



Johannes Volk

studierte Politikwissenschaften und Statistik in Bamberg. Im Statistischen Bundesamt ist er als wissenschaftlicher Mitarbeiter im Referat „Fragebogenpretest, Erhebungsmethoden“ für das Mixed-Mode-Projekt tätig.



Tim Hochgürtel

studierte Soziologie in Mainz und ist seit 2008 im Statistischen Bundesamt tätig, seit 2016 als Referent im Referat „Bevölkerungsstatistische Auswertungen und Analysen aus dem Mikrozensus“. Schwerpunkte seiner Arbeit sind Analysen sowie Publikationen zu privaten Haushalten und Familien.

DIE AUSWIRKUNG VON UNIT-NONRESPONSE IN STATISTISCHEN ERHEBUNGEN

Johannes Volk, Tim Hochgürtel

↳ **Schlüsselwörter:** Antwortausfälle – Verzerrungen – Korrekturverfahren – Gewichtung

ZUSAMMENFASSUNG

Der Unit-Nonresponse (vollständiger Antwortausfall) kann in Erhebungen einen erheblichen Einfluss auf die Qualität der Ergebnisse nehmen. Dies ist der Fall, wenn durch ihn eine verzerrte Nettostichprobe realisiert wird. Üblicherweise wird durch Gewichtungsverfahren versucht, diese Verzerrung zu korrigieren. Eine solche Korrektur setzt aber voraus, dass der Ausfallmechanismus bekannt ist und durch Variablen erklärt werden kann, welche für die Grundgesamtheit bekannt sind. Auch bei Verwendung von Gewichten können Ergebnisse noch systematische Fehler aufweisen, welche durch den Unit-Nonresponse bedingt sind.

↳ **Keywords:** unit-nonresponse – bias – adjustment methods – weighting

ABSTRACT

Unit-nonresponse can have a significant influence on the quality of survey results. This is the case if unit-nonresponse causes a biased net sample. Usually attempts are made to adjust these biases through weighting procedures. Such adjustments require non-response mechanisms that are familiar and can be explained through variables which are known for the entire survey population. Even after weighting-based adjustments, however, results can still be biased because of systematic unit-nonresponse errors.

1

Total Survey Error und Unit-Nonresponse

Bei der Durchführung von Befragungen gibt es eine Vielzahl potenzieller Fehlerquellen, welche die Qualität von Erhebungen mindern können. Nach dem Konzept des Total Survey Error begleiten diese Fehlerquellen den gesamten Prozess der Datengewinnung von der Stichprobenauswahl bis hin zum Erzeugen bereinigter Datensätze. Infolge solcher Fehler können in einer späteren Auswertung der betreffenden Erhebung die Schätzungen von Populationsparametern verzerrt sein (Faulbaum, 2014, hier: Seite 439).

Als eine Ursache für eine mögliche Verzerrung von Populationsparametern ist der Unit-Nonresponse zu nennen (Groves/Lyberg, 2010, hier: Seite 856). Ein Unit-Nonresponse kommt zustande, wenn Stichprobeneinheiten sich nicht an einer Erhebung beteiligen, obwohl sie Element der Bruttostichprobe sind.

Sofern Bruttostichproben als Zufallsstichproben realisiert werden, resultieren diese aus einem kontrollierten Auswahlprozess. Jedem Element der Grundgesamtheit wird im Rahmen des Auswahlprozesses eine Wahrscheinlichkeit zugewiesen, mit der das Element für die Bruttostichprobe berücksichtigt wird.

Die Nettostichprobe entspricht einer Teilmenge der Bruttostichprobe. Aus Sicht des Datenproduzenten kann nicht kontrolliert werden, welche Elemente der Bruttostichprobe zu Elementen der Nettostichprobe werden. Die Nettostichprobe konstituiert sich selbstselektiv, da die Elemente der Bruttostichprobe selbst darüber entscheiden, ob sie an einer Erhebung teilnehmen oder nicht.

