WHITE PAPER
March 2026

NORC's TrueNorth Calibration tool for probability and nonprobability samples: New Version 3.0 continues the state of the art and science of opt-in sample data quality.

TRUENORTH
Calibration Tool

NORC
at the UNIVERSITY of CHICAGO

The survey research landscape seems to encounter new challenges every decade. Two ongoing concerns with probability-based surveys are declining response rates and rising costs of data collection. As a result, individuals and businesses often turn to nonprobability-based surveys because they have lower costs and faster turnaround times. However, the quality of data from nonprobability surveys is a major concern[1]. Compared to probability surveys, nonprobability surveys suffer from serious representation issues, with samples missing key segments of study populations and over-representing other segments, and measurement errors, such as response data from bots, fake respondents, or satisficing respondents. It is not uncommon today for over half of respondents from nonprobability surveys to be discarded[2]. As a result, estimates from nonprobability surveys are biased and less accurate, and these coverage and measurement errors cannot be corrected through typical demographic weighting or quota approaches. How can trust be restored to this type of research?

TrueNorth® calibration is NORC's solution for improving data quality by combining the proven accuracy of probability-based research with the scalability of low-cost nonprobability samples. It uses a unique approach that is simple in concept but highly sophisticated in its statistical execution. The foundation of TrueNorth calibration is NORC's probability-based AmeriSpeak® Panel, the only probability household panel in the U.S. to use in-person/face-to-face household recruitment. By using the AmeriSpeak Panel, we know who the probability sample respondents are, and AmeriSpeak attains survey response rates greater than all other modern research panels. TrueNorth calibration reduces the bias of nonprobability samples at not only the topline level but also deep within key demographic groups.[3] Moreover, the approach is tailored to the topic of each survey to reduce bias. It is particularly useful for situations when more interviews are needed than AmeriSpeak or any probability solution at a given cost can provide, or alternatively, when a budget might not support solely using a high-quality approach like AmeriSpeak.

What is TrueNorth?

The TrueNorth process solves several problems inherent to nonprobability samples and creates a pseudo-probabilistic and less biased sample than nonprobability samples alone. This is mainly achieved by blending a higher-quality and lower-bias probability sample with a nonprobability sample. While any probability survey will work, be it a standalone telephone survey or address-based survey, we find that AmeriSpeak is the perfect candidate given its use of in-person recruitment, response rates superior to many other probability research designs, a price point already closer to nonprobability samples than other probabilistic research, and the ability to field quickly.

The real benefit to TrueNorth is in the sophisticated way the method combines the two samples. A nonprobability sample is not randomly selected. Rather, respondents are irregularly invited through a variety of means, driven primarily by convenience (in short, the survey provider has some "easy" means of finding people, such as purchasing a list from a company or through advertising on specific websites). The "types" of people in a nonprobability sample are unknown, and, just as concerning, the proportions of these types are unknown. Therefore, any method of weighting a nonprobability sample needs to

[1] Bilgen, Dutwin 2026
[2] Bilgen, Dutwin 2026
[3] Yang et al. 2018; Ganesh et al. 2017

typologize respondents into meaningful groups from which to weight and then know the proportions of people that belong in each group.
At its heart, this is what all weighting does. For example, nearly all samples are put into "types" by age group, sex, race/ethnicity, education, etc., and we can attain the correct proportions of each type from federal data sources. Raking or a similar typical weighting procedure will then create weights to ensure proper representation of each type of respondent. Unfortunately, multiple studies document in detail that weighting solely by demographics is necessary, but quite insufficient, for nonprobability samples and does not reduce the bias of such samples.[4]

Instead of weighting to just demographic parameters, TrueNorth does much more. The latest version of TrueNorth described in this paper – TrueNorth 3.0 – utilizes an ensemble tree-based non-parametric supervised learning algorithm to classify respondents into new types based on their survey responses. [5] TrueNorth leverages the fact that it has a companion probability sample that, properly weighted, is assumed to be generally unbiased. The TrueNorth 3.0 algorithm classifies the combined probability and nonprobability sample into types, beyond the usual basic demographic types, based on how the combined sample best clusters based on survey response data. It thus solves two problems for the nonprobability sample: It first creates types in that the tree-based analyses classify cases into distinct leaves (types), and second, the weighted, generally unbiased probability sample provides the estimated weighted proportion of each leaf in the overall tree.[6]

