese wählten Roosevelt deutlich seltener als die Gesamtbevölkerung (Döring/Bortz, 2016, 295). Eine Befragung mit den für bevölkerungsrepräsentative Umfragen üblichen 1.000 bis 2.000 Befragten kann Wahlpräferenzen präzise vorhersagen, wenn sie repräsentativ ausgewählt ist. Dies gilt in analoger Weise auch für Studien im religionspädagogischen Bereich.

Eine sehr große Stichprobe mit deutlich mehr als 2.000 Befragten ist aus forschungsökonomischer Sicht beispielsweise dann sinnvoll, wenn der tolerierte Stichprobenfehler (beispielsweise bei Studien im medizinischen oder epidemiologischen Kontext) sehr gering sein muss (Häder/Häder, 2019, 336-339) oder wenn repräsentative Ergebnisse auch für Teil-Grundgesamtheiten bereitgestellt werden sollen. Ein Beispiel hierfür sind die Studien zur Konfirmandenarbeit, bei denen jeweils etwa 10.000 Konfirmandinnen und Konfirmanden in Deutschland befragt wurden, weil die Ergebnisse nicht nur auf Ebene der Landeskirchen, sondern auch auf EKD-Ebene ausgewertet wurden (Schweitzer/Maaß/Lißmann/Hardecker/Ilg, 2015, 278-281).

3. Stichprobenverfahren

Für die Ziehung einer Stichprobe stehen grundsätzlich zwei Vorgehensweisen zur Verfügung: Vorzugswürdig, aber aufwändig in der Umsetzung ist die zufällige Auswahl der Stichprobe. Deutlich verbreiteter ist die einfachere, aber

fehleranfällige nicht-zufällige Auswahl.

3.1. Zufällige Auswahl (probabilistisches Verfahren)

Die *Zufallsstichprobe* (engl. *random sample*) ist diejenige Stichprobe, auf deren Basis Methoden der Inferenzstatistik, z.B. *Signifikanzberechnungen*, eingesetzt werden können. Für jede Form der zufälligen Auswahl gilt, dass die Grundgesamtheit bekannt sein muss, also beispielsweise eine Liste der Pfarrerinnen und Pfarrer in einer Landeskirche vorliegen muss, wenn eine Aussage über die Pfarrerschaft getroffen werden soll. Bei der *einfachen Zufallsstichprobe* hat jedes Element der Grundgesamtheit vor dem Auswahlprozess die gleiche Chance, in die Stichprobe aufgenommen zu werden. Eine Stichprobenauswahl erfolgt nach dem Prinzip „Ziehen ohne zurücklegen“: Eine Person, die aus der Grundgesamtheit ausgewählt wurde, kann nicht nochmals gezogen werden. Dieses Prinzip wird auch beim Ziehen der Lotto-Zahlen angewandt (Fahrmeir/Heumann/Künstler/Pigeot/Tutz, 2016, 23). Bei der *systematischen Zufallsstichprobe* werden alle Elemente einer Grundgesamtheit nach einem bestimmten Kriterium gelistet, beispielsweise alle Pfarrerinnen und Pfarrer der Landeskirche in alphabetischer Reihenfolge. Aus dieser Liste werden nun nach einer bestimmten Regel Elemente ausgewählt, beispielsweise jede dritte Person (Tachtsoglou/König, 2017, 263).

Für eine *geschichtete Zufallsstichprobe* (engl. *stratified random sample*) wird die Grundgesamtheit nach bestimmten Kriterien, z.B. Kirchenkreis, Geschlecht, Gemeinde- oder Funktionspfarramt in verschiedene Gruppen (auch Strata) eingeteilt. Diese Kriterien sollten mit der interessierenden Ausprägung hoch korrelieren. Von jeder Teilgruppe wird eine Zufallsstichprobe gezogen. Für eine proportionale Stichprobe kämen von jeder Gruppe gleich viele gezogene Elemente in die Stichprobe. Für eine disproportionale Stichprobe werden bestimmte Gruppen unter- bzw. übergewichtet. Von diesen Gruppen werden also weniger bzw. mehr Elemente als aus den anderen Gruppen in die Stichprobe aufgenommen. Dieses Verfahren ist beispielsweise dann interessant, wenn die Strata unterschiedliche Größen aufweisen (Fahrmeir/Heumann/Künstler/Pigeot/Tutz, 2016, 23; Tachtsoglou/König, 2017, 264). Bei der Berechnung der Gesamtergebnisse werden solche Disproportionalitäten durch entsprechende Gewichtungsfaktoren statistisch wieder ausgeglichen.

