

Alles Zufall oder? Ein Diskussionsbeitrag zur Qualität von face to face-Umfragen in Deutschland

Mohler, Peter Ph.; Koch, Achim; Gabler, Siegfried

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Mohler, P. P., Koch, A., & Gabler, S. (2003). Alles Zufall oder? Ein Diskussionsbeitrag zur Qualität von face to face-Umfragen in Deutschland. *ZUMA Nachrichten*, 27(53), 10-15. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-207710>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

ALLES ZUFALL ODER?

Ein Diskussionsbeitrag zur Qualität von face to face-Umfragen in Deutschland

PETER PH. MOHLER, ACHIM KOCH & SIEGFRIED GABLER

Der European Social Survey (ESS) setzt 70% Ausschöpfung als in Europa erreichbares Ziel an: "...the minimum target response rate – after discounting ineligibles ... – should be 70%. This figure is likely to be exceeded in certain countries and the ESS as a whole would be damaged if major national variations in response rates were to occur." (ESS 2001: 5, vgl. auch ESS 2003: 10). Dieses Ziel wird in Deutschland schon lange kaum noch erreicht. Wenn es aber anderswo in Europa als durchaus überraffbar gilt, ist es wohl an der Zeit, die öffentliche Diskussion über Qualität der Umfrageforschung in Deutschland zu verstärken. ZUMA beteiligt sich seit langem an dieser Diskussion, etwa durch das Symposium "Ausschöpfungen bei Umfragen" im Jahr 1997 (vgl. auch Porst 1996), Beteiligung an der DFG Publikation zur Qualität der Umfrageforschung (Kaase 1999) oder Beiträge in den ZUMA Nachrichten, wie zuletzt Koch (1998), Schräpler (2000), Diekmann/Jann (2001), Koch (2002) oder Schneekloth/Leven (2003) in diesem Heft. In diesem Zusammenhang sei auch auf die umfangreichen Publikationen von Schnell (insbes. Schnell 1997) verwiesen.

Dringlich wird die Diskussion deshalb, weil wir den Eindruck gewonnen haben, dass heutzutage in Deutschland auch qualitätsorientierte Umfragen faktisch kaum eine Ausschöpfungsquote von mehr als 50% erzielen – zumindest dann nicht, wenn das Stichprobenbrutto der Umfragen eindeutig definiert und die Umsetzung dieser Vorgaben im Feld penibel kontrolliert wird. Die oft genannten Gründe für das Absinken der Ausschöpfungsquote – wie schwierigere Erreichbarkeit von Zielpersonen (z.B. aufgrund steigender Frauenerwerbsquoten, der Zunahme von Einpersonenhaushalten u.ä.) oder ein verschlechtertes ‚Umfrageklima‘ (de Heer 1999) – können die schlechten Ausschöpfungsergebnisse nur bedingt erklären. Unsere Erfahrungen mit dem ALLBUS und anderen Studien deuten hier auf die *Form des Produktionsprozesses* als einem zentralen, bislang vernachlässigten Faktor hin. Welcher Produktionsprozess für eine Umfrage gewählt wird, hat offensichtlich nicht nur umfragetechnische, sondern auch betriebswirtschaftliche bzw.

organisatorisch/technische Gründe. Zufällig sind diese nicht, eher kann man sie als ‚strukturelle Bedingungen‘ ansehen. In unserer folgenden Skizze werden wir diesen Gedankengang näher ausführen.