The Process of TrueNorth 3.0 Weighting
To develop TrueNorth 3.0 weights, the first step is to fit a weighted ensemble tree model to the combined probability and nonprobability sample. Second, based on the fitted tree models, estimate the probabilities of inclusion in the combined probability and nonprobability sample and compute the initial weights as the inverse of the estimated probabilities. Third, poststratification adjustments, including calibration to demographic benchmarks and weight trimming, are made to the initial weights to create the final weights. These three steps are illustrated in this diagram and then described in more detail:

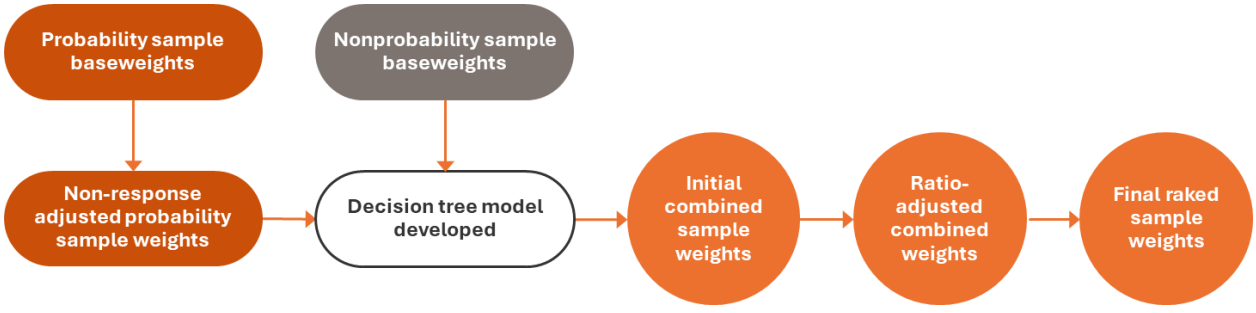

Figure 1 - TrueNorth Process

[4] Cornesse et al. 2020
[5] TrueNorth 1.0 calibrated the nonresponse sample with AmeriSpeak probability sample using small area model estimation. A White Paper outlining the methodology is available upon request. TrueNorth 2.0 used propensity score adjustment method, but it was not used by AmeriSpeak in production.
[6] Huang, Breidt 2024

Step 1: Starting Weights for Probability and Nonprobability Sample Members

We begin the TrueNorth process with nonresponse-adjusted sample weights for the probability sample members and a weight of 1 for all nonprobability sample members.

Step 2: Develop Weighted Ensemble Tree Model

A decision tree, a non-parametric supervised learning algorithm for classification, is next used to classify respondents in the combined sample into types based on their actual survey responses. To fit the weighted ensemble tree model, we use the starting weights from Step 1 for each member of the combined sample. The ensemble tree model is fitted with all observed survey response data and demographic covariates from the combined sample. The leaves in the final trees are assumed to be homogeneous with respect to the probabilities of inclusion in the nonprobability sample. Each combined sample member will be assigned to a single leaf in the tree model. The size of the leaves (i.e. number of sample members in each leaf) is determined to minimize a bias-variance score computed over all variables for the combined sample.

Step 3: Compute Initial Combined Sample Weights

In this step, we use the ensemble tree structure to estimate the overall probability that a unit is included in the combined sample, either through the probability sample or the nonprobability sample. This requires estimation of two quantities:

1) Probabilities of inclusion in the nonprobability sample, for both probability sample units and nonprobability sample units.
2) Probabilities of inclusion in the probability sample, only for the nonprobability sample units. (These are known for the probability sample units.)

For all combined sample units in each leaf, we estimate quantity 1, the probabilities of inclusion in the nonprobability sample among all sample units, as the ratio of the number of nonprobability sample units to the total weighted counts of the leaf. Note that the numerator is simply the number of nonprobability sample units and the denominator is the sum of the number of nonprobability sample units and the weighted total of probability sample units. Essentially, the estimated probability of inclusion in the nonprobability sample is the estimated population proportion of nonprobability units per leaf.

Because the leaves are expected to be homogeneous, we estimate quantity 2, the probability of inclusion in the probability sample among the nonprobability sample units, as the average design probability over all probability sample units. In other words, the nonprobability sample units in a leaf are assumed to have a probability of inclusion in the probability sample that is equal to the average inclusion probabilities among the probability sample units.