Bei der *Klumpen- oder Clusterstichprobe* (engl. *cluster sample*) haben die einzelnen Personen der Grundgesamtheit im Gegensatz zur Zufallsstichprobe nicht mehr dieselbe Chance, in die Stichprobe aufgenommen zu werden. Die Clusterstichprobe wird häufig bei Grundgesamtheiten angewandt, die in sich selbst in kleinere Gruppen geordnet sind. Es werden zunächst einzelne Cluster ausgewählt, dann werden alle Elemente der ausgewählten Cluster in die Stichprobe aufgenommen. Zu beachten ist jedoch, dass sich die Elemente

innerhalb eines Clusters häufig ähnlicher sind als die Elemente unterschiedlicher Cluster, was die Verallgemeinerbarkeit der Ergebnisse beeinträchtigt (Diekmann, 2013, 387-388). Ein Anwendungsbeispiel aus der Jugendarbeit ist die Panelstudie zu Jugendfreizeiten und internationalen Jugendbegegnungen: In einem ersten Schritt wird eine repräsentative Auswahl von Trägern definiert, von denen jeweils alle Mitarbeitenden und Teilnehmenden der entsprechenden Jugendgruppenfahrten eine Befragung durchlaufen. Die jährliche Erhebung bei einem unveränderlichen Panel schafft dabei die Voraussetzung für vergleichbare Ergebnisse über die Jahre hinweg (vgl. www.panelstudie.de).

Bei *mehrstufigen Zufallsstichproben (multi-stage random sample)* werden mehrere der genannten Verfahren kombiniert. Häufig wird aus der Grundgesamtheit zunächst eine Teilmenge gezogen (z.B. eine geschichtete Zufallsstichprobe oder eine Klumpenstichprobe), woraufhin einzelne Elemente aus der Teilmenge über die einfache bzw. systematische Stichprobe ausgewählt werden. Die Auswahl der Schülerinnen und Schüler in Deutschland, die an PISA 2018 teilnahmen, erfolgte vereinfacht anhand einer mehrstufigen Zufallsstichprobe: Zunächst wurden auf der ersten Stufe 234 Schulen ausgewählt. Auf der zweiten Stufe wurden an diesen Schulen mittels Zufallsstichproben 15-Jährige ausgewählt (Reiss/Weis/Klieme/Köller, 2019, 17).

3.2. Nicht-zufällige bzw. bewusste Auswahl (nicht-probabilistisches Verfahren)

Neben den unterschiedlichen Formen der zufälligen Stichprobenziehung werden in der quantitativen und qualitativen Forschung auch nicht-probabilistische Stichprobenverfahren angewandt. Zufällige Stichproben sind mit hohem Aufwand verbunden, wofür in der Forschungspraxis häufig die Ressourcen fehlen. Um dennoch Forschungsfelder untersuchen zu können, werden pragmatischere Verfahren durchgeführt, mit der Überzeugung, dass eine schwache Datenbasis immer noch besser sei als keine empirische Basis (Döring/Bortz, 2016, 307; Diekmann, 2013, 410). In religionspädagogischen Studien dominieren solche Sampling-Methoden aufgrund ihrer einfacheren Umsetzung, allerdings wird die Problematik solcher Stichprobenziehungen in den Publikationen oftmals zu wenig diskutiert.

Eine nicht-zufällige Form des Samplings bietet die *willkürliche Stichprobe* (auch *Gelegenheitsstichprobe* oder *Ad-hoc-Stichprobe*). Wie der Name sagt, erfolgt die Auswahl einzelner Elemente für die Stichprobe willkürlich: Häufig wird auf die Personen zurückgegriffen, die am einfachsten erreicht werden können. Gerade an Hochschulen kommt die willkürliche Stichprobenziehung häufig zum Einsatz: Studentische Versuchspersonen können hier einfach akquiriert werden. Eine Sonderform der willkürlichen Stichprobe ist die *Selbstselektions-Stichprobe*. Potenzielle Teilnehmende werden hierbei durch einen allgemeinen Aufruf gewonnen. Die Selbstselektion entsteht dadurch, dass die Teilnahme von der

Eigeninitiative der Personen abhängt – womit meist eine Verzerrung einhergeht, weil die einem Thema besonders verbundenen Personen zur Beantwortung deutlich motivierter sind als andere.