Umfragen, genauer gesagt, repräsentative Bevölkerungsumfragen im ‚Angesicht der zu befragenden Person‘ (face-to-face) entstehen durch einen hochkomplexen Produktionsprozess (Mohler 2000). In diesem Produktionsprozess hat der Zufall nur eine Chance: bei der Auswahl von zu befragenden Personen oder Haushalten, zumindest, wenn die Stichprobe repräsentativ und/oder über Zeit und Raum vergleichbar sein soll (Gabler/Häder 2003, Lohr 1999). Aber auch im Prozess der Auswahl gibt es aufgrund fehlender Listen in der Regel keine uneingeschränkte Zufallsauswahl. Komplexe Auswahlverfahren (geschichtet, mehrstufig) wurden entwickelt, um den Auswahlprozess innerhalb eines gegebenen Auswahlrahmens zu optimieren. Alle anderen Prozesse, beginnend beim Design der Untersuchung, der Fragebogenentwicklung, der Interviewersteuerung, der Dokumentation, der Datenaufnahme bis zur Veröffentlichung der Ergebnisse sind deterministisch festgelegt. Sie folgen wissenschaftlich fundierten Regeln, damit aus der Umfrage ein ‚wissenschaftliches Instrument‘ werden kann, dessen Eigenschaften wohl bekannt sind. Jeder einzelne dieser Prozesse folgt speziellen methodischen Prinzipien, besser gesagt, er sollte solchen Prinzipien folgen. Um Prinzipien und Regeln folgen zu können, müssen diese nicht nur aufgeschrieben, sondern auch wissenschaftlich als tragfähig belegt sein. Dies gilt weitgehend bislang nur für die Teilprozesse Stichprobe (vgl. z.B. Gabler/Hoffmeyer-Zlotnik 1997) und Fragebogenentwicklung (vgl. z.B. Sudman et al. 1996). Für so wichtige Prozesse, wie das Design, die Feldsteuerung oder die Dokumentation, gibt es noch keine durch methodische Diskurse fundierte Regeln. Stattdessen bestimmen Tradition und Common Sense, was wie gemacht wird. Das Resultat sind Schwankungen in der Ausschöpfung von 30 Prozentpunkten und mehr in ein und derselben Gemeinde beim gleichen Institut in verschiedenen Jahren (wie beim ALLBUS z.T. zu beobachten), die sich nicht mehr auf statistischen Zufall zurückführen lassen, sondern auf Mängel im Produktionsablauf hindeuten. Dies führt aber auch zu Unterschieden zwischen Instituten in einem Land (die sattsam bekannte ‚Institutshandschrift‘), zwischen Regionen und Ländern, wie sich zum Beispiel aus den Dokumentationen des ISSP und ESS ablesen lässt (www.issp.org, www.europeansocialsurvey.org).

Nun könnte man fragen, hat die Empirische Sozialforschung damit nicht lange gut gelebt? Was, außer möglicherweise übertriebenem Perfektionismus, sollte einen bewegen, sich den zugegebenermaßen schwierigen Fragen einer fundierten Prozesssteuerung zuzuwenden?

Nun, wer über den eigenen Tellerrand hinausblickt, wird feststellen können, dass in anderen europäischen Ländern wie Dänemark oder Schweden auch streng kontrollierte Erhebungen Ausschöpfungsquoten jenseits der 60%-Marke erreichen. Dabei soll nicht verschwiegen werden, dass es in Europa auch Länder wie die Schweiz gibt, die nicht einmal 50% Ausschöpfung erreichen (http://ess.nsd.uib.no/2003_Fworksummary.jsp).

Wenn also weder die unwilligen Bürger (Survey Climate) noch die verbesserten Kontrollen (wie z.B. bei den im ALLBUS eingesetzten Registerstichproben) hinreichende Gründe für die schlechten Ausschöpfungsergebnisse in Deutschland sind, ist es dann womöglich der zu niedrige Preis, die fehlende Bereitschaft für gute Daten auch gutes Geld zu bezahlen? Diese Frage hat zwei Antworten:

- die eine ist, dass es in Deutschland derzeit weder für ‚Geld noch gute Worte‘ eine hohe, nachvollziehbare Ausschöpfung bei face-zu-face Umfragen gibt, weil die dafür notwendigen Produktionslinien nicht vorhanden sind (s.u.),
- die andere ist, dass es in Deutschland viel zu wenig Erhebungen gibt, die auf eine hohe Qualität oder besser Präzision ihres Erhebungsinstrumentes Wert legen.

Damit fehlt die Nachfrage, welche es den Anbietern erlauben würde, auch die dafür notwendigen Produktionsprozesse aufzulegen. Diesen Teufelskreis gilt es zu durchbrechen, wenn deutsche Umfragen internationalen wissenschaftlichen Standards weiterhin genügen wollen.

Belege für die Bedeutung des Produktionsprozesses zeigt eine noch nicht veröffentlichte Studie in den Niederlanden (vgl. Stoop 2003). Dort wurden Prozesse bewusst verändert mit dem Ziel einer Verbesserung der Ausschöpfung bzw. der Informationen über Nichtteilnehmer. Stoop berichtet für diese landesweite Umfrage eine kontrollierte Ausschöpfung von annähernd 70% im Hauptfeld und weiteren 70% bei einer kleinen Zufallsauswahl von Nichtteilnehmern in der Nachbearbeitung. Die von de Heer (1999) bis dahin berichteten Werte für die Niederlande lagen bei 55% Ausschöpfung. Dies und andere Indikatoren legen die Bedeutung des Erhebungsprozesses für die Ausschöpfungsquote nahe. Je nach Ausgestaltung dieses Prozesses können danach gute Ausschöpfungsquoten sogar unter widrigen umfrage-klimatischen Bedingungen erzielt werden.

