

Estimators with Combined Probability and Nonprobability Samples Using Small Area Models

Yang, Michael, Nada Ganesh, Edward Mulrow, Evan-Herring Nathan, and Vicki Pineau
NORC at the University of Chicago, 4350 East-West Highway, 8th Floor, Bethesda, MD 20814

Abstract

Through case studies and simulations, the authors have evaluated a range of estimation methods for combining probability and nonprobability samples. Our earlier evaluations show that the Small Area Modeling method, a doubly robust method developed at NORC, tends to achieve greater bias reduction than the other methods, especially for the modeled response variables with large known biases associated with the nonprobability sample. Meanwhile, Statistical Matching and Propensity Weighting also demonstrate good properties. This paper expands our earlier studies to explore hybrid methods that integrate Statistical Matching, Propensity Weighting, and Small Area Modeling. Specifically, it reports comparative evaluations of four methods: (1) Matching Imputation Weighting without Small Area Modeling, (2) Matching Propensity Weighting without Small Area Modeling, (3) Matching Imputation Weighting with Small Area Modeling, and (4) Matching Propensity Weighting with Small Area Modeling. Evaluations of these methods are based on estimates of bias, confidence interval coverage, and effective sample size.

Key words: Nonprobability sample estimation, statistical matching, propensity weighting, small area estimation, doubly robust

1. Introduction

While probability sampling remains the gold standard for sample survey estimation, there has been growing interest in methods that combine probability and nonprobability samples in order to improve cost efficiency. Nonprobability samples may provide a lower cost alternative to probability samples; however, estimates based on nonprobability samples may be biased due to unknown coverage and selection biases.

In this paper, we present some results from our comparative analysis based on a Monte Carlo simulation study. Specifically, we compare bias, confidence interval coverage, and effective sample size associated with four different estimation methods that combine probability and nonprobability samples.

2. Methods Evaluated

Survey researchers and practitioners have proposed a range of estimation methods based on combined probability and nonprobability samples (Valliant 2020; Chen et al., 2019; Elliot and Valliant 2017; Kim and Hazia 2014). In our past evaluations (Ganesh et al., 2017; Yang et al., 2018, 2019, 2020; Mulrow et al., 2020), we reported comparative analysis of the following methods:

- Calibration: Calibrate total estimates to known population benchmarks
- Statistical Matching: Statistically match nonprobability and probability samples to derive pseudo weights
- Superpopulation Modeling: Use a linear superpopulation model to derive population estimates
- Propensity Weighting: Model the propensity of inclusion in a nonprobability sample

- Small Area Modeling: a doubly robust method that uses small area estimation to improve weighting adjustments

Based on our earlier evaluations using case studies and simulations, Statistical Matching, Propensity Weighting, and Small Area Modeling consistently outperform Calibration and Superpopulation Modeling (Yang et al., 2018, 2019, 2020). In this paper, we evaluate four hybrid methods that feature some combination of Statistical Matching, Propensity Weighting, and Small Area Modeling.

We start by matching each nonprobability sample unit to a probability sample unit using the R StatMatch package (D'Orazio, 2017). Statistical matching is carried out using a nearest neighbor hot deck algorithm based on a distance measure. The matching process resembles imputation in the sense that a donor from the probability sample is matched to a recipient from the nonprobability sample based on a set of matching variables (Bethlehem, 2015). Distances are measured using Gower's dissimilarity measure, which can use both categorical and continuous variables in the dissimilarity calculation. The matching variables are determined using Gradient Boosting (D'Orazio, Di, and Scanu, 2006).

Statistical matching divides the probability sample into two sets:

S_P^M : the set of probability sample units matched to a nonprobability sample unit

S_P^U : the set of probability sample units that are not matched

We develop an estimator that takes into account the matching pattern, as follows.

Let X_i be a variable of interest for respondent i , r_i . The combined estimator of the population total is:

$$\hat{X}_{comb} = \hat{X}_P^U + \lambda \hat{X}_P^M + (1 - \lambda) \hat{X}_{NP}, \text{ where } 0 \leq \lambda \leq 1$$

We assume that the matching pattern reflects the degree of undercoverage associated with the nonprobability sample. The population total is therefore expressed as the total of the unmatched portion plus the total of the matched portion of the population. For the unmatched portion, the only data source is the probability sample. For the matched portion, we have both a probability and a nonprobability sample and the total is estimated through a composite estimator with a combination factor λ . This combined estimator implies that we weight the combined sample as follows:

$$w^{comb}_i = \begin{cases} w_{Pi} & r_i \in S_P^U \\ \lambda w_{Pi} & r_i \in S_P^M \\ (1 - \lambda) w_{NPi} & r_i \in S_{NP} \end{cases}$$

The unmatched probability units retain their original weights, while the matched probability and nonprobability units receive a weight adjustment governed by λ .

We now consider two alternative methods for deriving the nonprobability sample weights w_{NPi} which are still unknown.

With *Matched Imputation Weighting*, we impute the nonprobability sample weights w_{NPi} as the weight of the matched probability sample unit. When a probability unit is matched to multiple nonprobability units, each matched nonprobability unit weight is imputed as the probability unit weight divided by the number of matches.

With *Matched Propensity Weighting*, we develop the nonprobability sample weights w_{NPi} via the following steps:

- Concatenate S_P^M and S_{NP}
- Create a dichotomous indicator variable, Y , 1 for nonprobability sample units S_{NP} and 0 for matched probability units S_P^M
- Fit a logistic regression model to predict the probability of inclusion for the nonprobability sample units
- The weight for nonprobability sample unit i , w_{NPi} , is computed as the reciprocal of the predicted probability

Using propensity models to develop nonprobability sample weights has been quite common in practice. The difference here is that our propensity model uses only the matched probability sample instead of the entire probability sample as the reference sample.

We now consider integrating Small Area Modeling in order to further improve the nonprobability sample weights. Small area estimation methods are used to jointly model domain-level estimates for one or more key response variables from the probability and the nonprobability sample (Ganesh et al., 2017). The model includes a set of covariates, fixed and random bias terms, and domain-level random effects. The combined sample weights are developed via the following steps:

- A Bivariate Fay-Herriot model (Rao, 2003; Fay and Herriot, 1979) is used to jointly model the weighted domain-level point estimates for a set of key response variables from the probability sample (y_d^P) and the nonprobability sample (y_d^{NP}):

$$y_d^P = \mathbf{x}'_d \beta + v_d + \varepsilon_d^P$$

$$y_d^{NP} = b + \alpha_d^{NP} + \mathbf{x}'_d \beta + v_d + \varepsilon_d^{NP}$$

- d is a demographic group (e.g. 18-34 year old, male, Hispanic)
 - \mathbf{x}_d is a vector of covariates
 - v_d 's are domain level random effects
 - b is a fixed effect bias term associated with the nonprobability sample estimate
 - α_d 's are random effect bias terms associated with the nonprobability sample estimate
 - $\varepsilon_d^P, \varepsilon_d^{NP}$ are the sampling errors associated with y_d^P, y_d^{NP}
- Predicted small area estimates for each domain are obtained using an Empirical Best Linear Unbiased Predictor (EBLUP).
 - Combined sample weights are derived such that the weighted sample estimates based on the combined sample match regular demographic benchmarks as well as the small area estimates for each domain for the key response variables.