Häufig werden die durch willkürliche Auswahl getroffenen Stichproben fälschlicherweise als zufällige Stichproben dargestellt. Dass dies keineswegs zutrifft, veranschaulicht folgende Überlegung: Werden für eine Elternbefragung in einer kirchlichen Kindertagesstätte eine bestimmte Anzahl an Personen beim Abholen der Kinder befragt, ergibt sich kein repräsentatives Meinungsbild der Elternschaft, denn die Elternteile haben verschieden hohe Wahrscheinlichkeiten, Teil des Samples zu sein. Systematisch unterrepräsentiert wären bei einer solchen Stichprobenziehung beispielsweise arbeitstätige Eltern sowie Eltern von Vorschulkindern, die den Weg nach Hause allein antreten können. Die Auswahlgesamtheit der Stichprobe liegt bei diesem Verfahren also nicht vor der Befragung fest, sondern wird erst durch die Auswahl selbst festgelegt (Döring/Bortz, 2016, 306). Wie problematisch die Ergebnisse von Studien mit einem willkürlichen Sampling sein können, wird beispielsweise an der sogenannten Kirchgangsstudie deutlich (Liturgische Konferenz, 2019). Die zunächst erfreulich hohe Zahl von über 10.000 Befragten wird durch die massiven Effekte der Selbstselektion konterkariert: So handelt es sich bei 22% der Befragten um beruflich bei der Kirche Beschäftigte – zu einem guten Teil wohl Pfarrerinnen und Pfarrer (Liturgische Konferenz, 2019, 7). Solche Ergebnisse auf Gottesdienstbesuchende insgesamt zu verallgemeinern führt zu einer massiven Schiefelage, die in den bisherigen Interpretationen der Studie kaum thematisiert wird. Ebenfalls mit einer Gelegenheitsstichprobe arbeitet die Studie „Generation Lobpreis“, bei der hochreligiöse Jugendliche im Rahmen christlicher Großveranstaltungen sowie ergänzend online befragt wurden (Faix/Künkler, 2018a). Der online bereitgestellte Forschungsbericht zu dieser Studie stellt eines der erfreulichen Beispiele für eine transparente Dokumentation auch methodischer Erwägungen zu Samplingfragen dar, was eine kritische Einordnung der Ergebnisse ermöglicht (Faix/Künkler, 2018b, 31-37).

Um den Vorwurf der Willkür abzuschwächen, gibt es Verfahren, welche die willkürliche Auswahl systematisieren. Zu diesen Formen zählt die *Quotenstichprobe* (auch *Quota-Verfahren* bzw. *quota sampling*). Der Auswahlprozess einzelner Elemente zur Stichprobe wird von der interviewenden Person bewusst systematisch vollzogen, um eine spezifische Repräsentativität der Stichprobe zu erreichen, zumeist nach sozio-demographischen Merkmalen. Dieses Verfahren setzt also Kenntnisse über die Grundgesamtheit voraus (Diekmann, 2013, 391). Anhand der Verteilung der interessierenden Merkmale in der Grundgesamtheit werden Quoten gebildet, die bei den Befragungen erfüllt werden sollen. Im Gegensatz zur geschichteten Zufallsstichprobe, die ebenfalls gewisse Merkmalsverteilungen abbilden soll, kann die Wahrscheinlichkeit einer Person, in die Stichprobe zu gelangen, nicht vor der Ziehung angegeben werden.

Bei der Quotenstichprobe hängt diese Wahrscheinlichkeit einer Person von der Teilnahmebereitschaft vorangegangener Fälle ab (Atteslander, 2010, 278-277). In der Jugendforschung arbeitet beispielsweise die Shell-Jugendstudie mit einer Quotenstichprobe, in der die Quotierungsmerkmale Alter, sozialer Status, Bundesland und Siedlungsstruktur sowie Personen mit/ohne Migrationshintergrund vorgegeben sind (Deutsche Shell, 2019, 325-332). Sofern die Interviewerinnen und Interviewer den vorgegebenen Quotenplan erfüllen, sind sie frei, beliebige Jugendlichen für die Studie zu rekrutieren – was dazu führen dürfte, dass schüchterne Menschen in Quotenstichproben eher unterrepräsentiert sind.