Als ein Qualitätsindikator einer Erhebung wird üblicherweise die Responserate angegeben. Die Responserate wird als Anteil des Nettostichprobenumfangs am Bruttostichprobenumfang gebildet (Bethlehem und andere, 2011, hier: Seite 9). Der komplementäre Anteil misst damit den Unit-Nonresponse. Die Größe des Unit-Nonresponse sagt prinzipiell noch nichts darüber aus, ob Ergebnisse verzerrt sind oder nicht. Auch bei geringer Ausschöpfung können theoretisch Ergebnisse ohne Bias erreicht werden. Die Höhe der Ausschöpfung sagt aber

sehr wohl etwas über das Potenzial möglicher Verzerrungen aus, das bei einem kleinen Unit-Nonresponse geringer ist als bei einem großen (Proner, 2011, hier: Seite 50).

Der vorliegende Aufsatz geht der Frage nach, welche Schritte unternommen werden können, um den (potenziell) verzerrenden Einfluss des Unit-Nonresponse auf die Schätzung von Populationsparametern möglichst gering zu halten.

2

Unit-Nonresponse und Nonresponse-Bias

Eine Responserate kann als mittlere Teilnahmewahrscheinlichkeit eines Merkmalsträgers aus der Bruttostichprobe verstanden werden. Eine unverzerrte Nettostichprobe wäre zu erwarten, wenn jede Stichprobeneinheit der Bruttostichprobe mit einem Münzwurf über die Teilnahme an einer Erhebung entscheiden würde. Hierbei entspräche die Wahrscheinlichkeit, dass die Münze eine Teilnahme ausspielt, der Responserate. Mit der komplementären Wahrscheinlichkeit würde die Münze die Nicht-Teilnahme entscheiden. Eine solche Nettostichprobe wäre „repräsentativ“ in der Hinsicht, dass die Populationsparameter, die auf Basis einer solchen Nettostichprobe geschätzt würden, lediglich zufälligen Schwankungen unterlägen, aber keine systematischen Verzerrungen vorlägen.

Die Annahme, dass Stichprobeneinheiten ihre Teilnahme über einen Münzwurf entscheiden, ist jedoch völlig unrealistisch. Sofern die Stichprobeneinheiten erfolgreich kontaktiert werden konnten und gesundheitlich wie sprachlich zu einer Teilnahme in der Lage sind, entscheiden sie anhand eigener Präferenzstrukturen über eine Teilnahme an einer Erhebung.

In der Regel resultiert aus dieser individuellen Entscheidung der Stichprobeneinheiten der Bruttostichprobe eine Verzerrung der Nettostichprobe (Nonresponse-Bias). Damit ist das Kriterium der „Repräsentativität“ verletzt. Wenn die Teilnahmepräferenz der Merkmalsträger der Bruttostichprobe mit den Merkmalen der Erhebung korreliert ist, ist durch die Nettostichprobe keine strukturtreue Abbildung der Grundgesamtheit mehr gegeben.

Der Grad der Verzerrung der einzelnen Erhebungsmerkmale der Nettostichprobe ist in der Regel unbekannt. Eine verzerrungsbedingte Abweichung von den „wahren“ Verteilungen der Erhebungsmerkmale ist nur dann gegeben, wenn die „wahren“ Verteilungen aus anderen Quellen bekannt sind. Für die meisten Erhebungsmerkmale liegen diese Verteilungen nicht vor. Wären die Verteilungen der Erhebungsmerkmale sowie die gemeinsame Verteilung bekannt gewesen, wäre die Erhebung nicht durchgeführt worden.

Von Interesse sind aber weniger die Eigenschaften der Nettostichprobe. Die Nettopopulation ist vielmehr eine Hilfspopulation, um mit inferenzstatistischen Verfahren auf die Eigenschaften der Grundgesamtheit zu schließen.

Um aus der Nettostichprobe Populationsparameter für die Grundgesamtheit zu gewinnen, muss der aus den ungleichen Teilnahmepräferenzen der Merkmalsträger der Bruttostichprobe resultierende Bias ausgeglichen werden.

Hierzu werden üblicherweise Gewichtungsverfahren zum Einsatz gebracht. Damit werden die einzelnen Merkmalsträger der Nettostichprobe bei Analysen ungleich stark mit einbezogen. Verzerrungen, die aus den ungleichen Teilnahmepräferenzen von Merkmalsträgern der Bruttostichprobe resultieren, sollen damit ausgeglichen werden.