The small domains are typically defined by cross-classifying a set of demographic variables such as age, gender, education, race/ethnicity, etc.

We considered two methods for developing nonprobability sample weights: matched imputation weighting and matched propensity weighting, and we then considered applying small area modeling. Therefore, our current evaluation compares these four methods:

- Matched Imputation Weighting with Small Area Modeling
- Matched Propensity Weighting with Small Area Modeling
- Matched Imputation Weighting without Small Area Modeling
- Matched Propensity Weighting without Small Area Modeling

3. Monte Carlo Simulation Setup

As described in Benoit-Bryan and Mulrow (2021), simulation samples were generated using survey completes from a large-scale national study, *Culture & Community in a Time of Crisis (CCTC): A Special Edition of Culture Track*, a survey of the general public and culturally active Americans. The full population frame (Frame 1) for the simulation consists of 113,549 U.S. adult survey completes. From this frame, 1,000 stratified random samples of size 1,000 were selected. These are considered probability samples.

To create the nonprobability sample frame, a set of records from the full population frame were removed based on some descriptive variables in the file. Additional records were removed based on a highly skewed binomial distribution, resulting in a skewed nonprobability sample frame with 74,202 records. From this frame (Frame 2), 1,000 samples of size 4,000 were selected. These are considered nonprobability samples.

Both the probability and nonprobability samples contain a large number of demographic variables and survey response variables. CCTC response variables of interest are categorical by nature. Therefore, statistics related to these variables are proportions related to the categories of each variable. For each response variable of interest, let $P_{Frame\ 1}$ and $P_{Frame\ 2}$ be the population proportion computed from the probability and nonprobability frame, respectively. The *known absolute bias* associated with the nonprobability frame, B_{pop} , is calculated as the difference of population proportions between the two frames:

$$B_{pop} = |P_{Frame\ 1} - P_{Frame\ 2}|$$

Information on the B_{pop} for 12 behavioral and two attitudinal variables that are used to evaluate the properties of the estimation methods are provided in Benoit-Bryan and Mulrow (2021).

4. Simulation Results

Each of the four methods is used to develop a set of combined weights for the probability and nonprobability samples, and these weights are used to derive combined sample estimates. In this section, we evaluate the properties of these estimates.

4.1 Bias reduction

For each outcome variable, the estimated bias associated with iteration m is defined as the difference between the combined estimate per iteration $\hat{p}_{Comb,m}$ and the true population proportion $P_{Frame\ 1}$,

$$\hat{b}_m = P_{Frame\ 1} - \hat{p}_{Comb,m}$$

Absolute estimated bias is the average absolute bias across the 1,000 iterations:

$$\bar{b} = \frac{1}{1000} \sum |\hat{b}_m|$$

Figures 1-3 plot B_{pop} and \bar{b} to show the magnitude and direction of bias reduction under each of the four methods for the outcome variables, where the outcome variables are ordered by the size of Absolute Known Bias.

The magnitude and direction of bias reduction for each variable is represented by a vector, with the starting point representing the Absolute Known Bias and the ending point the Absolute Estimated Bias. Green vectors indicate bias reduction and orange vectors indicate bias increase. For each method, the dashed vertical line represents the average Absolute Estimated Bias across the outcome variables.

In general, all four methods achieved great bias reduction for the outcome variables as the Absolute Estimated Bias is very small for all variables under all methods. Bias increased slightly for a few variables, but all these variables have small known bias to begin with and the Absolute Estimated Bias remains very small despite the bias increase. Note that, Matched Propensity without Small Area Modeling led to slightly larger average bias than the other methods; even so the bias remains small.