Trivial? Die Antwort ist ja und nein zugleich. Einerseits ist es offensichtlich, dass die Qualität eines Produktes wesentlich vom Produktionsprozess abhängt. Weniger offensichtlich ist, dass auch Umfragen Ergebnis eines mit anderen Wirtschaftsgütern vergleichbaren Produktionsprozesses sind (Mohler 2000). Wie die Leadership Group on Quality in the European Statistical System formuliert: "The product is generated by an underlying process. It is unlikely that the product will have good quality if the underlying process is not up to par. Therefore the concept of process quality comes into play." (Lyberg et al.

2001). Deshalb gilt, dass mit einem auf Massenware ausgerichteten Fließband qualitativ hochwertige Waren nicht erzeugt werden können. Die Begrifflichkeit von „Massenware“ und „hochwertige Produkte“ ließe sich natürlich auch auf Umfragen übertragen. Jedoch möchten wir hier anstelle des abwertenden Begriffes „Massenware“ eher von „unterschiedlicher Präzision“ von Erhebungen sprechen. Diese Begrifflichkeit verweist auf andere technische Instrumente bzw. Messgeräte, wo die Präzision sich aus der angezielten Verwendung ergibt. Die dafür notwendigen Produktionsprozesse unterscheiden sich ganz erheblich.

In diesem Sinne wurden bis in die neunziger Jahre auch Umfragen, je nach der angestrebten Präzision getrennt, produziert. Es gab Interviewerstäbe für Marketingerhebungen, Sozialforschung, Gesundheitsforschung usw. Danach fand eine ‚technische Revolution‘, wahrscheinlich angeregt durch die hohen Investitionskosten für computerunterstützte face-to-face Erhebungen (CAPI) statt, indem an die Stelle von speziellen Interviewerstäben (und auch Feldsteuerungen) gemeinsame Stäbe und eine einheitliche Feldsteuerung traten. Von da an wurde ein und derselbe Interviewer für alle möglichen thematische Bereiche, vor allem aber für höchst unterschiedliche Präzision bei der Erfüllung des Stichprobenplanes eingesetzt (von ‚vorgegebener Anschrift einer Person‘ bis zu ‚random irgendwie walk‘). Krasse ‚random-irgendwie-walks‘ laufen darauf hinaus, dass es für die Interviewer einen mehr oder weniger zufallsgesteuerten Pfad durch ein Stadtviertel gibt, der so lange verfolgt wird, bis genügend Interviews erzielt werden. Da durch die Begehungsunterlagen und Begehungsregeln bestimmte Gebiete (z.B. Neubaugebiete, Adressen mit hohen Hausnummern in längeren Straßen, kleine Orte) praktisch keine Auswahlchance haben, verabschiedet man sich faktisch von einer Zufallsstichprobe mit berechenbaren Inklusionswahrscheinlichkeiten. Die Zahl der Nichtkontakte, Verweigerungen etc. bleibt bei diesem Vorgehen unerheblich. Das Verfahren legt den Interviewern nahe, möglichst schnell viele Interviews zu machen und dabei nicht an der Zahl kontaktierter Haushalte zu sparen. Geht man mit einer solchen Strategie an ein Design, in dem die Zielpersonen explizit mit Anschrift vorgegeben werden, wird die Stichprobe nicht nachdrücklich genug bearbeitet. Das Resultat ist eine hohe Zahl von ‚verbrannten‘ Adressen, d.h. eine suboptimale Ausschöpfung (von den weiteren Problemen bei der Durchführung der Interviews soll hier erst gar nicht die Rede sein). Dies bedeutet, dass die ‚technische Revolution‘ der Zusammenlegung von Interviewerstäben zwangsnotwendig zu einer Verringerung der maximal erreichbaren Präzision des wissenschaftlichen Instruments ‚Repräsentative Bevölkerungsumfrage‘ führen musste.

Dies ist aber nur einer von mehreren Strukturbrüchen, welche die Umfrageforschung in den letzten Jahren erlebt hat. Ein weiterer für den Produktionsprozess wichtiger Strukturbruch ist die Gesetzgebung zur Scheinselbständigkeit. Diese erzwingt, dass Interviewer

gleichzeitig für mehrere Erhebungsinstitute tätig sein müssen. Damit hat kein Feldchef mehr alleinigen Zugriff auf bestimmte Interviewer, er kennt deren persönliche Auslastung nicht mehr, die Bindung an das Institut ist, vorsichtig ausgedrückt, schwach. Hinzu kommt eine für hohe Präzisionsleistung eigentlich nicht angemessene Bezahlung. Das führt zu erheblichen Rekrutierungsproblemen bzw. in bestimmten Regionen zu einem Interviewermangel, was unter den gegebenen Bedingungen den Interviewern als selbständigen Unternehmern erlaubt, die weniger gewinnbringenden Studien links liegen zu lassen. Produktionsintensive, sehr präzise Erhebungen werden damit offensichtlich erschwert.

