ASK THE EXPERTS

What are the conditions under which various survey designs that do not use probability samples might still be useful for making inferences to a larger population?

J. Michael Brick

1. INTRODUCTION

Researchers routinely collect data collection using a variety of methods and techniques. For making inference that can be generalized to a finite population, probability sampling is generally accepted as the most appropriate method. With a probability sample, every unit in the population has a known, non-zero chance of being sampled, and these probabilities are the basis for the inferences. Almost all official statistics have used this methodology for many years. It is the theory that is described in most sample survey textbooks

But probability sampling is not the only method for drawing samples and making inferences. Quota samples that only require samples meet target numbers of individuals with specific characteristics such as age and sex have been used for many years, especially in market research. In the last decade or so, widespread access to the Internet has resulted in a huge shift to online surveys. Many of these surveys draw their samples from "opt-in" panels comprised of large numbers of people to have "opted in" to do surveys. These types of surveys typically are not probability samples.

The popularity of the online, opt-in surveys is largely driven by cost. The cost per completed interview is generally a much lower than it would be for a probability sample, even if the probability sample uses a lower cost method such as mail. At the same time, probability samples around the world have been suffering due to rising nonresponse and concerns about the coverage of sampling frames, especially with the rise of cell phones. There are genuine concerns about the validity of inferences from a probability sample with significant undercoverage and high nonresponse. Is it still a probability sample when the basic sampling theory requires full coverage and response?

The next section briefly summarizes a review of non-probability sampling by a task force that I co-chaired. The last section discusses some recent work that continues this discussion.

2. AAPOR TASK FORCE

Reg Baker and I chaired a Task Force for the American Association of Public Opinion Research (AAPOR) "to examine the conditions under which various survey designs that do not use probability samples might still be useful for making inferences to a larger population." The task force completed its report in early 2013 and the full report can be downloaded www.aapor.org. The Journal of Survey Statistics and Methodology published a summary of the report, with comments from five experts in the field.

Our main conclusions are listed below. The details supporting these conclusions are omitted because of space limitations, but they are more informative than the headlines alone. I encourage you to read the report or the journal article.

i. Unlike probability sampling, there is no single framework that adequately encompasses all of non-probability sampling.
ii. Researchers and other data users may find it useful to think of the different non-probability sample approaches as falling on a continuum of expected accuracy of the estimates.
iii. Transparency is essential.
iv. Making inferences for any probability or non-probability survey requires some reliance on modeling assumptions.
v. The most promising non-probability methods for surveys are those that are based on models that attempt to deal with challenges to inference in both the sampling and estimation stages.
vi. One of the reasons model-based methods are not used more frequently in surveys may be that developing the appropriate models and testing their assumptions is difficult and time-consuming, requiring significant statistical expertise.
vii. Fit for purpose is an important concept for judging survey data quality, but its application to survey design requires further elaboration.
viii. Sampling methods used with opt-in panels have evolved significantly over time, and, as a result, research aimed at evaluating the validity of survey estimates from these sample sources should focus on sampling methods rather than the panels themselves.
ix. If non-probability samples are to gain wider acceptance among survey researchers there must be a more coherent framework and accompanying set of measures for evaluating their quality.
x. Although non-probability samples often have performed well in electoral polling, the evidence of their accuracy is less clear in other domains and in more complex surveys that measure many different phenomena.
xi. Non-probability samples may be appropriate for making statistical inferences, but the validity of the inferences rests on the appropriateness of the assumptions underlying the model and how deviations from those assumptions affect the specific estimates.

3. RECENT DEVELOPMENTS

Much of the survey research community continues to hold that only probability sampling can be used to make inferences to a population and that online non-probability sampling is not appropriate when inference is the goal (e.g., Bethlehem and Cooben, 2013). I would hope that even those who hold this opinion realize that there needs to be room for continued research into using online sample sources for all of the reasons cited above.

Second, applications continue to be published, and not all take the approach suggested by the task force to control both sampling and improve weighting. Here are two examples that differ drastically from the task force suggestion. Barratt, Ferris and Lenton (2014) explore an online sample for a rare subpopulation. They find that an online sample of this population differs in important ways from an offline probability sample of the same group. They do not describe their estimation methods, but suggest the online sample can be useful in addition to a probability sample. Wang et al. (2014) go in the opposite direction and use a sample of Xbox users that they know is not representative of voters in the U.S. elections. They rely fully on estimation methods and show they can produce estimates with small biases despite the problems with the sample. While we might not see these as the road to the type of applications that could easily be generalized, it does show the field is dynamic and will continue to advance in many ways.

There also are efforts to examine and control the quality of online surveys. For example, draft guidelines for online sampling quality are being developed by ESOMAR, the World Association for Social, Opinion and Market Research, and the Global Research Business Network.

This is clearly an area that is undergoing explosive and unpredictable growth. By the time the ink dries on any review like this one, something new is likely to be out there. As this work continues, it is still unclear whether online sampling will gain a firmer theoretical basis and become more acceptable for official statistics.

References

Baker, Reg, J. Michael Brick, Nancy A. Bates, Mike Battaglia, Mick P. Couper, Jill A. Dever, Krista J. Gile, and Roger Tourangeau. "Summary Report of the AAPOR Task Force on Non-probability Sampling." Journal of Survey Statistics and Methodology 1,
(2013): 90-143.

Barratt, Monica J., Jason A. Ferris, and Simon Lenton. "Hidden Populations, Online Purposive Sampling, and External Validity Taking off the Blindfold." Field Methods (2014): 1525822X14526838.

Bethlehem, Jelke, and Fannie Cooben. "Web Panels for Official Statistics?" Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong. Downloaded on May 1, 2014 from
http://2013.isiproceedings.org/Files/IPS064-P1-S.pdf.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman." Forecasting Elections with Non-Representative Polls." (preprint of article submitted to International Journal of Forecasting). Downloaded on May 1, 2014 from http://www.stat.columbia.edu/~gelman/research/unpublished/forecasting-with-nonrepresentative-polls.pdf.