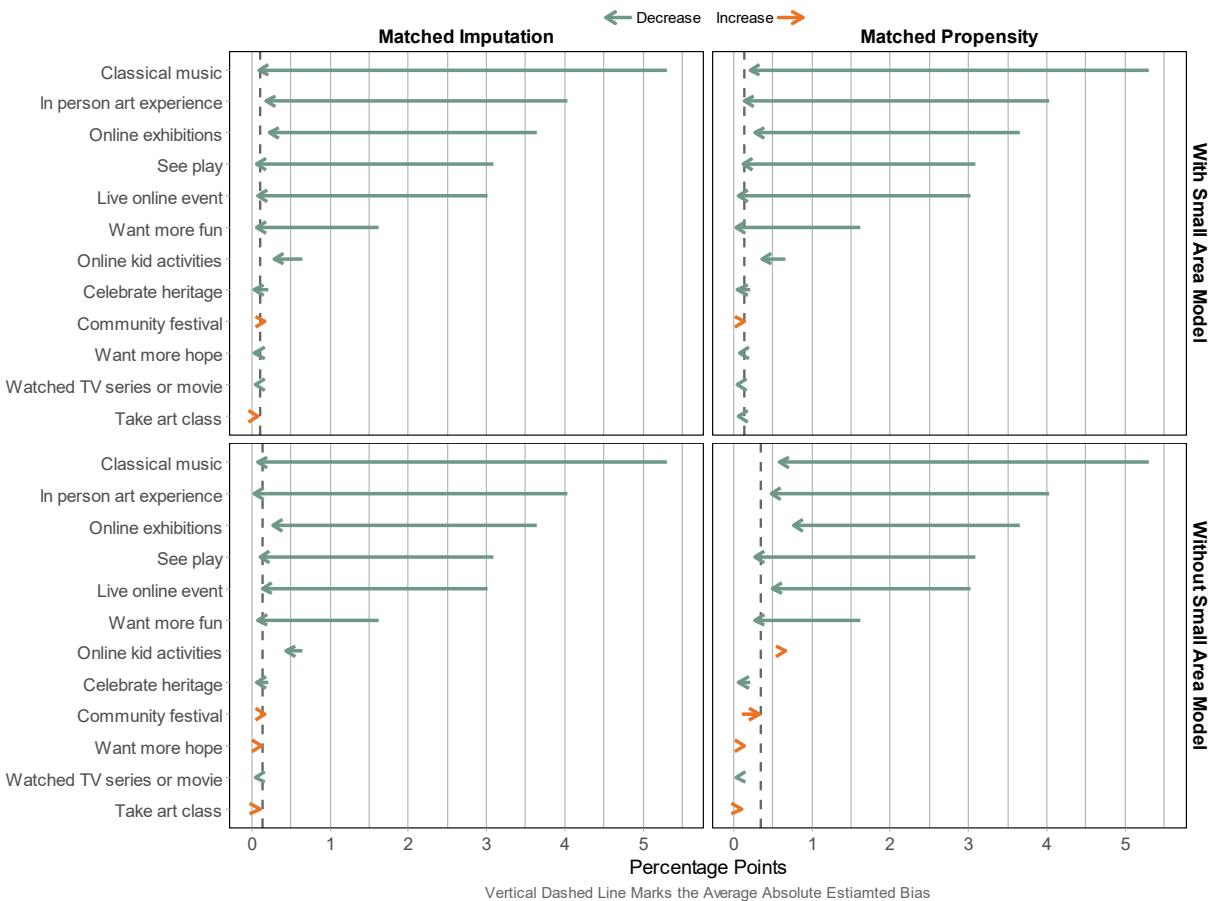


Figure 1: Bias reduction for the behavioral variables.

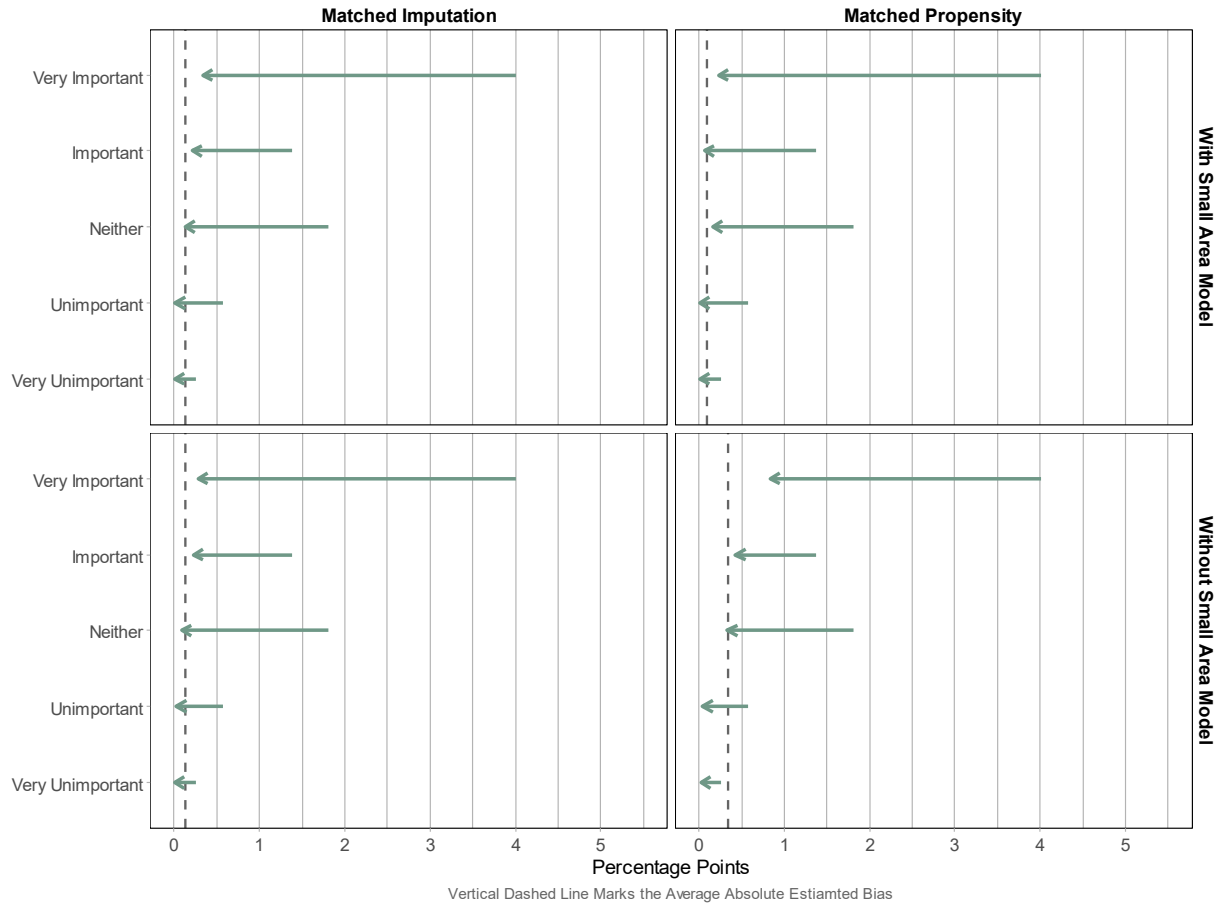


Figure 2: Bias reduction for attitudinal variable Q17. During COVID-19, how important are arts & culture organizations?