Um einen Weg aus dieser, wie wir meinen, strukturellen Krise der Umfrageforschung zu finden, wäre zuerst einmal eine fundierte öffentliche Diskussion über die für unterschiedliche Präzisionen notwendigen Produktionsprozesse erforderlich. Diese mag vielleicht ergeben, dass die von uns genannten Faktoren nicht die entscheidenden sein mögen. Dafür sind wir sicher offen. Weniger offen wären wir gegenüber entschuldigenden Argumenten, die sinkende Ausschöpfung ergäbe sich aus einer zufälligen Häufung unglücklicher Umstände. In diesem Sinne würden wir die Frage „alles Zufall oder?“ mit einem klaren „Nein“ beantworten.

Literatur

de Heer, W., 1999: International Response Trends: Results of an International Survey. *Journal of Official Statistics*, 15 (2): 129-142.

Diekmann, A./Jann, B., 2001: Anreizformen und Ausschöpfungsquoten bei postalischen Befragungen. Eine Prüfung der Reziprozitätshypothese. *ZUMA-Nachrichten* 48: 18-27.

ESS, 2001, 2003: European Social Survey – Specification for participating countries. http://naticent02.uuhost.uk.uu.net/proj_spec/index.htm

Gabler, S./Häder, S., 2003: Auswahl der Untersuchungseinheiten
http://www.politikon.org/preview.php?sid=a27b9aa5fcd1b95d17babe0abd507705&ilias_url=http%3A%2F%2Fwww.politikon.org/preview/course.php%3Fco_id%3D150%26co_inst%3D1

Gabler, S./Hoffmeyer-Zlotnik, J.H.P. (Hrsg.), 1997: *Stichproben in der Umfragepraxis*. Opladen: Westdeutscher Verlag.

Kaase, M. (Hrsg.), 1999: Qualitätskriterien der Umfrageforschung. Denkschrift der Deutschen Forschungsgemeinschaft. Berlin: Akademie Verlag.

Koch, A., 1998: Wenn "mehr" nicht gleichbedeutend mit "besser" ist: Ausschöpfungsquoten und Stichprobenverzerrungen in Allgemeinen Bevölkerungsumfragen. ZUMA-Nachrichten 42: 66-90.

Koch, A., 2002: 20 Jahre Feldarbeit im ALLBUS. Ein Blick in die Black-Box. ZUMA-Nachrichten 51: 9-38.

Lohr, S., 1999: Sampling: Design and Analysis. Pacific Grove: Duxbury Press.

Lyberg, L. et al., 2001: Summary Report from the Leadership Group (LEG) on Quality. <http://amrads.jrc.it/WPs%20pages/Quality/Documents/LEGsummary.pdf>

Mohler, P. Ph., 2000: Umfrageforschung. S. 310-322 in: B. Held/S. Russ-Mohl (Hrsg.): Qualität durch Kommunikation sichern - Vom Qualitätsmanagement zur Qualitätskultur. Erfahrungsberichte aus Industrie, Dienstleistung und Medienwirtschaft. Frankfurt/M: F.A.Z.-Institut für Management-, Markt- und Medieninformationen.

Porst, R., 1996: Ausschöpfungen bei sozialwissenschaftlichen Umfragen. Die Sicht der Institute. ZUMA-Arbeitsbericht 96/07.

Schneekloth, U./Leven, I., 2003: Woran bemisst sich eine "gute" allgemeine Bevölkerungsumfrage? Analysen zu Ausmaß, Bedeutung und zu den Hintergründen von Nonresponse in zufallsbasierten Stichprobenerhebungen am Beispiel des ALLBUS. ZUMA-Nachrichten 53: 16-57.

Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen: Leske + Budrich.

Schräpler, J.-P., 2000: Was kann man am Beispiel des SOEP bezüglich Nonresponse lernen? ZUMA-Nachrichten 46: 117-149.

Stoop, I., 2003: Surveying Nonrespondents. In: Field Methods (in press).

Sudman, S./Bradburn, N.M./Schwarz, N., 1996: Thinking about Answers: The Application of Cognitive Processes to Survey Methodology. San Francisco: Jossey-Bass.