Der Berechnung von Gewichtungsfaktoren liegen – implizit oder explizit – Erklärungsmodelle des Ausfallmechanismus zugrunde. Hierbei werden Merkmale berücksichtigt, die das Teilnahmeverhalten der Merkmalsträger erklären sollen. Eine Gewichtung kann daher immer nur so gut sein wie das zugrunde liegende Erklärungsmodell.

3

Erklärungen des Ausfallprozesses

Damit der Einsatz von Gewichtungs- und Korrekturmethode den Nonresponse-Bias, der durch den Unit-Nonresponse entstanden ist, ausgleichen kann, ist ein umfassendes Verständnis des Ausfallprozesses notwendig. Der Ausfallprozess muss bei der Berechnung von

Gewichten modelliert werden (Proner, 2011, hier: Seite 50). Grundsätzlich muss beim Einsatz von Gewichtungsverfahren immer der Ausfallmechanismus der Stichprobeneinheit (zum Beispiel Unternehmen, Betriebe, Haushalte, Personen) erklärt werden. Im Folgenden wird beispielhaft der Ausfallprozess bei einer Befragung von Personen diskutiert.

Anhand ausgewählter Studien mit unterschiedlichen Datengrundlagen¹ lässt sich zeigen, dass Ausfälle mit bestimmten soziodemografischen Merkmalen, wie Alter, Geschlecht, Bildung, Erwerbsstatus, korrelieren.

Es gibt auch Erklärungsversuche, die andere soziodemografische Eigenschaften von Personen für das Teilnahmeverhalten an Erhebungen heranziehen. Hierbei wird etwa das Teilnahmeverhalten abhängig von der sozialen Partizipation, der zur Verfügung stehenden Zeit und der Lebensweise der Personen gedeutet (Stoop und andere, 2010, hier: Seite 126 f.). Doch auch wenn es in Umfragen relativ stabile Korrelationen von Teilnahmebereitschaft und soziodemografischen Eigenschaften von Personen zu geben scheint, ist dies noch kein Erklärungsmodell. Dazu müsste man die dahinterliegenden Kosten-Nutzen-Kalküle aufdecken, die an bestimmte Handlungsbedingungen anknüpfen (Schnell, 1997, hier: Seite 215). Demografische Variablen (Geschlecht, Alter, Schicht, Bildungsstand) sind nicht nur als methodologisch unvollständige Erklärungen des Teilnahmeverhaltens zu interpretieren (Schnell, 1997, hier: Seite 208 f.), sondern auch Erklärungsmodelle, die auf soziodemografischen Variablen basieren, „sind nicht in der Lage, handlungstheoretische Beweggründe individuellen Verhaltens angemessen zu erklären“ (Haunberger, 2011, hier: Seite 104).

Über die Soziodemografie hinaus beeinflussen auch andere Faktoren die Struktur des Unit-Nonresponse einer Erhebung. Das Teilnahmeverhalten wird auch von Charakteristika der Umfrage selbst (zum Beispiel Umfang des Frageprogramms, Komplexität der Fragen) und der Form der Feldarbeit (Erhebungsmode, Ansprache durch den Interviewer, Zeitpunkt der Befragung) beeinflusst

1 Zum Beispiel die Studie von Behr und anderen (2002), die Daten des European Community Household Panels (ECHP), des Sozio-ökonomischen Panels (SOEP) und des British Household Survey Panels (BHSP) in Bezug auf Nonresponse untersuchten. Ebenso Schnell, der US-Census-Match-Studien und verschiedene britische Umfragen untersuchte (Schnell, 1997, hier: Seite 204 f.), sowie Stoop und andere (2010, hier: Seite 122 ff.).