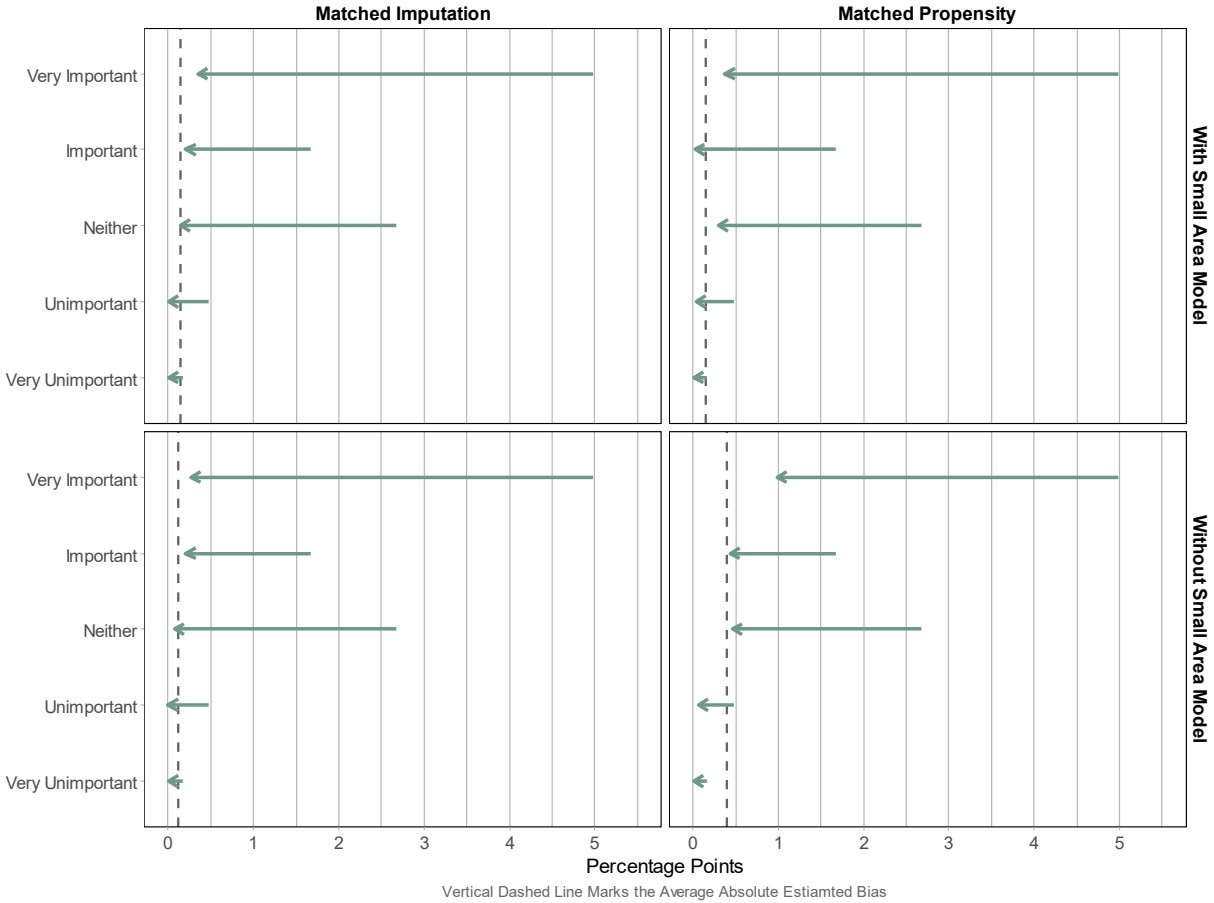


Figure 3: Bias reduction for attitudinal variable Q18. Before COVID-19, how important were arts & culture organizations?

4.2 Confidence interval coverage

For each outcome variable, we compute the standard error of the combined estimate as

$$\widehat{SE}_{Comb} = \sqrt{\frac{1}{1000} \sum (\hat{p}_{Comb,m} - \bar{\hat{p}}_{Comb})^2}$$

We then construct a 95 percent confidence interval around the combined estimate per iteration $\hat{p}_{Comb,m}$ using \widehat{SE}_{Comb} as standard error.

Finally, we calculate the 95 percent confidence interval coverage rate as the percentage of the 1,000 intervals that contain the population true value. The results are reported in Figures 4-6.

In Figures 4-6, the dashed vertical line represents the average confidence interval coverage rate per method across the estimates. All coverage rates are quite close to the nominal 95%. However, Matched Propensity without Small Area Modeling gives slightly lower average coverage and there is more variation and less consistency across variables.

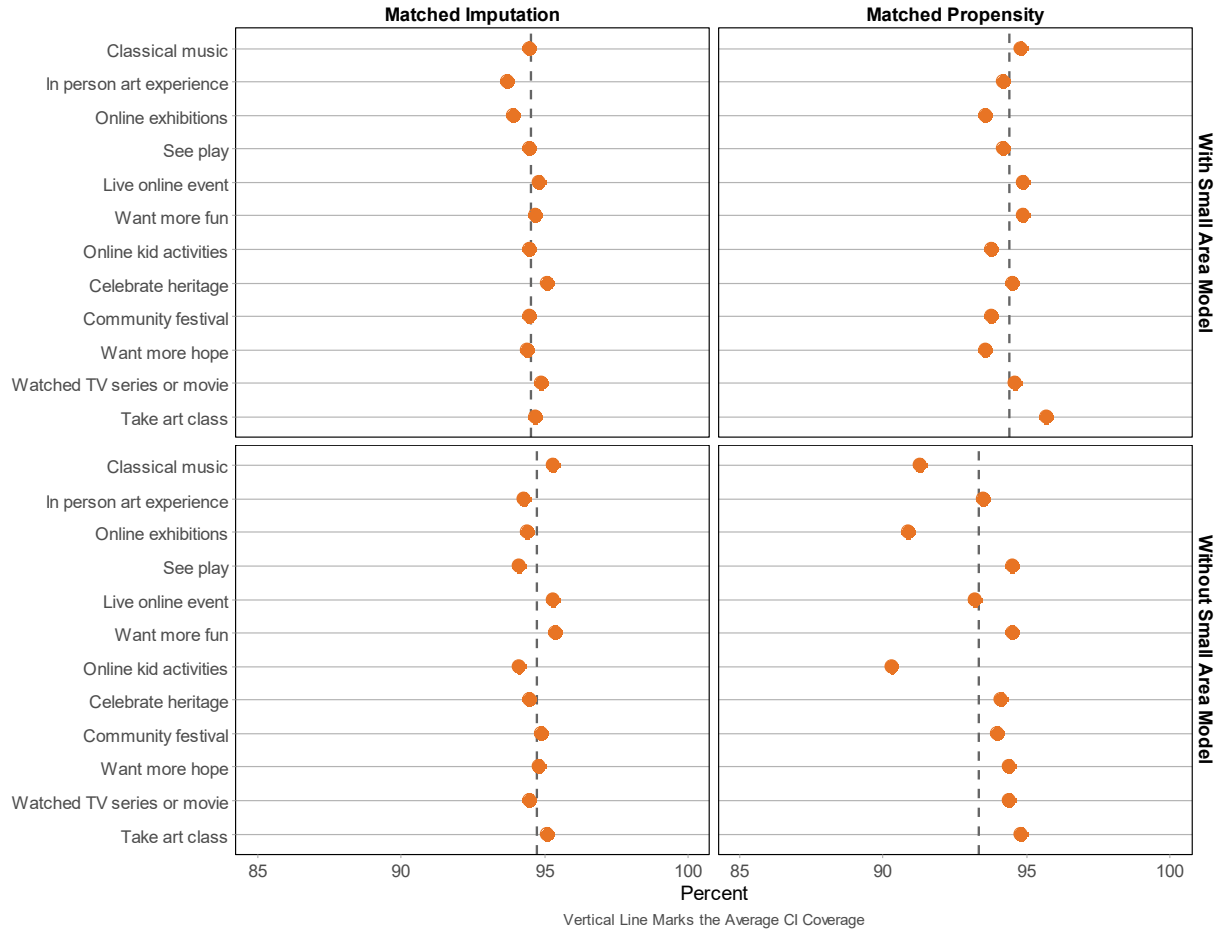


Figure 4: 95 percent confidence interval coverage for behavioral variables.

