Probability and Nonprobability Sampling: Representative Surveys of hard-to-reach and hard-to-ask populations. Current surveys between the poles of theory and practice

Survey Methods: Insights from the Field, Special issue: Probability and Nonprobability Sampling: Sampling of hard-to-reach survey populations

Johann Bacher | Johannes Lemcke | Andreas Quatember | Patrick Schmich

How to cite this article : Bacher J., Lemcke J., Quatember A., & Schmich P. (2019). Probability and Nonprobability Sampling: Representative Surveys of hard-to-reach and hard-to-ask populations. Current surveys between the poles of theory and practice. Survey Methods: Insights from the Field. Retrieved from https://surveyinsights.org/?p=12070

DOI : 10.13094/SMIF-2019-00018

Copyright : © the authors 2019. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Introduction

In times of increasing demand for evidence based data and accelerating social change, survey research faces numerous challenges. Survey research is still confronted with decreasing and stagnating response and cooperation rates. Furthermore, the emergence and diffusion of relatively new technologies in combination with different usage styles makes it necessary to adapt the framework in which surveys are administrated. Due to the rapid diffusion of the internet as communication platform, an increasing share of surveys is conducted online. This development intensifies the ongoing discourse about probability vs. non-probability sampling. While within survey research there is still a dispute between the representativeness of probability and non-probability sampling, research has long since been overtaken by reality. For example, serious and influential newspapers are now equally using non-probability samples for their surveys (which does not necessarily mean that this approach is valid). In this context research is needed to evaluate, under which conditions non-probability sample surveys can obtain valid and reliable information on characteristics such as attitudes and behaviors of a target population. A potentially valuable of non-probability samples could be the case of hard-to-reach populations. In this case, different non-probability sampling techniques, such as snowball sampling, respondent-driven sampling or quota sampling in non-probability web panels, could increase the integration of these hard-to-reach populations.

Following Willis et al. (2014) typical hard-to-reach survey groups are: ethnic minorities, migrant populations, highly mobile populations, homeless and refugee populations, sexual minorities, populations affected by natural disasters or armed conflicts and stigmatized populations. Bonevski et al. (2014) found in a systematic literature review with 116 epidemiological studies, that the three most frequently investigated hard-to-reach groups are: ethnical/racial groups, African American and substance users. Since those people are part of particularly vulnerable groups, it is of tremendous importance to collect valid data of their opinions, living conditions and needs. Therefore the question: “How do we reach the

‘hard-to-reach’?” is crucial for survey methodologists. Various aspects in the survey process are affected by this question.

In this sense Tourangeau (2014) summarized five theoretical categories: Hard-to-reach survey populations are hard to reach because they are **hard to sample, hard to identify, hard to find or contact, hard to persuade** and **hard to interview.** Each of these reasons contribute to the Total Survey Error (Biemer, 2010). In this respect, populations which are considered hard to sample, hard to identify or hard to find/contact can increase the coverage and selection error and thus undermine the representativeness of a survey. Furthermore hard-to-reach groups, which are considered hard to persuade can increase the non-response error and thus undermine the representativeness of a survey as well. Added to this, hard-to-interview survey groups can undermine the validity and reliability (if no appropriate adaptions are made in the measurement process) and thus increase the measurement error of a survey.

Special Issue

With the following collection of articles, ‘Survey Methods: Insights from the Field’ aims to give an overview about the current state of the hard-to-reach research and the ongoing dispute between the two above mentioned sampling methods. Moreover, this Special Issue attempts to combine theoretical discussions, methodological considerations with experiences from the fields. It offers insights into possible links between non-probability sampling and hard-to-reach populations on the one hand, and, on the other hand, different approaches to address the aforementioned problems via the praxis of each methodology.

This special issue was inspired by a PUMA-Symposium 2017, which was organized by two of the guest editors of the Johannes Kepler University Linz (JKU), Johann Bacher and Andreas Quatember, within the PUMA-project of the Austrian social sciences. In this context different sampling issues and different solution attempts were discussed (https://puma.univie.ac.at/home/). The editorship by two researchers of the Robert Koch Institute (RKI), Johannes Lemcke und Patrick Schmich, was guided by the Institute´s interest to a better integration of the hard-to-reach groups into its health monitoring system. The RKI as the public health institute in Germany has the obligation to monitor the health status of the *whole population* in order to offer policy makers valid information within the decision process. As a result of this commitment, various feasibility studies were carried out, which are presented in this special issue alongside initial results. In this respect, the inclusion of elderly and people with migration background is a crucial task for the RKI.

