



Mini Review

Copy Right@ Kyu-Seong Kim

Methodology of Non-probability Sampling in Survey Research

Kyu-Seong Kim*

Department of Statistics, University of Seoul, South Korea

*Corresponding author: Kyu-Seong Kim, Professor of Department of Statistics, University of Seoul, South Korea.

To Cite This Article: Kyu-Seong Kim. *Methodology of Non-probability Sampling in Survey Research*. *Am J Biomed Sci & Res*. 2022 - 15(6). *AJBSR*. MS.ID.002166. DOI: [10.34297/AJBSR.2022.15.002166](https://doi.org/10.34297/AJBSR.2022.15.002166)

Received: 📅 March 14, 2022; **Published:** 📅 March 21, 2022

Introduction

Since the mid20th century the probability sampling paradigm has become a mainstream methodology for sampling and inference in most surveys [1]. Especially large-scale national surveys conducted in national statistical offices or institutions are mostly based on this paradigm because objective statistics in the basis of this paradigm would be given to these institutions. Usually, probability sampling is subject to well-constructed frame, sampling design and high rate of response.

Recently, the probability sampling paradigm is faced with a great challenge due to decreasing population coverage rate and increasing non-response rate coupled with rising costs of sample surveys. Also, the number of sample surveys using non-probability samples like web survey is growing. In these situations, concerns about non-probability sampling paradigm as an alternative to probability sampling paradigm has been increasing [1,2].

Sample surveys with non-probability samples as well as probability samples has been carried out consistently. Non-probability samples have the merit of the faster speed of data collection, lower survey cost, and easier accessibility to the potential respondents. But lack of control of selection bias as well as the difficulty of statistical inference are the weakness of these samples. So, the overall use of non-probability samples is controversial in survey research area. Some of current dominant view of sampling are as follows, “researchers should avoid non-probability online panels when one of the research objectives is to accurately estimate population values [3].” or “statistical inference

is impossible without probability sampling or that the sampling method is irrelevant to inference [1].”

Nevertheless, non-probability samples have been commonly used in area of case-control study, clinical trial, observational study and so on. It is because of the research situation under which convenience or inevitability of non-probability samples is required. And with natural results, if the number of non-probability sample surveys is increasing, there will be a growing need for development of methodology based on non-probability samples.

Traditionally in the field of survey research, development of theory followed rather than driving realistic demands. As a typical example, surveys with sample have replaced the complete enumeration in the early 20th century. It is not because of the theoretical excellence of sample surveys, but because of the rapidly rising demands on much faster results through sample surveys. Then the theory of sample surveys has been established over time.

The theoretical development of non-probability sampling in survey research would go through a similar process as in sample surveys. If the number of surveys with non-probability samples is increasing, a corresponding theoretical development is expected. Such an expectation is hopeful because sampling theory is only a strategy not a dogma [4,5]. That is, the sampling theory is not an absolute principle, but a great strategy for obtaining an objective result in survey research. So, if we fully understand the principle of sampling as a strategy, then we can seek an appropriate methodology with non-probability samples in survey research.

Non-probability sampling in Survey Research

In survey research, randomization means the process of random allocation of units in experiments or random selection of sampling units in sample surveys. This randomization contributes two things to survey research. First, the objectivity of survey results may be guaranteed through randomization because researcher's subjective selection bias can be removed by randomization. This is a great contribution to survey research as well as science [6].

The next contribution is that the sampling distribution generated by randomization may provide a basis of statistical inference to survey research [6,7]. Such inference is called randomization-based inference or design-based inference. In the strict sense, randomization distribution is different from the distribution of uncertainty of things, so there is an argument that the inference based on randomization distribution is not valid even though randomization distribution itself is valid in the sense of sampling distribution [8].

Non-probability sampling is defined as a sampling, not a probability sampling [9]. It occurs if either the sample is not selected randomly or the inclusion probability of unit is unknown even under random sampling [9,10]. For example, quota sampling, judgment sampling, and volunteer sampling are considered as non-probability sampling.