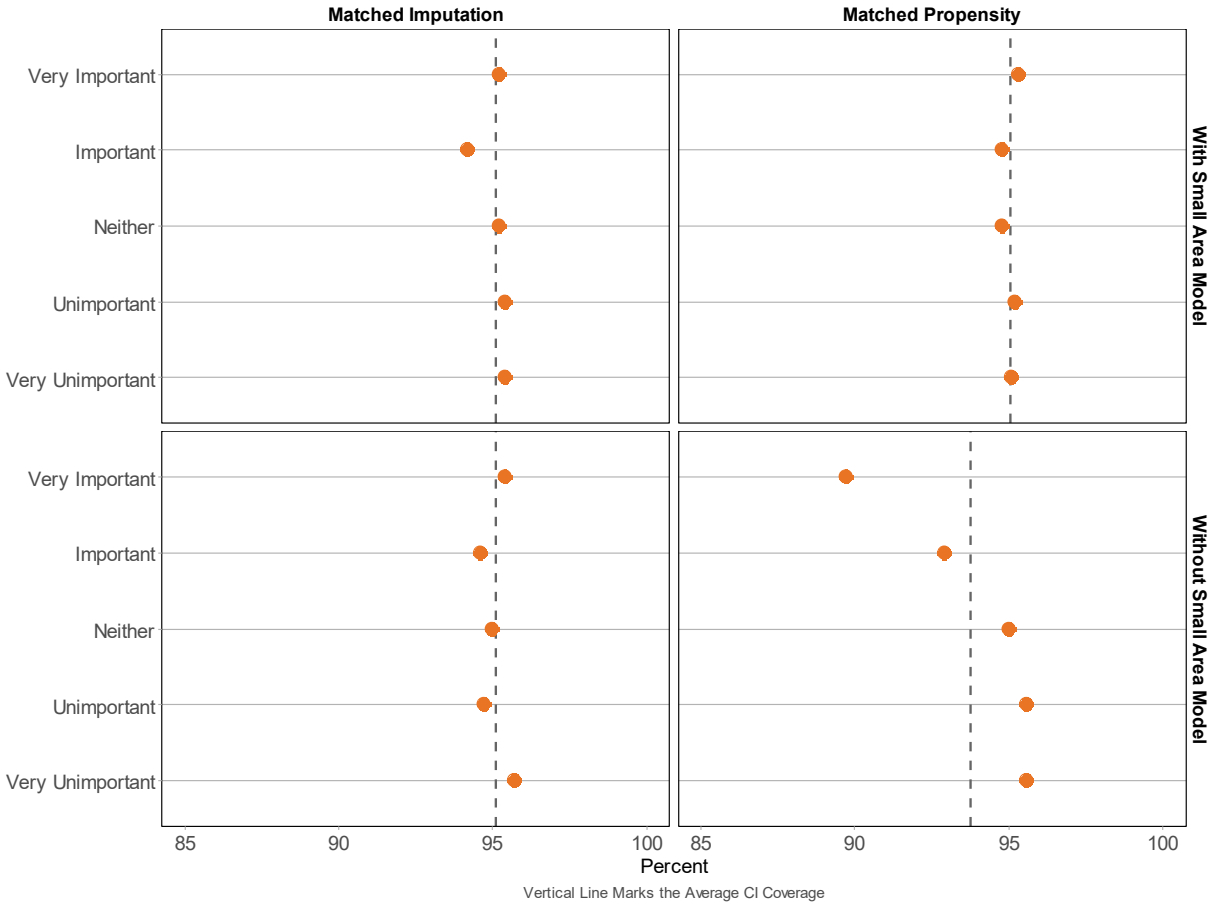


Figure 5: 95 percent confidence interval coverage for attitudinal variable Q17. During COVID-19, how important are arts & culture organizations?