For all sample units, the inclusion probability in the combined sample is estimated as the probability of inclusion in the probability sample plus the probability of inclusion in the nonprobability sample given that they are not selected into the probability sample. The inverse of this value is the initial sample weight for each unit in the combined sample.

Step 4: Compute Ratio-adjusted Combined Weights

Next, we ratio-adjust the initial combined sample weights so that the sum of the weights over all units is the same as the sum of the nonresponse-adjusted weights for all probability sample units. This ratio adjustment does not change the initial weight for leaves that contain probability sample units only (there are typically few, if any, of these leaves). For leaves that contain nonprobability sample units only, all units retain their starting weight of 1 (there are typically none of these leaves in the ensemble tree structure). For leaves that have both probability and nonprobability units, the ratio adjustment resembles a poststratification adjustment that forces the total weight to match the sum of the nonresponse-adjusted weight for all probability sample units.

Step 5: Compute Final TrueNorth Raked Sample Weights

A final raking adjustment is applied to the ratio-adjusted combined weights from Step 4. The raking dimensions are specific to each study and typically come from national federal sources such as the Current Population Survey or the American Community Survey. The raked weights are the final TrueNorth weights for the combined sample.

Empirical Evaluation of Effect of TrueNorth

NORC has run hundreds of simulations and has fielded over 100 surveys to confirm the effectiveness of TrueNorth, a process which has evolved continuously since its inception in 2019. One major study fielded by NORC for the purpose of evaluating TrueNorth is our Proof-of-Concept survey, which fielded a large sample size (n=9,573) including major oversamples and low-incidence groups to assess the ability of TrueNorth to not just reduce bias in overall samples but in subgroups as well. In addition, a range of questions were asked across a variety of topics to assess how TrueNorth performs when tasked with a survey that ranges in topical variety.

Figure 2 demonstrates TrueNorth's effectiveness overall, as well as by demographic subgroup, by showing the reduction in mean absolute relative bias across three samples: nonprobability sample alone, combined probability and nonprobability sample weighted using TrueNorth 1.0, and combined probability and nonprobability sample weighted using TrueNorth 3.0. We calculate the relative biases across survey measures using the AmeriSpeak probability sample estimates as the benchmark for comparison.

We include the comparison results from TrueNorth 1.0 to document the degree to which our evolving methods have further improved bias reduction[7]. Overall, TrueNorth 1.0 was able to reduce bias by a little over one-third, but the current version TrueNorth 3.0 reduces bias by over one half compared to the nonprobability sample estimates.

[7] As TrueNorth 2.0 was not used by AmeriSpeak in production, comparisons to TrueNorth 2.0 are not shown in this paper.

As data is broken into demographic subgroups, bias goes up significantly, but the effectiveness of TrueNorth, relative to the bias in the nonprobability samples, remains generally consistent.