The contributions can be divided into four groups

Of course, the assignment of a paper to a certain group is not distinct in every case. Nonetheless, the grouping below should provide some guidance.

1. Two papers discuss the fundamental theoretical aspects of our topic.
2. Two papers provide a comparison of probability and non-probability sampling from an applied perspective.
3. Reports about experiences of concrete studies build the majority of the paper. Nine papers belong to this group. However, these are not only reports in a strict sense; they provide literature reviews, describe the designs and the underlying assumptions and reflect on the design. They cover migrants and refugees as hard-to-reach-groups mentioned by Willis et al. (2014) as well the elderly as one group sometimes ignored.

4. The fourth group contains papers that can be labeled as “reflections and methodological proposals”. One paper suggests a model to increase the recruitment of old people; the second one applies simulations methods.

Short description of the papers:

1. **Quatember (University Linz)** discusses the inferential quality of an available data set under the standard of the representativeness of a sample and focuses on the assumptions that are made, when calculating an estimator of a certain population characteristic using a specific sampling method. Thus the author offers a theoretical framing for the special issue. In the paper of **Kohler (University Potsdam)**, the usability of non-probability and probability sampling is described under six different research scenarios, which are taken from the practice of empirical research in the social sciences. He considers the conditions, which allows the application of nonprobability sampling methods.
2. In the second group of papers, the article from **Geary et al. (London School of Hygiene and Tropical Medicine)** gives an impression about the shortcomings of probability sampling, when small populations are of interest (in this case men who have sex with men and people of Black African ethnicity in Britain). In this study the research tried to boost the sample size of these groups by utilizing non-probability sampling of a non-probability web-panel. The contribution from **Marks (RTI International (formerly)) & Rhodes (RTI International)** used a similar approach in a different hard-to-reach population (Africans, African Americans, Cambodians, Hispanics, Koreans, and Whites in Los Angeles in Study 1 and African American and White households with individual(s) who have been incarcerated in Study 2). In their paper the authors present the sampling strategy of these two studies.
3. In the third group, the paper of **Koschollek et al. (Robert Koch Institute, Berlin)** describes a project, in which a sample of migrants from sub-Saharan Africa is surveyed in Germany on highly stigmatized topic of HIV. The chosen approach involved members of the hard-to-reach target group throughout the entire process from the planning of the study to data collection. **Zeisler et al. (Robert Koch Institute, Berlin** report about a multilingual feasibility study conducted in two German federal states. The target populations were people with migration background. Different modes of administration and interventions (study hotline, home visits) were used sequentially and evaluated whether they are able to increase participation. In the article from **Prandner & Weichbold (University Linz & University Salzburg)**, the authors present lessons learned in the Austrian Immigrant Survey. In this study the research tackled the sampling issue of immigrants in Austria by building a sampling frame via an onomastic approach. **Hipp et al. (Berlin Social Science Center & University Potsdam)** evaluated another sampling technique. The authors highlight recent findings from an ongoing research project, utilizing respondent-driven sampling. All the contributions in the following section deal with the problem of insufficient or missing sampling frames. In their contribution **Steinhauer et al. (Leibniz Institute for Educational Trajectories, Bamberg)** also describe a sampling procedure for migrants. They provide insights of how to sample refugee Kindergarten children and students in secondary education. The contribution of **Kühne et al. (University Bielefeld)** reports on a project to survey a nationwide sample of refugee households in Germany. As in other papers that focus on immigrants or refugees, problems such as the coverage gap of the Central Register of Foreigners or the high mobility of the target group are discussed. Moreover, the paper presents alternative survey instruments applied to this hard-to-reach subpopulation, a group that is hard to interview. The last three papers in the group discuss problems of hard-to-reach-population for older people. **Gaertner et al. (Robert Koch Institute, Berlin)**, within a sequential mixed-mode design, compare different approaches to include elderly in nursing homes. Amongst the applied methods, face-to-face contact appears to be the only feasible contact mode if postal contact fails. In a second paper, **Gaertner et al. (Robert Koch Institute, Berlin)** analyses the effects of their mixed-mode design in more detail and identify different reasons for non-participation. The last

article in this group from **Kutschar & Weichbold (Paracelsus Medical University & University Salzburg)** addresses the aspect of interviewing elderly in nursing homes (and thus the category of hard-to-interview in the hard-to-reach framework). In their contribution, the authors evaluate the data quality in a survey of elderly people and examine, which respondent-, survey-, and item characteristics predict item non-response, as one indicator of data quality.