(Haunberger, 2011, hier: Seite 49). Ältere Frauen sind beispielsweise weniger bereit, eine (fremde) Person für eine Face-to-Face-Befragung in ihre Wohnung zu lassen (Stoop und andere, 2010, hier: Seite 123). Auch das Thema der Erhebung kann in Zusammenhang mit der Teilnahmebereitschaft stehen. In einer Analyse der deutschen Stichprobe des European Social Survey (ESS) zeigte sich bei der Berechnung verschiedener logistischer Regressionsmodelle, dass politisches Interesse einen starken, signifikanten Einfluss auf die Teilnahmebereitschaft hat und zuvor festgestellte soziodemografische Effekte insignifikant werden ließ (Weinhardt/Liebig, 2015, hier: Seite 76).²

Die Ausdehnung der Bandbreite an Variablen, die zur Erklärung von Unit-Nonresponse herangezogen werden, löst jedoch nicht das grundsätzliche Problem des theoretischen Verständnisses. Es existieren nach wie vor keine theoretischen Modelle, die die Gründe von Nonresponse und die sozialen Mechanismen des Teilnahmeprozesses verlässlich erklären können (Haunberger, 2011, hier: Seite 97). Ein solches Modell müsste das Teilnahmeverhalten von Personen auch im Kontext der Eigenheiten der jeweiligen Erhebung erklären können.

4

Korrekturtechniken

Kommt es aufgrund von Unit-Nonresponse zu Verzerrungen, kann man mit verschiedenen Techniken versuchen, diese zu korrigieren. Dazu muss der Ausfallmechanismus anhand von (Hilfs-)Variablen erklärt werden, die sowohl in der Erhebung abgefragt als auch für die Grundgesamtheit bekannt sein müssen (Stoop und andere, 2010, hier: Seite 212). Üblicherweise werden hierzu Gewichtungungsverfahren eingesetzt. Die Verteilung der Hilfsvariablen der Stichprobe wird der Verteilung dieser Hilfsvariablen in der Grundgesamtheit angepasst, indem jeder Stichprobeneinheit ein Korrekturgewicht zugewiesen wird. Damit ist die Stichprobe bezogen auf diese Hilfsvariablen korrigiert.

² Diese logistischen Regressionen waren nur möglich, weil nicht teilnahmebereite Personen einen Kurzfragebogen beantworteten, der auch politisches Interesse erfragte. Von 4 095 Verweigerern konnten 1 070 Kurzfragebogen ausgewertet werden (Weinhardt/Liebig, 2015, hier: Seite 62 f.).

Problematisch wird der Stichproben-Bias besonders dann, wenn sich der Ausfallmechanismus über Phänomene erklärt, über die eine Erhebung Auskunft geben soll. Dies ist beispielsweise gegeben, wenn in einer Erhebung zum Thema „Einkommenssituation“ die Teilnahmebereitschaft mit dem Einkommen der Personen der Bruttostichprobe korreliert ist (Glaser/Kafka, 2015, hier: Seite 482). In einer solchen Situation stehen häufig keine geeigneten Hilfsvariablen zur Verfügung, da die Erkenntnisse über das Phänomen aus der Erhebung selbst kommen sollen.

Eine bekannte Form der Gewichtungsverfahren ist die Poststratifizierung (auch Redressment genannt). Die Idee dabei ist, die Grundgesamtheit bezüglich einer Anzahl von Hilfsvariablen in homogene Schichten einzuteilen. Ebenso werden alle Befragten nach dem gleichen Verfahren in die gleichen Schichten eingeteilt. Allen Befragten innerhalb einer solchen Schicht wird ein Gewicht zugeschrieben, das die (verzerrte) Schichtstruktur der Stichprobe an die (wahre) Schichtstruktur der Grundgesamtheit anpasst (Bethlehem und andere, 2011, hier: Seite 241; Stoop und andere, 2010, hier: Seite 212; Schnell und andere, 2013, hier: Seite 306). Poststratifizierung kann gut funktionieren, wenn die Schichten homogen sind, was jedoch nicht immer der Fall ist (Bethlehem und andere, 2011, hier: Seite 240). Dennoch können zwei Probleme auftreten: Zum einen kann es Schichten geben, für die keine Beobachtungen in der Stichprobe vorliegen, zum anderen kann es einen Mangel an adäquaten Informationen in den Hilfsvariablen geben. Da in der Praxis häufig eben solche geeigneten Hilfsvariablen fehlen, die mit den interessierenden Variablen ausreichend korrelieren, ist der Wert der Poststratifizierung als generelle Methode, um Unit-Nonresponse-Verzerrungen zu korrigieren, limitiert (Stoop und andere, 2010, hier: Seiten 212 und 301).