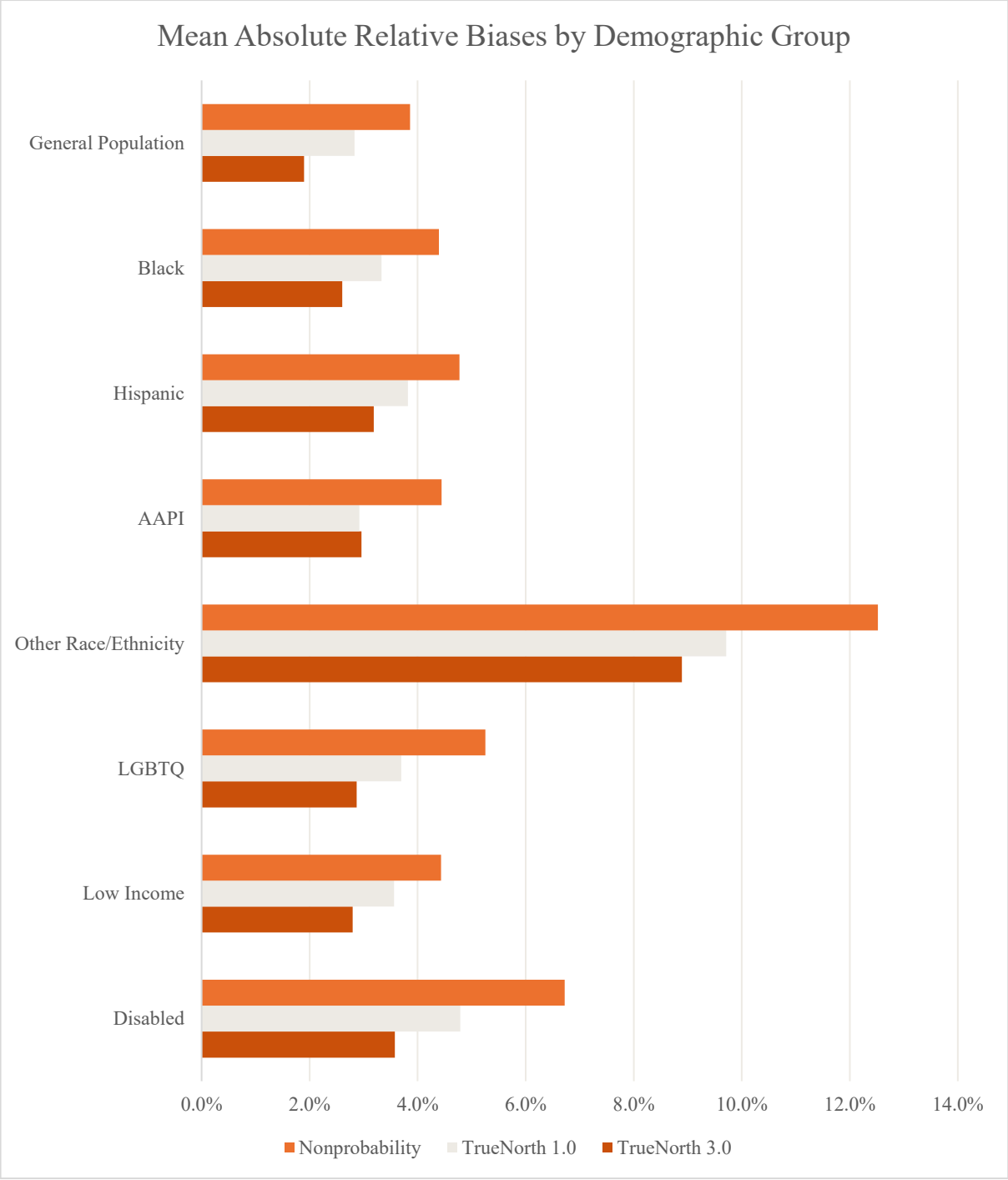

Figure 2 - Bias Reduction of TrueNorth 3.0 compared to TrueNorth 1.0 by Demographic Group

As we look at topical areas, there are similar results in which TrueNorth exhibits the ability to reduce bias by half or more. In some instances, such as with questions on mental health and health insurance cost, bias is reduced by almost two-thirds.