4. Two papers were assigned to the last group of papers. Based on their experiences of a quantitative and qualitative study of older people, **Kammerer et al. (Institute for Gerontological Research, Berlin)** develop a model for recruitment to increase the participation of hard-to-persuade and hard-to-interview people. They label their model TIBAR, refereeing to the four important elements of the model, namely building trust, offering incentives, identifying barriers and being responsive. In the last paper of the Special Issue, **Schanze & Zins (Leibniz Institute for the Social Sciences, Mannheim)** used the framework of a simulation study to evaluate the risk of biased estimates, if elderly institutionalized populations are excluded from the sampling frame or are just insufficiently integrated. Thus this study gives an idea of what to expect if the used sampling frame suffers from undercoverage and describes a method to analyze the effects of sampling procedure by simulation studies.

In summary, the papers enable the following conclusions:

- Random sampling is still the gold standard of survey methods. However, there are situations and scenarios where it is (i) not possible to draw a random sample, where (ii) other sampling techniques are appropriate for the analyzed research question and where (iii) other sampling techniques provide acceptable results – in some cases better results – in practice. Examples for all three circumstances are given in this volume.
- Random sampling is based on a set of assumptions. It is important to take these assumptions not for granted. They must be tested in the concrete research project. This requires to collect appropriate data for these task (see below).
- Sampling and non-random sampling are not mutually exclusive. In many cases, multi-stage or mixed-mode designs prove to be most successful; random and not-random sampling methods can be combined in praxis.
- It is important to document the research process and to collect additional information (e.g. metadata) that enables one to evaluate the quality of the realized sample. It should be kept in mind, that the hard-to-reach-problem is multi-dimensional. It cannot be reduced to sampling. Hence, information to all aspects that influence quality should be gathered.
- Finally, “good” surveys require resources. Surveys of hard-to-reach-populations require more resources in order to reach the same quality as in surveys of the general population.

Some remarks on the term “hard-to-reach”

At the end of the introduction the editors want to emphasize that the term “hard-to-reach” is currently considered controversial by some researchers. To cite Nicola Brackertz: ”The problem with using the term ”hard-to-reach” is that it implies a homogeneity within distinct groups, which does not necessarily exist” (Brackertz, 2007, p. 1). Furthermore this term could imply that the problem of reachability is one within the group of hard-to-reach populations and not within the researcher’s approaches to survey hard-to-reach groups (Smith, 2006). There are certainly different social contexts and personality traits which make it hard to survey certain populations. This, however, does not mean that the researches have a passive role. The opposite is true. It becomes just more expensive and resource demanding to integrate

these populations. Willis et acknowledges al. (2014) this fact with the remark: “We must recognize that, from the respondent’s point of view, he or she may not be at all ‘hard-to-reach.’” (Willis et al., 2014, p. 175).

For these reasons, some researches offered different terms to describe the phenomenon of underrepresented survey groups. For example Atkinson & Flint (2001) use “hidden populations” as alternative. This term highlights the sampling issues in such populations. Nevertheless, we still use the term hard to reach in the special issue. We understand the term “hard to reach” multidimensionally. This means that hard to reach groups are difficult to reach with conventional survey methods.

Linz and Berlin in April 2019

Johann Bacher

Johannes Lemcke

Andreas Quatember

Patrick Schmich

Editors of the special issue

References

1. Atkinson, R., & Flint, J. (2001). Accessing hidden and hard-to-reach populations: snowball research strategies. Social Research Update Issue 33. *Social Research Update*, *33*(1), 1–4.
2. Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, *74*(5), 817–848. https://doi.org/10.1093/poq/nfq058
3. Bonevski, B., Randell, M., Paul, C., Chapman, K., Twyman, L., Bryant, J., … Hughes, C. (2014). Reaching the hard-to-reach: A systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC Medical Research Methodology*, *14*(1), 1–29. https://doi.org/10.1186/1471-2288-14-42
4. Brackertz, N. (2007). Who is hard to reach and why? *ISR Working Paper*, (January), 1–7.
5. Smith, G. (2006). “Hard-to-reach groups don”t exist’. Retrieved from http://www.delib.co.uk/dblog/hard-to-reach-groups-don-t-exist
6. Tourangeau, R. (2014). Defining Hard to Survey Populations. In B. Edwards, R. Tourangeau, T. P. Johnson, K. M. Wolter, & N. Bates (Eds.), *Hard to Survey Populations*(pp. 3–21). Cambridge University Press.
7. Willis, G. B., Smith, T. W., Shariff-Marco, S., & English, N. (2014). Overview of the Special Issue on Surveying the Hard-to-Reach. *Journal of Official Statistics*, *30*(2), 171–176.