Bei speziellen Grundgesamtheiten kann es sein, dass zufällige Auswahlen kaum realisierbar wären. Das ist dann der Fall, wenn die Zahl der betreffenden Personen sehr gering ist und sich das Auffinden entsprechender Personen als schwierig gestaltet – beispielsweise, wenn Jugendliche mit einem bestimmten Krankheitsbild Auskunft zu ihrer Religiosität geben sollen. Am häufigsten wird dann das *Schneeballverfahren* eingesetzt, das bei einer untereinander gut vernetzten Population geeignet ist: Sobald die erste Person gefunden ist, kann diese über bestehende Vernetzungen, beispielsweise entsprechende Internetforen, weitere Zielpersonen ansprechen. Der Name leitet sich von einem im Schnee rollenden Schneeball ab, der anfangs langsam, später schneller an Größe gewinnt (Diekmann, 2013, 400; Döring/Bortz, 2016, 308).

Ein anderes Verfahren bei Grundgesamtheiten, die weniger gut vernetzt sind, beispielsweise bei Drogenkonsumenten, ist der *Targeted-Sampling-Ansatz*: Hier werden typische Aufenthaltsorte der interessierenden Personen ausgemacht und ein Stichprobenplan erstellt, der angibt, wie viele Personen pro Treffpunkt zu rekrutieren sind (Döring/Bortz, 2016, 308).

4. Besonderheiten in der qualitativen Forschung

Qualitative Studien zielen, anders als die oben beschriebenen quantitativen Studien, in der Regel nicht auf repräsentative Ergebnisse, die sich für eine Grundgesamtheit quantifizieren lassen (→ [qualitative Sozialforschung](#)). Vielmehr besteht ihr besonderer Forschungsbeitrag in der tiefergehenden Erkundung individueller Wahrnehmungen und Zusammenhänge. Für das Sampling bei qualitativen Studien wird eine bewusste Auswahl von Fällen angewendet, wobei drei Herangehensweisen unterschieden werden können: Bei der *Theoretischen Stichprobenbildung* (engl. *theoretical sampling*) wird das Sample nicht zu Beginn der Untersuchung festgelegt, sondern ergibt sich in einem iterativen Prozess auf Grundlage der bisherigen Ergebnisse, bis eine *theoretische Sättigung* erreicht wird (→ [Grounded Theory](#)). Eine andere Methode besteht, analog zur oben

dargestellten Quotenstichprobe, in der Verwendung eines qualitativen Stichprobenplans (engl. *qualitative sampling plan*). Für eine Befragung kirchlicher Jugendreferentinnen und Jugendreferenten könnte hierbei beispielsweise festgelegt werden, dass die Kriterien Geschlecht, Ausbildungsstätte und Anstellungsträger berücksichtigt werden und für jede Zelle zwei Personen befragt werden sollen. Eine dritte Herangehensweise bietet sich an, wenn eine sehr spezielle Zielgruppe untersucht werden soll, beispielsweise junge Menschen, die als Jugendsynodale tätig sind. Die Stichprobe muss in solchen Fällen aufgrund der geringen Fallzahl klein bleiben und gibt aufgrund der Thematik einen bestimmten Rekrutierungsweg oftmals vor.

Neben der Stichprobenziehung bei der *Auswahl der Befragten* müssen bei qualitativen Studien auch zwei weitere Auswahllebenen berücksichtigt werden. Angesichts oft umfangreicher Interviewmitschnitte kann es forschungsökonomisch geboten sein, nur einen Teil des Gesagten zu transkribieren. Hier findet also eine Stichprobenziehung bei der *Datenauswertung* statt. So wurden bei der oben genannten Kirchgangsstudie nur ein Bruchteil der offenen Antworten codiert (Liturgische Konferenz, 2019, 20). Eine ähnliche Selektionsaufgabe ergibt sich für die *Datenpräsentation*, wenn beispielsweise die Ergebnisse von hunderten Seiten transkribiertem Text in einem Vortrag oder Fachartikel zusammengefasst werden sollen (Akremi, 2019, 313-315).

Angesichts dieser besonderen Herausforderungen wird deutlich, warum qualitative Forschung nicht auf verallgemeinerbare Ergebnisse im Sinne repräsentativer Aussagen zielen kann. Am Beispiel der in Deutschland gängigen Jugendstudien kann gezeigt werden, dass solche forschungsmethodischen Fragen in der Rezeption kaum reflektiert werden. So werden inzwischen die qualitativen Jugendstudien des Sinus-Instituts (mit lediglich 72 Befragten im Bundesgebiet) medial oft gleichrangig wie die etablierten quantitativen Shell-Studien (mit ca. 2.500 Befragten) fälschlicherweise als Repräsentativstudien für Jugendliche in Deutschland wahrgenommen. Dazu trägt auch die unpräzise Formulierung der Sinus-Autoren bei, die Ergebnisse seien zwar „nicht im statistischen, wohl aber im psychologischen Sinne repräsentativ“ (Calmbach/Flaig/Edwards/Möller/Borchard/Schleer, 2020, 22; für den evangelischen Bereich auch Kopp/Hügin/Kaup/Borchard/Calmbach, 2013).