Ein anderer Zweig der Gewichtungungsverfahren kann unter dem Begriff „Kalibrierung“ zusammengefasst werden. Um Gewichte zu bestimmen, werden dabei Regressionen verwendet, die es erlauben, die Randverteilungen von Hilfsvariablen der Stichprobe an viele verschiedene Randverteilungen dieser Hilfsvariablen in der Grundgesamtheit anzupassen (Schnell und andere, 2013, hier: Seite 308).

Lineare Gewichtungungsverfahren legen beispielsweise ein lineares Regressionsmodell zugrunde, bei dem die abhängige Variable (Gewicht) aus einem Set von Hilfs-

variablen bestimmt wird. Die Korrekturgewichte der Zielvariablen werden aus der Summe von Gewichtungskoeffizienten bestimmt (Bethlehem und andere, 2011, hier: Seite 241).

Im Unterschied dazu werden bei multiplikativen Gewichten (auch raking ratio estimation oder iterative proportional fitting) die Korrekturgewichte aus dem Produkt der Gewichtungsfaktoren gebildet (Bethlehem und andere, 2011, hier: Seite 231). Sie basieren auf einer Anpassungsprozedur, die wiederholt nacheinander die Gewichtungsfaktoren der einzelnen Kreuzklassifizierungen so lange anpasst, bis die gewichtete Stichprobenverteilung der Hilfsvariablen der Verteilung in der Grundgesamtheit entspricht. Es wird dabei ein loglineares Modell für die Zelhäufigkeiten angenommen.

Welches Verfahren angemessen ist, entscheidet sich danach, welche Modellannahmen zum Ausfallmechanismus angenommen werden (Bethlehem und andere, 2011, hier: Seiten 237 und 241).

Wie auch bei anderen Gewichtungsverfahren, geht man bei allen Kalibrierungsmethoden immer von einem bestimmten Modell aus, innerhalb dessen die Hilfsvariablen, die zur Gewichtung herangezogen werden, den Unit-Nonresponse ausgleichen sollen. „Dies ist mit den zur Verfügung stehenden Variablen in der Praxis häufig schwierig zu erreichen und wird sehr selten nachgewiesen, sondern meist nur angenommen“ (Schnell und andere, 2013, hier: Seite 308).

Eine etwas andere Idee verfolgen Response-propensity-Verfahren. Ausgehend von Hilfsvariablen, die auch für Nicht-Befragte vorliegen (müssen), wird meist mithilfe von multivariaten logistischen Modellen und logistischen Regressionen die Antwortneigung (= response propensity) als abhängige Variable geschätzt. Die Antwortneigung ist also die abhängige Wahrscheinlichkeit, dass eine Stichprobeneinheit an einer Umfrage teilnimmt, bei gegebenen verfügbaren Hintergrundinformationen. Man geht dabei davon aus, dass Stichprobeneinheiten, die die gleiche Antwortneigung haben, die gleiche Verteilung an Hilfsvariablen aufweisen, unabhängig davon, ob sie tatsächlich teilgenommen haben oder nicht. Im weiteren Verfahren kann die Antwortneigung auf unterschiedliche Art als direktes Gewicht in der Berechnung der Schätzwerte (response propensity weighting; response propensity GREG³) oder indirekt,

als Grundlage zur Bildung von Gewichtungsgruppen (response propensity stratification), genutzt werden (Bethlehem und andere, 2011, hier: Seite 327 ff.).

In der Praxis erweisen sich Response-propensity-Techniken besonders dann als schwache Korrekturverfahren mit geringen Verbesserungsmöglichkeiten (Bethlehem und andere, 2011, hier: Seite 348; Schnell und andere, 2013, hier: Seite 308), wenn der modellierte Ausfallmechanismus die empirischen Ausfälle nur unzulänglich erklärt und die interessierenden Variablen nur schwach mit den eingesetzten Variablen korreliert sind (Stoop und andere, 2010, hier: Seite 212; Blumenstiel/Gummer, 2015, hier: Seite 17).