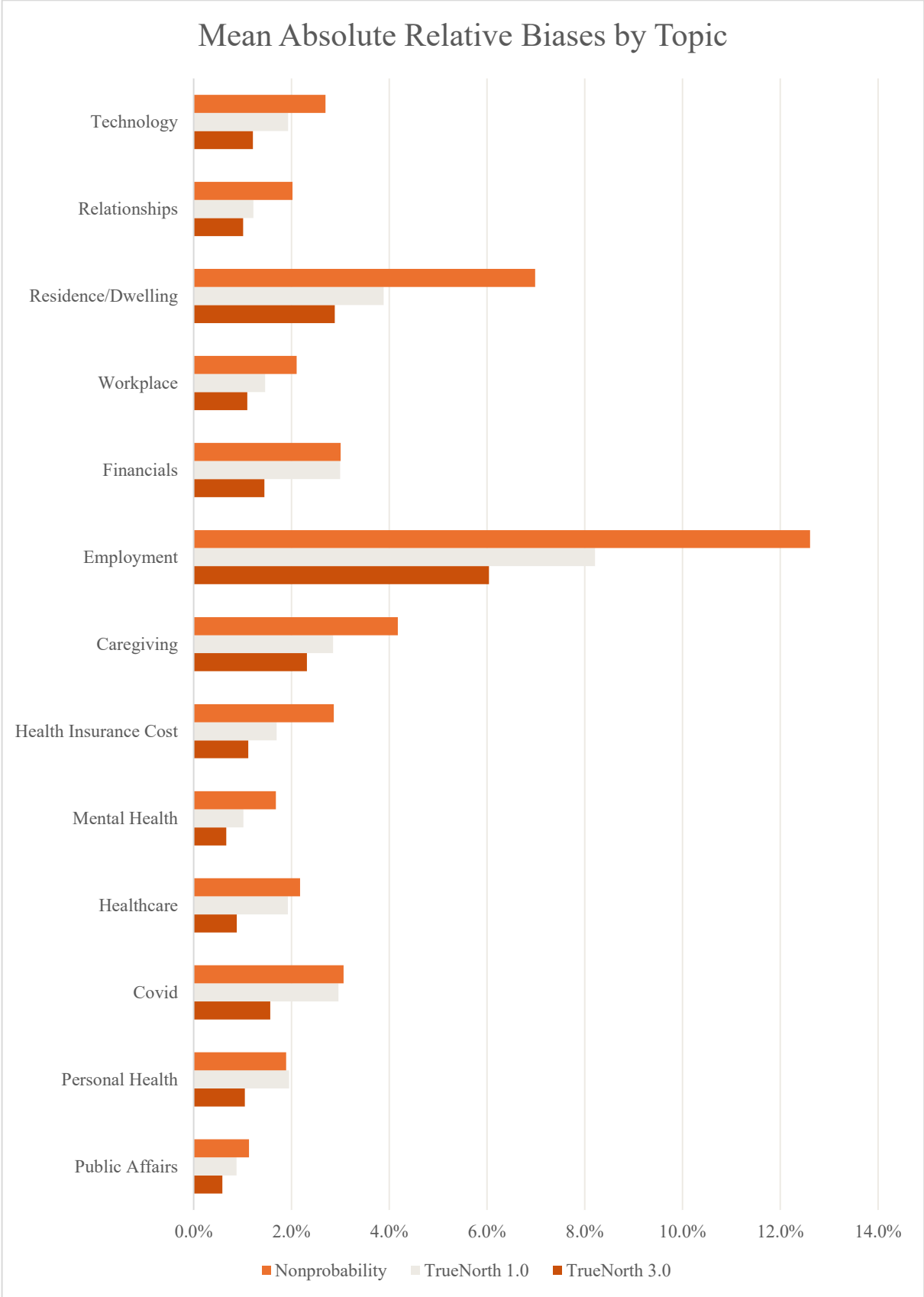

Figure 3 - Bias Reduction of TrueNorth 3.0 compared to TrueNorth 1.0 by Question Topic Area

Conclusion

As survey researchers confront declining response rates, rising costs, and growing reliance on nonprobability samples, methods that can restore data quality are no longer optional, they are essential. TrueNorth calibration represents a rigorously tested, empirically validated solution that moves beyond traditional demographic weighting by leveraging rich survey response data and a high-quality probability sample source to substantially reduce bias, both overall and within critical demographic subgroups. Results from extensive simulations and large-scale field studies show that TrueNorth 3.0 achieves meaningful improvements over earlier versions, consistently cutting bias by one-half or more across a wide range of topics and populations. By combining methodological rigor with practical scalability, TrueNorth enables researchers to confidently integrate nonprobability samples into modern research designs while maintaining the standards of accuracy and credibility that decision-makers require.

References

Bilgen, Ipek and David Dutwin (2026). Using Nonprobability Survey Samples Can Be a Dangerous Gamble. NORC at the University of Chicago. Retrieved from www.norc.org.

Cornesse, Connie, et al (2020). A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research. Journal of Survey Statistics and Methodology, Volume 8, Issue 1, February 2020, Pages 4–36, https://doi.org/10.1093/jssam/smz041.

Ganesh, N., V. Pineau, A. Chakraborty, and J. M. Dennis (2017). Combining probability and non-probability samples using small area estimation. In *Proceedings of the Section on Survey Research Methods*, Alexandria, VA, pp. 1657–1667. American Statistical Association.

Huang, C.-M. and F. J. Breidt (2024). A Tree-Based Dual-Frame Estimation Approach for Combining Probability and Non-probability Samples. In *Proceedings of the Section on Survey Research Methods*, Portland, OR, https://doi.org/10.5281/zenodo.13942558. American Statistical Association.

Yang, M., N. Ganesh, E. Mulrow, and V. Pineau (2018). Estimation methods for nonprobability samples with a companion probability sample. In *Proceedings of the Section on Survey Research Methods*, Alexandria, VA, pp. 1715–1723. American Statistical Association.