

## SAMPLING STRATEGIES FOR PROBABILITY SAMPLES USED TOGETHER WITH NON-PROBABILITY DATA

Lyndon Ang<sup>1</sup>, Robert Clark<sup>2</sup>, Bronwyn Loong<sup>3</sup>, and Anders Holmberg<sup>4</sup>

<sup>1</sup> Australian National University, Australia  
e-mail: [lyndon.ang@anu.edu.au](mailto:lyndon.ang@anu.edu.au)

<sup>2</sup> Australian National University, Australia  
e-mail: [robert.clark@anu.edu.au](mailto:robert.clark@anu.edu.au)

<sup>3</sup> Australian National University, Australia  
e-mail: [bronwyn.loong@anu.edu.au](mailto:bronwyn.loong@anu.edu.au)

<sup>4</sup> Australian Bureau of Statistics, Australia  
e-mail: [anders.holmberg@abs.gov.au](mailto:anders.holmberg@abs.gov.au)

### Abstract

There is a growing trend among statistical agencies to explore alternative data sources for producing more timely and more detailed statistics, while reducing costs and respondent burden. These data sources may include administrative records, big data (or “found data”) such as bank transactions or supermarket scanner data, and non-probability surveys such as online web panels. Coverage and measurement error are two issues that may be present in these types of data. These errors may be corrected using available auxiliary information relating to the population of interest, such as from a census or a reference probability sample.

In this paper, we discuss considerations for how a reference probability sample should be designed for the purpose of treating an imperfect data source. We consider in particular the case where the imperfect data relate to businesses. The multiple frame and cut-off sampling frameworks are explored in this context as alternatives to the usual optimal allocation for a single frame sample. A simulation study is conducted to examine the performance of various estimators under these frameworks.

**Keywords:** Non-probability sampling, Sample design, Multiple Frame, Cut-off Sampling

### References

- Beaumont, F.-F. (2020) Are probability surveys bound to disappear for the production of official statistics? *Survey Methodology*, 46 (1), 1-28.
- Lohr, S. L. (2021). Multiple-frame surveys for a multiple-data-source world. *Survey Methodology*, 47 (2), 229–263
- Wu, C. (2022). Statistical inference with non-probability survey samples. *Survey Methodology*, 48 (2), 283–311