5

Fallbeispiele zur Korrektur des Nonresponse-Bias

Im Folgenden sollen kurz einige Studien gezeigt werden, die sich praxisorientiert mit der Anwendung von Korrekturtechniken zur Korrektur von Nonresponse-Verzerrungen befasst haben.

Gabler/Häder (1997) versuchten, die Schichtungsstrukturen zweier Eurobarometer-Stichproben (zum einen der Standard-Eurobarometer mit persönlich mündlicher Befragung, zum anderen eine ergänzende Telefonumfrage) mithilfe von Gewichtungsverfahren an die der Grundgesamtheit (gegeben durch amtliche Zahlen) anzupassen, um somit die Verteilung der in beiden Stichproben vorhandenen Zielvariable „subjektive Schichteinstufung“ anzunähern. Sie zogen dafür Daten aus Belgien, Frankreich, Spanien und Deutschland heran und wählten zwei unterschiedliche Wege, die Gewichte zu konstruieren: zum einen eine lineare Methode (Kleinste-Quadrate-Regressionsschätzer), zum anderen eine multiplikative Methode (iterative proportional fitting). Die benutzten Hilfsvariablen hatten jedoch kaum Einfluss auf die Zielvariable „subjektive Schichteinstufung“, weshalb die unterschiedlichen Randverteilungen der Zielvariable zwischen den beiden Studien durch die Gewichtungsverfahren nicht angeglichen werden konnten. Dies trifft für alle betrachteten Länder sowie für beide Gewichtungsverfahren zu (Gabler/Häder, 1997, hier: Seite 243).

3 Generalized Regression Estimator.

Blumenstiel/Gummer (2015) beschäftigen sich in einem Papier mit den Möglichkeiten, Unit-Nonresponse-Verzerrungen durch Response-propensity-Methoden zu reduzieren. Sie kommen zu dem Schluss, dass zwar verschiedene Verfahren in der Lage sind, Verzerrungen zu reduzieren, in einzelnen Fällen aber auch das Gegenteil der Fall sein kann: Die Verzerrungen wurden durch den Einsatz der Methoden verstärkt. Es bleibt aus ihrer Sicht fraglich, „ob das relativ geringe Ausmaß der Bias-Reduzierung den [...] Aufwand rechtfertigt“ (Blumenstiel/Gummer, 2015, hier: Seite 39).

Glaser/Kafka (2015) haben verschiedene Korrekturtechniken getestet. Im Wesentlichen verglichen sie verschiedene Formen von Response-propensity-Methoden mit einigen Kalibrierungstechniken. Im Ergebnis „bleibt bei allen untersuchten Schätzern, welche auf unterschiedlichen Gewichtungsverfahren beruhen, Nonresponse Bias in unterschiedlichem Ausmaß bestehen“ (Glaser/Kafka, 2015, hier: Seite 481). Dabei konnten Kalibrierungsverfahren etwas bessere Ergebnisse erzielen als Response-propensity-Methoden.

Die genannten Beispiele zeigen, dass eingesetzte Korrekturverfahren nicht immer zum gewünschten Erfolg führen. Bezüglich der den Korrekturtechniken zugrunde liegenden Modellannahmen muss in diesem Zusammenhang beachtet werden, dass „an attempt to use a correction technique for which the underlying model does not hold will not help to reduce the bias“ (Bethlehem und andere, 2011, hier: Seite 237). Doch gerade die Frage, ob die getroffenen Modellannahmen zutreffend sind, ist äußerst kritisch. „In other words, we can never be sure whether the distributions of the target variables represent the population values better after weighting. We can only assume that if different adjustment methods using different types of auxiliary data provide comparable estimators, weighting will improve the results“ (Stoop und andere, 2010, hier: Seite 213). Dabei kann man nicht feststellen, welcher der Schätzer besser ist, da es keinen fixen Richtwert, keinen Benchmark gibt, „because each of the adjustment schemes is based on a number of untestable assumptions“ (Stoop und andere, 2010, hier: Seite 213). Es besteht daher die Notwendigkeit nach „weiterer Forschung zur Verbesserung der zugrunde liegenden Schätzmodelle und der Variablenauswahl“ (Blumenstiel/Gummer, 2015, hier: Seite 38).

