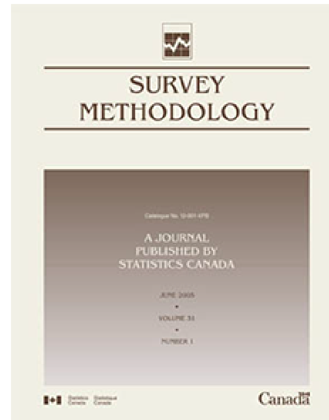
Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Authors' response to comments on "Exploring the assumption that commercial online nonprobability survey respondents are answering in good faith"



by Courtney Kennedy, Andrew Mercer and Arnold Lau

Release date: June 25, 2024

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2024
Vol. 50, No. 1, pp. 33-36
Statistics Canada, Catalogue No. 12-001-X

Authors' response to comments on "Exploring the assumption that commercial online nonprobability survey respondents are answering in good faith"

Courtney Kennedy, Andrew Mercer and Arnold Lau[1]

Abstract

Our comments respond to discussion from Sen, Brick, and Elliott. We weigh the potential upside and downside of Sen's suggestion of using machine learning to identify bogus respondents through interactions and improbable combinations of variables. We join Brick in reflecting on bogus respondents' impact on the state of commercial nonprobability surveys. Finally, we consider Elliott's discussion of solutions to the challenge raised in our study.

Key Words: Commercial nonprobability surveys; Survey panels; Benchmarking studies.

We thank the journal's leadership for hosting this dialogue and the discussants for offering their thoughtful comments. Each brings a unique perspective. Sen connects our study to other trends in survey statistics. Brick offers sobering reflections on the state of commercial non-probability surveys and helps to situate our work within that. Elliott advances discussion of solutions to the challenge raised in our study.

Sen observes that the demographic groups highlighted in our study were subjective and not exhaustive, as researchers could also look at education, geography, etc. We agree with the overall point and acknowledge that new insights could be gained from casting a wider net for variables correlating with bogus response. We also appreciate her pointing to machine learning as a possible means of identifying bogus respondents through interactions and improbable combinations of variables. The fact that machine learning is scalable and may be adaptive to changing respondent behavior makes it a potentially fruitful avenue for future research. On the other hand, we are skeptical that small area modeling and doubly robust estimators are likely to move the needle on accuracy. Past studies have found that for opt-in samples, such methods offer only marginal improvements over more common calibration methods (Mercer, Lau and Kennedy, 2018; Valliant, 2020). It may be that their limited utility stems from the fact that while such methods are excellent for correcting problems related to selection, they are poorly suited to address the problem of bogus respondents, which is fundamentally about measurement error.

Brick offers several high-level industry reflections that resonate with us. He notes, "The search for a high-quality opt-in panel is something that many pursued for over a decade, but the evidence thus far suggests this is a chimera." Indeed, our study along with Geraci (2022), Enns and Rothschild (2022) and others suggest that the emergence of such an opt-in panel is growing less likely not more. Previously, statisticians in this space focused on modeling to make commercial nonprobability samples more representative. Now they face the added challenge of determining which interviews are real and which are bogus.

1. Courtney Kennedy, Andrew Mercer and Arnold Lau, Pew Research Center, 1615 L St., NW, Washington D.C., Suite 800, 20036, U.S.A. E-mail: CKennedy@PewResearch.org.

We also appreciate Brick highlighting the role of the data supplier and how remarkable it is that the client (e.g., the researcher) bears the burden of identifying and remedying the types of errors we document. Our study suggests that data cleaning claims appearing on supplier websites give a false sense of protection from this threat. Indeed, if the suppliers' quality checks worked, bogus respondent wouldn't appear in client samples, and studies like ours wouldn't exist. It is imperative that researchers are aware of the threat posed by bogus respondents, particularly to domain estimates and full population estimates of rare outcomes. In our view, this threat has become so severe that researchers publishing point estimates using commercial non-probability samples should include a fulsome discussion of their approach for dealing with bogus respondents. Journal editors likely have a role in fostering that practice.

One of Brick's comments specific to our study was particularly intriguing. Reflecting on Table 3.2, he notes how the literature on program participation shows that the likelihood of false negative reporting tends to be significantly higher than the likelihood of false positive reporting. But Table 3.2 shows the opposite pattern in dramatic fashion for opt-in samples and in a more muted but still noticeable fashion for online panels recruited via address-based sampling (ABS). We agree with Brick that this contrarian finding indicates that online panels (both opt-in and ABS-recruited) perform differently than more rigorous probability-based samples on these outcomes. For opt-in panels, we have a reasonably strong hypothesis: bogus respondents tend to report in the affirmative (e.g., "Yes", "Agree") regardless of their true status because they want to qualify for future surveys and make more money. For online panels recruited by ABS, however, we are not aware of any hypothesis that would predict false positive reporting. Our suspicion is that the differences between rigorous probability samples and probability-based online panels are not fundamental differences in kind, and are likely a function of mode differences, panel conditioning and other well-known phenomena from the survey methods literature. That being said, we agree with Brick that identifying the precise mechanisms driving these differences seems like fertile ground for theoretical development and future research.