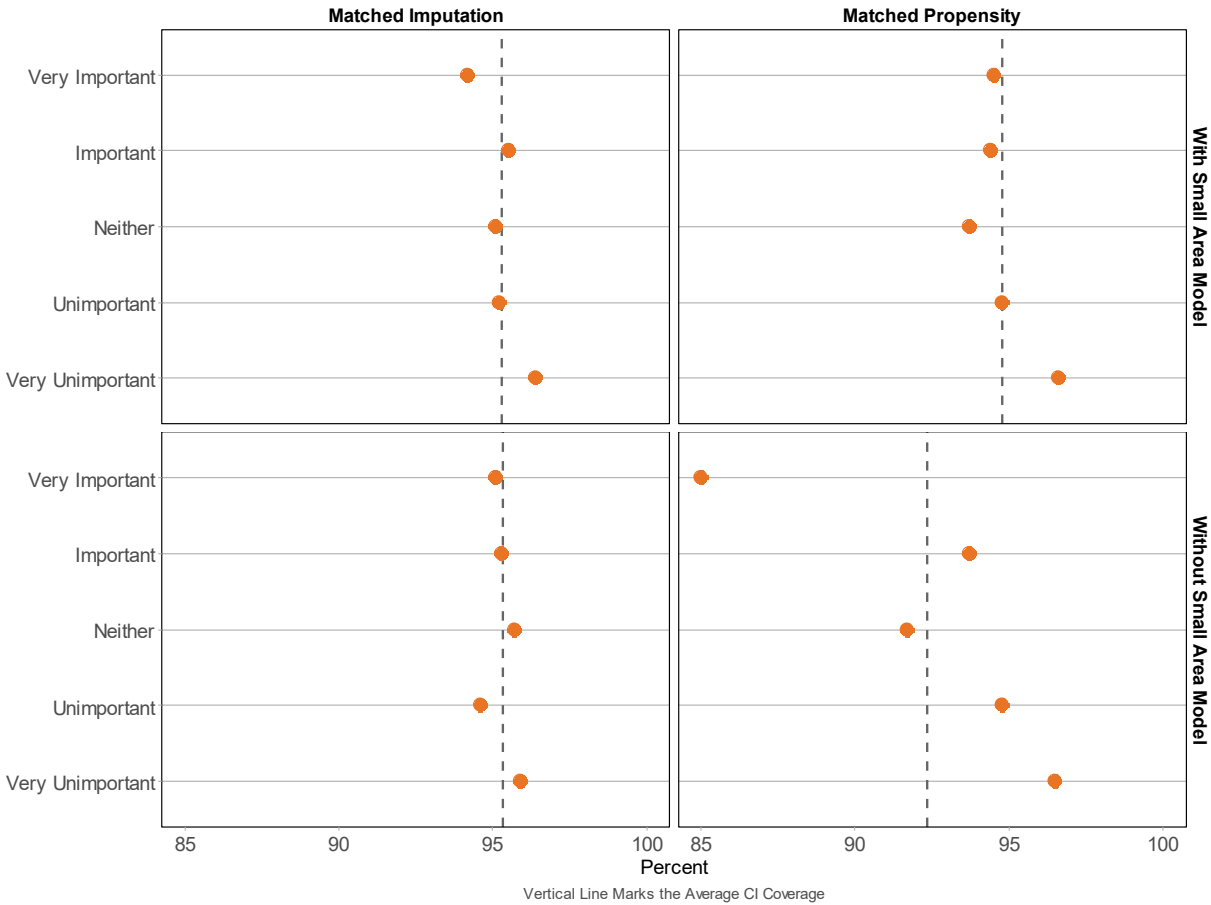


Figure 6: 95 percent confidence interval coverage for attitudinal variable Q18. Before COVID-19, how important were arts & culture organizations?

4.3 Effective sample size ratio

For each outcome variable, we compute an effective sample size ratio under each method, defined as the effective sample size for the combined sample divided by the effective sample size for the probability sample only. Effective sample size is the nominal sample size divided by design effect due to weighting. We use this ratio to measure the marginal contribution of the nonprobability sample. It turns out that the effective sample size ratio is also the mean squared error ratio.

The effective sample size ratio is,

$$\frac{n_{E(Comb)}}{n_{E(P)}} = \frac{5000/DEFF_{Comb}}{1000/DEFF_P} = \frac{\widehat{MSE}_P}{\widehat{MSE}_{Comb}}$$

where $\widehat{MSE}_P = \frac{1}{1000} \sum (\hat{p}_{P,m} - P_{Frame 1})^2$ is the mean squared error with the probability sample, and $\widehat{MSE}_{Comb} = \frac{1}{1000} \sum (\hat{p}_{Comb,m} - P_{Frame 1})^2$ is the mean squared error with the combined sample. Figures 7-9 show the results of the effective sample size ratio comparisons.

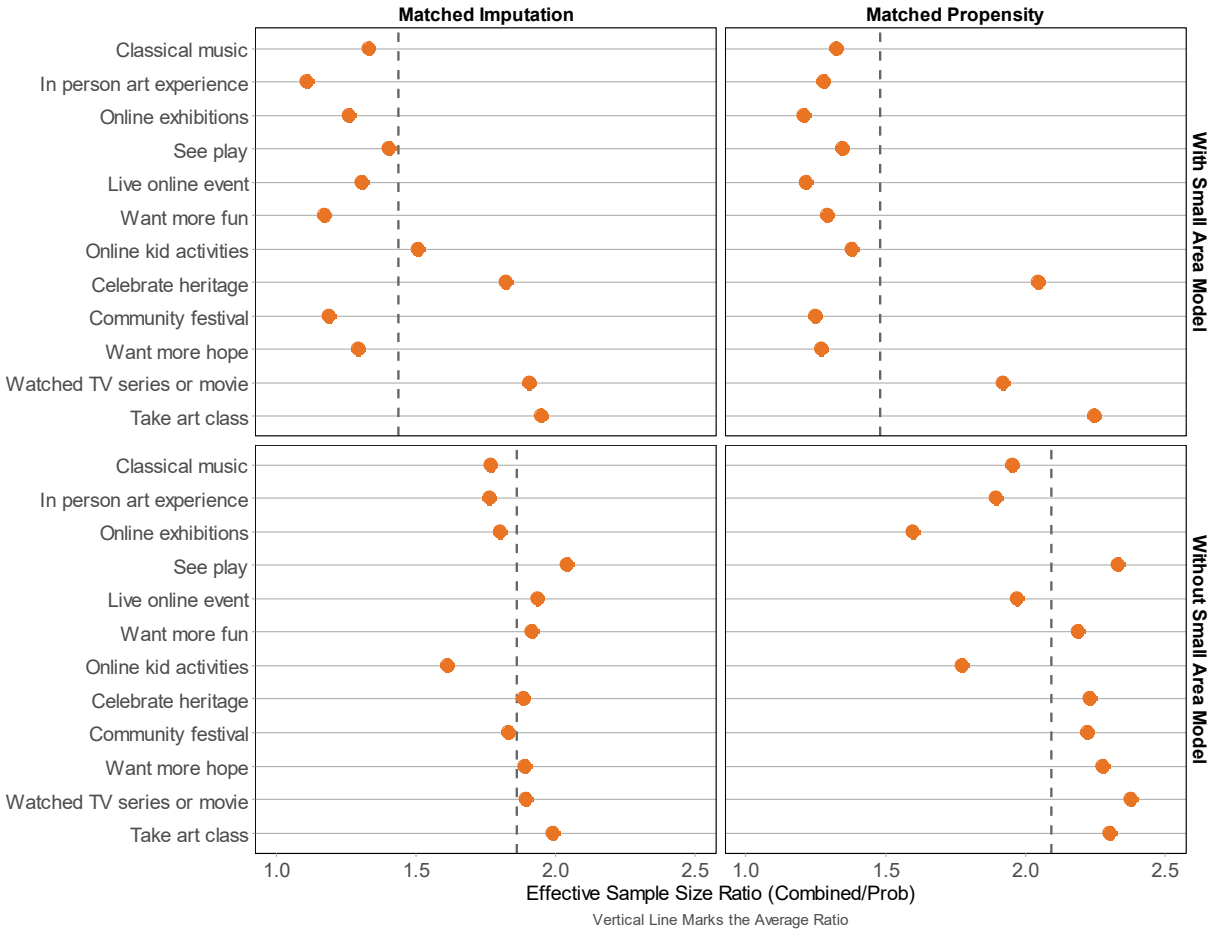


Figure 7: Effective sample size ratio for behavioral variables.