Um also verzerrungsfreie Daten zu bekommen, empfiehlt es sich, den Unit-Nonresponse so klein wie möglich zu halten, da dadurch bereits das Potenzial von Unit-Nonresponse-Verzerrungen minimiert wird.

6

Fazit

Weist eine Erhebung einen Unit-Nonresponse auf, ist damit immer die Möglichkeit verbunden, dass die Schätzungen von Populationsparametern durch diesen Unit-Nonresponse verzerrt werden. Gewichtungsverfahren können die Verzerrungen reduzieren und sollten daher zum Einsatz gebracht werden. Der Grad, in dem es gelingt, mittels Gewichtungsverfahren die durch den Unit-Nonresponse verursachten Verzerrungen zu reduzieren, ist im Allgemeinen unbekannt. Die Korrekturleistung einer Gewichtung ist im Wesentlichen davon abhängig, in welchem Maße das im Gewichtungsmodell enthaltene Erklärungsmodell in der Lage ist, die empirischen Ausfallmechanismen abzubilden.

Hierbei erweist es sich als hinderlich, dass immer noch keine allgemein anerkannten Erklärungsmodelle existieren, welche in der Lage wären, die Ausfallmechanismen zu erklären, die zum Unit-Nonresponse einer Erhebung führen. Damit entfällt die Möglichkeit, den Ausfallmechanismus über eine standardisierte Vorgehensweise auf qualitativ befriedigende Weise zu kompensieren.

Die Befunde der Methodenforschung müssen vielmehr dahingehend verstanden werden, dass der Ausfallmechanismus eine hohe Abhängigkeit von der jeweiligen Erhebung aufweist. Das Thema der Befragung, die verwendeten Modes sowie die Organisation der Feldarbeit haben einen entscheidenden Einfluss auf den Ausfallmechanismus. Es ist daher nicht davon auszugehen, dass es gelingt, ein allgemein gültiges Erklärungsmodell zu entwickeln, aus dem sich Standards ableiten lassen, um den Unit-Nonresponse-Bias über Gewichtungsverfahren zielsicher abzufangen.

Daneben muss beachtet werden, dass sich Erklärungsmodelle zur Korrektur des Unit-Nonresponse-Bias nur dann zur Anwendung bringen lassen, wenn für die erklärenden Merkmale Eckwerte aus anderen Quellen

zur Verfügung stehen und im Merkmalskranz der Erhebung enthalten sind. Gewichtungungsverfahren passen die Verteilung ausgewählter Merkmale der Erhebung, die im Idealfall den Ausfallprozess erklären, an die „wahren“ Werte der Grundgesamtheit an. Damit ist der Bias für die Korrekturmerkmale kompensiert. Diese Kompensation wirkt sich auch für andere Merkmale positiv auf die Bias-Reduktion aus. Je höher die Korrelation mit den Korrekturmerkmalen ausfällt, desto stärker ist die Bias-Reduktion gegeben.

Letztlich bleibt der Unit-Nonresponse-Bias, der nach der Gewichtung die Schätzung von Populationsparametern verzerrt, aber als systematischer Fehler unbekannt. Für die Schätzung von Populationsparametern kann ein zufälliger Fehler bestimmt werden. Dieser zufällige Fehler kann eine gute Qualität von Schätzungen auch dann ausweisen, wenn diese vom Unit-Nonresponse-Bias verzerrt sind.

Da insgesamt der Erfolg nachträglicher Korrekturverfahren im Wesentlichen unbekannt ist, sollte intensiv darauf hingearbeitet werden, den Unit-Nonresponse – und somit das Potenzial für den Unit-Nonresponse-Bias – von vornherein möglichst klein zu halten. Dies gilt insbesondere dann, wenn eine sehr hohe Genauigkeit äußerst wichtig ist. [\[1\]](#)

LITERATURVERZEICHNIS

Behr, Andreas/Bellgardt, Egon/Rendtel, Ulrich. Extent and determinants of panel attrition in the European Community Household Panel. In: CHINTEX Working Paper #7, Work-package 6. Frankfurt am Main 2002.