5. Fazit

Die Stichprobenziehung gehört bei empirischen Untersuchungen zu den zentralen Stellschrauben für die Aussagekraft der Ergebnisse. Viele Rezipienten und Rezipientinnen religionspädagogischer Studien verhalten sich im Bild gesprochen jedoch wie der Käufer eines Gebrauchtwagens, der sich von der blank polierten Karosserie verzaubern lässt und dabei den Blick unter die Motorhaube vergisst. Weil diese oberflächliche Rezeption von Studien sich an

vielen Stellen leider etabliert hat, wird bei der Darstellung von Studien zunehmend auf eine transparente Darstellung des Samples verzichtet.

Die im vorliegenden Artikel dargestellten Grundlagen des Samplings bieten Interessierten wichtiges Handwerkszeug für einen kritischen Blick auf Untersuchungsergebnisse. Zugleich können sie bei der eigenen Durchführung von Studien dazu motivieren, das Sampling mit der dafür notwendigen Detailtiefe durchzuführen. Insbesondere der Versuchung, auf die einfach zu gewinnenden Gelegenheitsstichproben zu setzen, sollte nach Möglichkeit widerstanden werden, um sich so weit als möglich einer zufälligen Stichprobenauswahl anzunähern.

[Angaben zu Autor / Autorin finden Sie hier](#)

Empfohlene Zitierweise

Ilg, Wolfgang, Rutkowski, Mirjam, Art. Stichprobe/Sampling, in: Wissenschaftlich Religionspädagogisches Lexikon im Internet (www.wirelex.de), 2021

Literaturverzeichnis

- Akremi, Leila, Stichprobenziehung in der qualitativen Sozialforschung, in: Baur, Nina/Blasius, Jörg (Hg.), **Handbuch Methoden der empirischen Sozialforschung**, Wiesbaden 2. Aufl. 2019, 313-331.
- Atteslander, Peter, Methoden der empirischen Sozialforschung, Berlin 13. Aufl. 2010.
- Bedford-Strohm, Heinrich/Jung, Volker (Hg.), Vernetzte Vielfalt. Kirche angesichts von Individualisierung und Säkularisierung. Die fünfte EKD-Erhebung über Kirchenmitgliedschaft, Gütersloh 2015. Online unter: https://www.ekd.de/download/20151120_kmu_v_auswertungsband.pdf, abgerufen am 19.08.2020.
- Calmbach, Marc/Flaig, Bodo/Edwards, James/Möller, Slawinski/Borchard, Inge/Schleer, Christoph, Wie ticken Jugendliche? SINUS-Jugendstudie 2020. Lebenswelten von Jugendlichen im Alter von 14 bis 17 Jahren in Deutschland. Bonn: Bundeszentrale für politische Bildung, Bonn 2020. Online unter: <https://www.bpb.de/shop/buecher/einzelpublikationen/311857/sinus-jugendstudie-2020-wie-ticken-jugendliche>, abgerufen am 19.08.2020.
- Deutsche Shell (Hg.), Jugend 2019. Eine Generation meldet sich zu Wort. 18. Shell Jugendstudie, Weinheim 2019.
- Diekmann, Andreas, Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen, Reinbeck bei Hamburg 2013.
- Döring, Nicola/Bortz, Jürgen, **Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften**, Berlin/Heidelberg 5. Aufl. 2016.
- Fahrmeir, Ludwig/Heumann, Christian/Künstler, Rita/Pigeot, Iris/Tutz, Gerhard, Statistik. Der Weg zur Datenanalyse, Berlin/Heidelberg 2016.
- Faix, Tobias/Künkler, Tobias, Generation Lobpreis und die Zukunft der Kirche. Das Buch zur empirica Jugendstudie 2018, Vluyn 2018a.
- Faix, Tobias/Künkler, Tobias, Empirica Jugendstudie 2018. Forschungsbericht, Kassel 2018b. Online unter: https://www.cvjm-hochschule.de/fileadmin/2_Dokumente/5_FORSCHUNG/empirica/2018_Jugendstudie-Forschungsbericht.pdf, abgerufen am 19.08.2020.
- Gehring, Uwe W./Weins, Cornelia, Grundkurs Statistik für Politologen und Soziologen, Wiesbaden 5. Aufl. 2009.
- Gennerich, Carsten, Lebensstile Jugendlicher. Beteiligung an Angeboten kommunaler, vereinsorganisierter und kirchlicher Jugendarbeit, Leverkusen-Opladen 2018.
- Häder, Michael, Empirische Sozialforschung: Eine Einführung. Wiesbaden 2019 (insb. S. 145-197).
- Häder, Michael/Häder, Sabine, Stichprobenziehung in der quantitativen