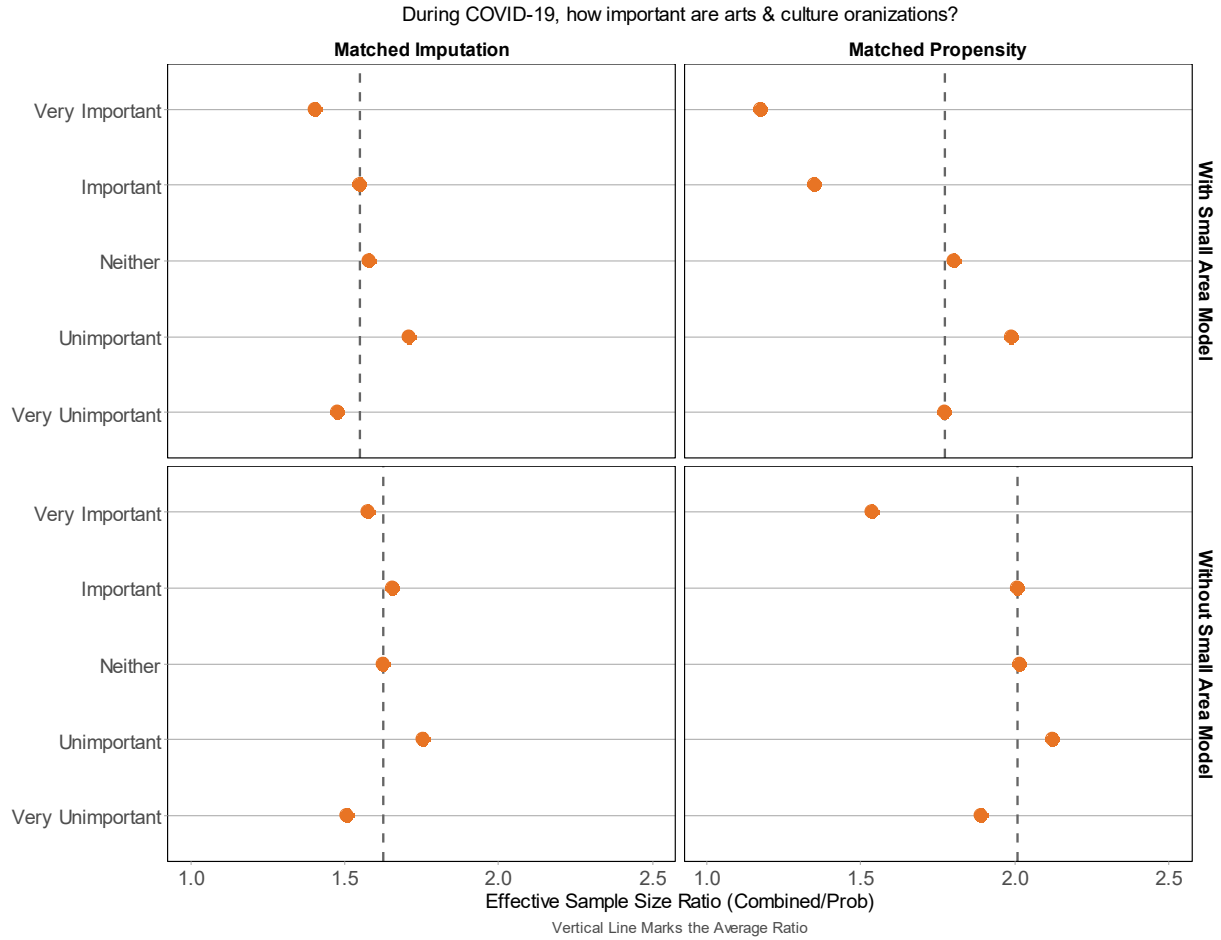


Figure 8: Effective sample size ratio for attitudinal variable Q17. During COVID-19, how important are arts & culture organizations?

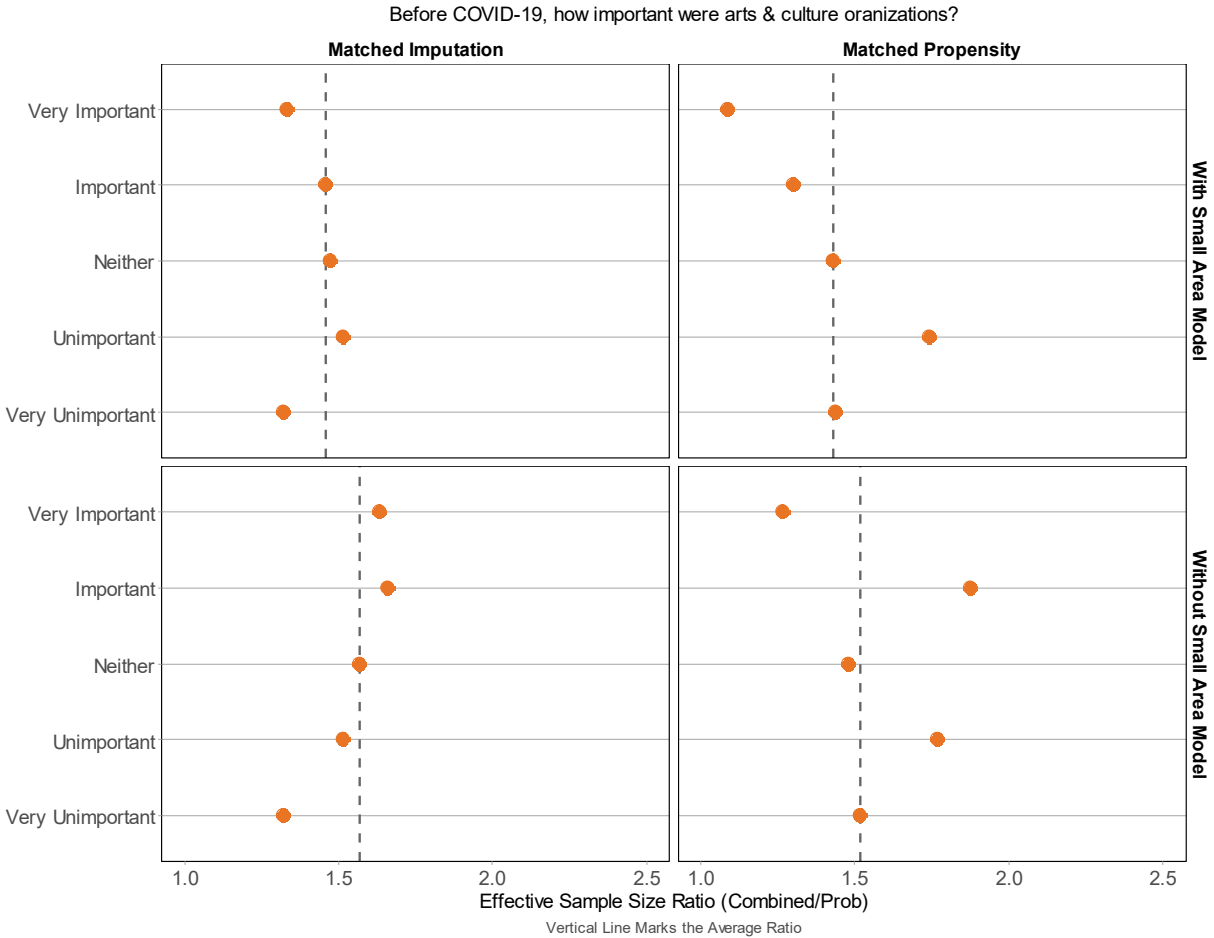


Figure 9: Effective sample size ratio for attitudinal variable Q18. Before COVID-19, how important were arts & culture organizations?