Bethlehem, Jelke/Cobben, Fannie/Shouten, Barry. Handbook of Nonresponse in Household Surveys. Hoboken 2011.

Blumenstiel, Jan Eric/Gummer, Tobias. Prävention, Korrektur oder beides? Drei Wege zur Reduzierung von Nonresponse Bias mit Propensity Scores. In: Schupp, Jürgen/Wolf, Christof (Herausgeber). Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen. Wiesbaden, Opladen 2015, Seite 13 ff.

Faulbaum, Frank. Total Survey Error. In: Baur, Nina/Blasius, Jörg (Herausgeber). Handbuch Methoden der empirischen Sozialforschung. Wiesbaden 2014, Seite 439 ff.

Gabler, Siegfried/Häder, Sabine. Wirkung von Gewichtungen bei Face-to-Face- und Telefonstichproben. Eurobarometer-Experiment 1994. In: Gabler, Siegfried/Hoffmeyer-Zlotnik, Jürgen H. P. (Herausgeber). Stichproben in der Umfragepraxis. Opladen 1997, Seite 221 ff.

Glaser, Thomas/Kafka, Elisabeth. Analyse und Behebung von selektivem Bias – EU-SILC Österreich. In: Schupp, Jürgen/Wolf, Christof (Herausgeber). Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen. Wiesbaden, Opladen 2015.

Groves, Robert M./Lyberg, Lars. Total Survey Error. Past, Present, and Future. In: Public Opinion Quarterly. Jahrgang 74, Ausgabe 5, 2010, Seite 849 ff.

Hauemberger, Sigrid. Teilnahmeverweigerung in Panelstudien. Wiesbaden 2011.

Proner, Hanna. Ist keine Antwort auch eine Antwort? Die Teilnahme an politischen Umfragen. Wiesbaden 2011.

Rendtel, Ulrich. Lebenslagen im Wandel. Panelausfälle und Panelrepräsentativität. Frankfurt am Main 1995.

Schnell, Rainer. Nonresponse in Bevölkerungsumfragen: Ausmaß, Entwicklung und Ursachen. Opladen 1997.

Schnell, Rainer/Hill, Paul B./Esser, Elke. Methoden der empirischen Sozialforschung. München 2013.

Stoop, Ineke/Billiet, Jaak/Koch, Achim/Fitzgerald, Rory. Improving Survey Response: Lessons learned from the European Social Survey. Chichester 2010.

Weinhardt, Michael/Liebig, Stefan. Teilnahmeverhalten und Stichprobenverzerrung in der deutschen Stichprobe des European Social Survey. In: Schupp, Jürgen/Wolf, Christof (Herausgeber). Nonresponse Bias. Qualitätssicherung sozialwissenschaftlicher Umfragen. Wiesbaden 2015, Seite 47 ff.

Herausgeber

Statistisches Bundesamt, Wiesbaden

www.destatis.de

Schriftleitung

Dieter Sarreither, Präsident des Statistischen Bundesamtes

Redaktionsleitung: Kerstin Hänsel

Redaktion: Ellen Römer

Ihr Kontakt zu uns

www.destatis.de/kontakt

Erscheinungsfolge

zweimonatlich, erschienen im August 2016

Das Archiv aller Ausgaben ab Januar 2001 finden Sie unter www.destatis.de/publikationen

Print

Einzelpreis: EUR 18,- (zzgl. Versand)

Jahresbezugspreis: EUR 108,- (zzgl. Versand)

Bestellnummer: 1010200-16004-1

ISSN 0043-6143

ISBN 978-3-8246-1046-4

Download (PDF)

Artikelnummer: 1010200-16004-4, ISSN 1619-2907

Vertriebspartner

IBRo Versandservice GmbH

Bereich Statistisches Bundesamt

Kastanienweg 1

D-18184 Roggentin

Telefon: +49 (0) 382 04 / 6 65 43

Telefax: +49 (0) 382 04 / 6 69 19

destatis@ibro.de

Papier: Metapaper Smooth, FSC-zertifiziert, klimaneutral, zu 61% aus regenerativen Energien

© Statistisches Bundesamt, Wiesbaden 2016

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.