Sozialforschung, in: Baur, Nina/Blasius, Jörg (Hg.), Handbuch Methoden der empirischen Sozialforschung, Wiesbaden 2. Aufl. 2019, 333-348.

- Ilg, Wolfgang, Notwendige Horizonterweiterungen für die Kirchenmitgliedschaftsuntersuchungen. Möglichkeiten und Grenzen empirischer Annäherungen an gemeindepädagogische Arbeitsfelder am Beispiel der Kirchenmitgliedschaftsuntersuchungen und der Studien zur Konfirmandenarbeit, in: Zeitschrift für Pädagogik und Theologie (ZPT) 69 (2017) 4, 317-329. DOI: <https://doi.org/10.1515/zpt-2017-0036>.
- Ilg, Wolfgang/Heinzmann, Gottfried/Cares, Mike (Hg.), Jugend zählt! Ergebnisse, Herausforderungen und Perspektiven aus der Statistik 2013 zur Arbeit mit Kindern und Jugendlichen in den Evangelischen Landeskirchen Baden und Württemberg, Stuttgart 2014.
- Kopp, Hansjörg/Hügin, Stefanie/Kaupp, Stefan/Borchard, Inga/Calmbach, Marc, Brücken und Barrieren. Jugendliche auf dem Weg in die Evangelische Jugendarbeit, Stuttgart 2013.
- Kreitzscheck, Mathis/Haensch, Anna-Carolina, „Klopfet an, so wird euch aufgetan?“ Teilnahmeverweigerung und Nonresponse Bias in der fünften Kirchenmitgliedschaftsuntersuchung, in: Praktische Theologie 54 (2019) 1, 43-51.
- Kruse, Jan, Qualitative Interviewforschung: ein integrativer Ansatz. Weinheim 2. Aufl. 2015 (insb. Kapitel IV).
- Liturgische Konferenz, Kirchengangsstudie 2019. Erste Ergebnisse, Hannover 2019. Online unter https://www.liturgische-konferenz.de/download/Kirchengangsstudie_2019_Ergebnispapier.pdf, abgerufen am 19.08.2020.
- Pirner, Manfred L./Rothgangel, Martin (Hg.), Empirisch forschen in der Religionspädagogik. Ein Studienbuch für Studierende und Lehrkräfte. Stuttgart 2018.
- Reiss, Kristina/Weis, Mirjam/Klieme, Eckhard/Köller, Olaf (Hg.), PISA 2018. Grundbildung im internationalen Vergleich, Münster/New York 2019.
- Schnell, Rainer/Hill, Paul B./Esser, Elke, Methoden der empirischen Sozialforschung, München 10. Aufl. 2013.
- Schweitzer, Friedrich/Wissner, Golde/Bohner, Annette/Nowack, Rebecca/Gronover, Matthias/Boschki, Reinhold, Jugend – Glaube – Religion. Eine Repräsentativstudie zu Jugendlichen im Religions- und Ethikunterricht, Münster 2018.
- Schweitzer, Friedrich/Maaß, Christoph H./Lißmann, Katja/Hardecker, Georg/Ilg, Wolfgang, Konfirmandenarbeit im Wandel – Neue Herausforderungen und Chancen. Perspektiven aus der Zweiten Bundesweiten Studie, Konfirmandenarbeit erforschen und gestalten 6, Gütersloh 2015.
- Tachtsoglou, Sarantis/König, Johannes, Statistik für Erziehungswissenschaftlerinnen und Erziehungswissenschaftler. Konzepte, Beispiele und Anwendungen in SPSS und R, Wiesbaden 2017.

Impressum

Hauptherausgeberinnen:

Prof. Dr. Mirjam Zimmermann (Universität Siegen)

Prof. Dr. Heike Lindner (Universität Köln)

„WiReLex“ ist ein Projekt der Deutschen Bibelgesellschaft

Deutsche Bibelgesellschaft

Balinger Straße 31 A

70567 Stuttgart

Deutschland

www.bibelwissenschaft.de