By this definition, non-probability sampling is not free from selection bias by researcher and does not provide randomization distribution where theoretical inference takes place. Therefore, these two things should be considered in developing theories of non-probability sampling.

Methodology of Non-probability Sampling

Little is known about non-probability sampling methodology for controlling selection bias. Instead, if we recognize the existence of selection bias in non-probability sampling, we may think of two response strategies against that.

The first strategy is about sampling mechanisms that do not affect statistical inference. In such a mechanism, the non-probability sample does not cause selection bias [11,12]. In volunteer sampling, for example, if some characteristics of sample members are similar as those of non-sample members, then the problem of selection bias does not arise.

The second is to adjust the selection bias in the process of statistical inference after selecting a non-probability sample. This strategy may be classified into a pseudo-design-based framework as well as model-based framework [5]. Combinations of both frameworks are also possible afterward.

In the pseudo-design-based framework, non-probability samples are regarded as probability samples. But the design-weights are not available because the sampling process of the non-probability sample is unknown. In this framework, such unknown or undefined design weights are replaced by the corresponding surrogate weights called pseudo-design weights. Here pseudo-weights are usually constructed by using propensity weighting [13] or calibration weighting [14]. Sample estimates are then calculated using non-probability sample data with these pseudo-weights.

In contrast, the model-based framework uses the non-probability sample to fit a prediction model for the population. The predicted model is then used for estimation and inference on the population parameters [15,16].

Summary

Unlike the probability sampling framework, a single framework that encompasses the non-probability sampling has not been established yet. So non-probability sampling framework is still under controversy. Nevertheless, if the major form of sample surveys would be transferred from survey with probability samples to surveys with non-probability samples in this century, then, similarly to the previous century's sample survey, it is likely to be due to the soaring demand for non-probability sample surveys. Based on this trend of development, more theories related to non-probability sampling will be developed and supplemented. More useful research on non-probability sampling methodology is expected.

References

1. Baker R, Brick JM, Bates NA, Battaglia M, Couper MP, et al (2013) Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology* 1: 90-143.
2. Kim KS (2017) A study of non-probability sampling methodology in sample surveys. *Survey Research* 18: 1-29.
3. Baker R, Blumberg SJ, Brick JM, Couper MP, Courtright M, et al. (2010) AAPOR report on online panels. *Public Opinion Quarterly* 74: 711-781.
4. Kish L (1965) *Survey Sampling*. John Wiley and Sons.
5. Lenau S, Marachetti S, Munnich R, Pratesi M, Salvayi N, et al. (2021) Methods for sampling and inference with non-probability samples. Deliverable D11.8, Leuven, InGRID-2 project 730998-H2020.
6. Smith TMF (1983) On the validity of inferences from non-random samples. *Journal of the Royal Statistical Society Series A* 146: 394-403.
7. Cox DR (2006) *Principles of Statistical Inference*. Cambridge.
8. Royall RM (1983) Comment on an evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association* 78: 794-796.
9. Sarndal CE, Swenson B, Wretman J (1992) *Model assisted survey sampling*. Springer.
10. Statistics Canada (2010) *Survey methods and practices*. Catalogue no. 12-587-X.

11. Little RJA (1982) Model of nonresponse in sample surveys. *Journal of the American Statistical Association* 77: 237-250.
12. Sugden RA and Smith TMF (1984) Ignorable and informative designs in survey sampling. *Biometrika* 71: 495-506.
13. Rosenbaum PR and Rubin DB (1983) The central role of the propensity score in observational studies for casual effects. *Biometrika* 70: 41-55.
14. Chen JKT, Valliant RL and Elliott MR (2018) Model-assisted calibration of non-probability sample survey data using adaptive LASSO. *Survey Methodology* 44: 117-144.
15. Kim JK, Park S, Chen Y and Wu C (2021) Combining non-probability and probability survey samples through mass imputation. *Journal of the Royal Statistical Society Series A* 184: 1-23.
16. Chen Y, Li P and Wu C (2020) Doubly robust inference with non-probability survey samples. *Journal of the American Statistical Association* 115: 2011-2021.