Effective sample size ratios are all greater than 1, which means that the addition of the nonprobability sample helped to increase the effective sample size of the combined sample in all cases. However, the increase is quite modest. We brought in 4,000 additional nonprobability cases, but only increased the effective sample size by about 50 percent. There may be a point of diminishing returns for the size of the nonprobability sample, which needs to be investigated in relation to the cost model. Increasing the nonprobability sample size beyond a certain point may not increase statistical or cost efficiency.

Among the four methods, Matched Propensity Weighting without Small Area Modeling gives the largest effective sample size ratio. In general, without Small Area Modeling did better than with Small Area Modeling on this measure. Both Statistical Matching and Small Area Modeling tend to produce greater bias reduction but larger weight variations per our earlier investigations (Yang et al., 2018). Therefore, it should not be all that surprising to see only modest increase in effective sample size. If we consider bias reduction as the major objective in nonprobability sample estimation, we would prefer methods that generate greater bias reduction. However, we need to better understand the complex tradeoffs between bias and variance. Furthermore, the major advantage of Small Area Modeling is that it produces more reliable estimates for small domains, which may also require greater weighting adjustments and need to be investigated.

5. Summary and future research

In this study, we evaluated the properties of combined probability and nonprobability sample estimates based on four estimation or weighting methods. Our simulations show that all four methods result in significant bias reductions. Methods with Small Area Modeling led to greater bias reduction than those without Small Area Modeling. Meanwhile, confidence interval coverage rate for all methods is close to the nominal 95%, except that Matched Propensity Weighting without Small Area Modeling achieved lower and less consistent coverage rate across the outcome variables. However, methods with Small Area Modeling resulted in smaller effective sample sizes. Therefore, no obvious best method emerged from these simulations and our research will continue.

Our current simulations use a ratio of 1:4 between probability and nonprobability samples. In future research, we intend to investigate estimator properties with different probability and nonprobability sample size ratios. Given the modest increase in effective sample size observed here, we believe a smaller sample size ratio may be more efficient.

The standard error of the combined estimates is currently based on the variance of estimates across the 1,000 iterations. In future research, we will also consider the variance of estimates within each iteration, which may be estimated via a jackknife procedure (Yang et al., 2020).

Small Area Modeling is primarily designed to enable better small domain estimation. As such, we would like to expand our current investigation to compare combined estimates for small domains under the different methods. Generating more reliable estimates for small domains may be considered important for some studies and justify some increase in overall variance.

Finally, we would like to investigate the impact of weight trimming on estimator properties. In general, methods that lead to greater bias reductions also tend to produce larger weight variations. Therefore, it is important to understand the potentially complex bias-variance tradeoffs to inform decision making in practice.

References

- J. Benoit-Bryan and E. Mulrow (2021) “Exploring Nonprobability Methods with Simulations from a Common Data Source: Culture and Community in a Time of Crisis.” To appear: *Joint Statistical Meetings 2021 Proceedings*.
- Bethlehem J. (2015) “Solving the nonresponse problem with sample matching?” *Social Science Computer Review*, Vol. 34, Issue 1, pp. 59 – 77.
- Chen, Yilin, Pengfei Li, Changbao Wu. (2019). “Doubly Robust Inference with Non-probability Survey Samples,” Paper presented at the Joint Statistical Meetings, Denver, Colorado.
- D’Orazio, M. (2017). StatMatch: Statistical Matching. R package version 1.2.5. <https://CRAN.R-project.org/package=StatMatch>.
- D’Orazio, M., Di Zio M., Scanu M. (2006) *Statistical matching: Theory and practice*. Wiley, Chichester.
- Elliot, M. R., Valliant, R. (2017). “Inference for Nonprobability Samples,” *Statistical Science* 2017, Vol. 32, No. 2, 249–264.
- Fay, R.E., and Herriot, R.A. (1979). “Estimates of income for small places: An application of James-Stein procedures to Census data,” *Journal of the American Statistical Association*, v. 74 (366), pp. 269-277.
- Ganesh, N., Pineau, V., Chakraborty, A., Dennis, J.M., (2017). “Combining Probability and Non-Probability Samples Using Small Area Estimation.” *Joint Statistical Meetings 2017 Proceedings*.
- Kim, J.K., and Haziza, D. (2014), “Doubly Robust Inference with Missing Data in Survey Sampling,” *Statistica Sinica*, 24, 375-394.

- Mulrow, Edward, Nada Ganesh, Vicki Pineau, Michael Yang. (2020). "Using Statistical Matching to Account for Coverage Bias When Combining Probability and Nonprobability Samples," *Joint Statistical Meetings 2018 Proceedings*.
- Rao, J.N.K. (2003). *Small Area Estimation*, John Wiley & Sons, Inc.
- Rivers, D. (2007), "Sampling for Web Surveys", Proceedings of the Survey Research Methods Section, Joint Statistical Meetings, American Statistical Association, Alexandria, VA, 1-26.
- Valliant, R. (2020). "Comparing Alternatives for Estimation from Nonprobability Samples," *Journal for Survey Statistics and Methodology*, vol. 8, 231-263.
- Yang, Y. Michael, Nada Ganesh, Ed Mulrow, and Vicki Pineau. (2018). "Estimation Methods for Nonprobability Samples with a Companion Probability Sample," *Joint Statistical Meetings 2018 Proceedings*.
- Yang, Y. Michael, Nada Ganesh, Ed Mulrow, and Vicki Pineau. (2019). "Evaluating Estimation Methods for Combining Probability and Nonprobability Samples through a Simulation Study," *Joint Statistical Meetings 2019 Proceedings*.
- Yang, Y. Michael, Nada Ganesh, Ed Mulrow, and Vicki Pineau. (2020). "Variance Estimation for Combined Probability and Nonprobability Samples," *Joint Statistical Meetings 2020 Proceedings*.