All discussants offered thoughts on possible solutions to the data quality problem explored in our study. As Brick's remarks suggest, one solution is to simply decide not to use commercial nonprobability samples. While they are undisputedly cheaper and faster, a sizable literature (e.g., Dutwin and Buskirk, 2017; KML; MacInnis, Krosnick, Ho and Cho, 2018; Pennay, Neiger, Lavrakas and Borg, 2018; Yeager, Krosnick, Chang, Javitz, Levendusky, Simpser and Wang, 2011) shows they are less accurate. With Brick, we do not endorse using opt-in samples and assuming one can weed out the bogus cases. We agree with his observation that, given sufficient motivation, bad actors will continue to find ways to circumvent inspection tools in online sources that allow people to opt-in to the process.

Sen raises the possibility of down-weighting respondents found likely to be bogus using detective questions. Prospects for that approach seem to depend on how much of the data provided by the bogus respondents is valid. For nonprobability samples in which the measurement error is likely to stem more from satisficing than fraud, this approach sounds promising. For commercial opt-in samples showing signs of bogus responding (e.g., cases answering “yes” regardless of the question), it is less clear that retaining bogus cases even in a down-weighted capacity would improve mean square errors. Fortunately, these are testable questions, and with Sen, we’d welcome a deeper look into this.

Elliott joins Wu (2022) in advocating for ongoing surveys rigorous enough to produce high quality benchmarks for use in calibrating less rigorous surveys. We enthusiastically second this proposal. At Pew Research Center, we have taken modest steps along these lines, creating an annual, multi-mode address-based survey designed to produce timely benchmark estimates for Americans’ political party affiliation, religious affiliation, and technology use (Pew Research Center, 2022). This multi-modal study reflects the highest rigor we can achieve with our institution’s resources, but much more enhanced designs (e.g., with an in-person stage of data collection) could be possible with the type of investment Elliott proposes. It is clear to us that such new benchmarking studies are needed to improve very low response rate probability-based samples like the three in our study. Whether benchmarking studies can rescue commercial non-probability samples is, to our minds, an open question given the challenge posed by respondents intentionally misreporting their status on both weighting and outcome variables.

References

Dutwin, D., and Buskirk, T.D. (2017). Apples to oranges or gala versus golden delicious? comparing data quality of nonprobability internet samples to low response rate probability samples. *Public Opinion Quarterly*, 81, 213-239.

Enns, P., and Rothschild, J. (2022). Do you know where your survey data come from? Outsourcing data collection poses huge risks for public opinion. Medium, available at https://medium.com/3streams/surveys-3ec95995dde2.

Geraci, J. (2022). *Poll-arized: Why Americans Don’t Trust the Polls and How to Fix Them Before It’s Too Late*. Houndstooth Press, p. 153.

MacInnis, B., Krosnick, J.A., Ho, A.S. and Cho, M. (2018). The accuracy of measurements with probability and nonprobability survey samples: Replication and extension. *Public Opinion Quarterly*, 82, 707-744.

Mercer, A., Lau, A. and Kennedy, C. (2018). For weighting online opt-in samples, what matters most? *Pew Research Center.* http://www.pewresearch.org/2018/01/26/for-weighting-online-opt-in-samples-what-matters-most/.

Pennay, D.W., Neiger, D., Lavrakas, P.J. and Borg, K. (2018). The online panels benchmarking study: A total survey error comparison of findings from probability-based surveys and nonprobability online panel surveys in Australia (2e Eds.) The Australian National University. https://csrm.cass.anu.edu.au/research/publications/methods-research-papers?search_term=The+online+panels+benchmarking.

Pew Research Center (2022). National Public Opinion Reference Survey (NPORS). Available at https://www.pewresearch.org/methods/fact-sheet/national-public-opinion-reference-survey-npors/.

Valliant, R. (2020). Comparing alternatives for estimation from nonprobability samples. *Journal of Survey Statistics and Methodology*, 8(2), 231-263.

Wu, C. (2022). Statistical inference with non-probability survey samples. *Survey Methodology*, 48, 2, 283-311. Paper available at https://www150.statcan.gc.ca/n1/pub/12-001-x/2022002/article/00002-eng.pdf. With discussion available at https://www150.statcan.gc.ca/n1/pub/12-001-x/12-001-x2022002-eng.htm.

Yeager, D.S., Krosnick, J.A., Chang, L., Javitz, H.S., Levendusky, M.S., Simpser, A. and Wang, R. (2011). Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opinion Quarterly*, 75, 709-747.