

**WHITE PAPER**

**NORC's TrueNorth Calibration tool for  
probability and nonprobability samples:  
New Version 2.0 even more effective**

**TRUE↑NORTH**  
Calibration Tool

 **NORC** at the  
University of  
Chicago

# TrueNorth: An Advanced Calibration Tool for Combining Probability and Nonprobability Samples

Updated August 2021

The survey research landscape seems to have two consistencies in its approximately 60-year history: every few decades it encounters new challenges, and in that same timespan, innovations have paved the way to solutions that in the long run, have perhaps just made the quality of the research even stronger. The challenge of the past decade has of course been declining participation. And while modern research<sup>1</sup> finds that the quality of data continues to be strong, what is undeniable is that costs have skyrocketed.

NORC provides a strong suite of solutions, featuring the probability-based [AmeriSpeak® Panel](#). AmeriSpeak is the only probability household panel in the U.S. to use in-person/face-to-face household recruitment. As a result, AmeriSpeak attains survey response rates greater than most typical modern standalone telephone surveys, yet at well under half the cost. NORC also fields address-based designs for research that requires even greater response rate requirements, and nonprobability samples for consumer research.

Yet NORC, like many across the research industry, understands the near unanimous research on nonprobability (also called opt-in internet panels or internet convenience panels) surveys, which finds, simply put, significantly more bias than any comparable probability research, and nearly four times the variance within estimates. This means, in short, a nonprobability survey will almost surely have estimates with little or no bias, but also estimates with very large biases, and no way to detect which estimates fall into either camp.<sup>2</sup> The risk in making decisions based on such a pattern of bias is not unlike playing Russian Roulette. But what is a researcher to do if their price-point for research makes even low-cost probabilistic research like the AmeriSpeak panel unattainable?

The NORC solution, [TrueNorth® Calibration](#), is to leverage the best of both worlds, the stability and historic accuracy of probabilistic research with low-cost nonprobability sample, combined with a solution unique within the research industry, parsimonious in its conceptualization but sophisticated in its statistical execution. TrueNorth Calibration reduces the bias of nonprobability samples at not only the topline level, but also deep within key demographic groups.<sup>3</sup> Moreover, the approach is tailored to the particular topic of each survey to reduce bias that might be specific to the given survey.

<sup>1</sup> Dutwin, D., and Buskirk, T. (2020). Telephone Sample Surveys: Dearly Beloved or Nearly Departed? Trends in Survey Errors in the Age of Declining Response Rates.” *Journal of Survey Statistics and Methodology*. 8, 2: 1-32.

<sup>2</sup> See table 1 in Carina Cornesse, Annelies G Blom, David Dutwin, Jon A Krosnick, Edith D De Leeuw, Stéphane Legleye, Josh Pasek, Darren Pennay, Benjamin Phillips, Joseph W Sakshaug, Bella Struminskaya, Alexander Wenz, 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>. The four times bias is a conclusion in David Dutwin, Trent D. Buskirk, Apples to Oranges or Gala versus Golden Delicious? Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples, *Public Opinion Quarterly*, Volume 81, Issue S1, 2017, Pages 213–239, <https://doi.org/10.1093/poq/nfw061>

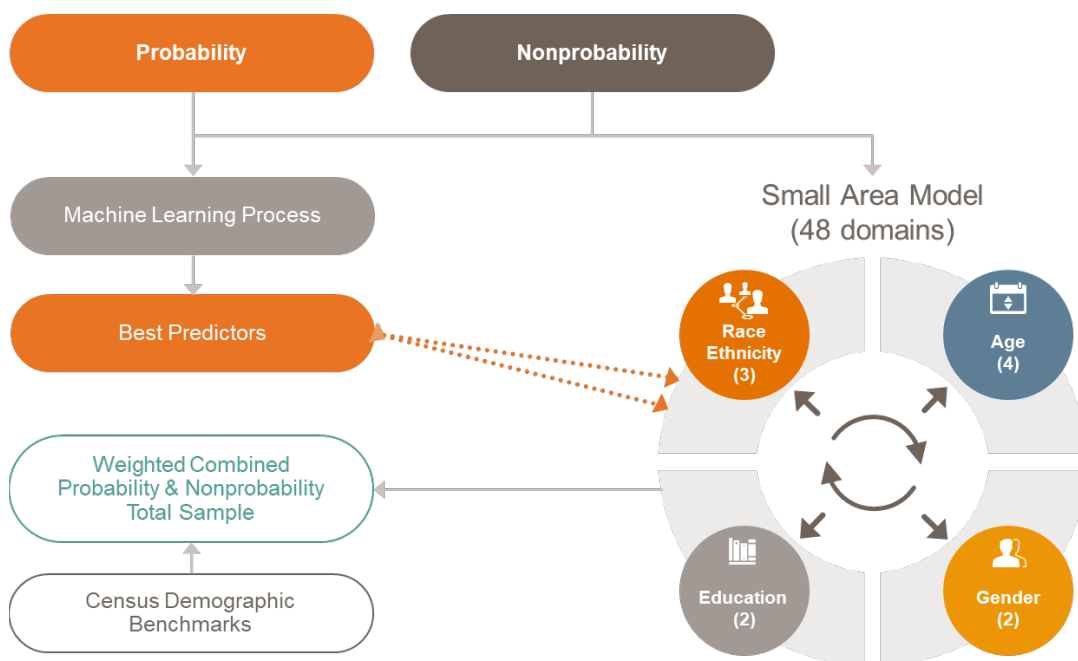
<sup>3</sup> Yang et al. 2018; Ganesh et al. 2017; Gupta et al 2019

The “chassis” for the advanced calibration is of course a high-quality probability survey. 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, as noted earlier, its use of in-person recruitment, response rates superior to many other probability research designs, and a price point already closer to nonprobability samples than other probabilistic research.

To implement TrueNorth Calibration, we combine the probability sample from the AmeriSpeak Panel with the nonprobability sample, and then calibrate it using small area estimation. The process is as follows.

1. We identify two-to-four key analytical variables from the survey to target for reducing bias. These are identified using a machine learning approach called random forest modeling, combined with a correlational assessment. The technique therefore both identifies which variables have the most bias between the two samples as well as the greatest promise to correct for bias across all survey estimates within a given survey.
2. We establish 20-40 domains in the data, where each domain is a specific, relevant subgroup. We typically use demographics as a starting point to create the domains, such as African-American males age 18 to 34 with a college degree. We create even deeper dimensionality by also defining domains with key questions that are specific to the particular study, such as whether respondents purchase a product or live in a particular market area.
3. Within each domain, we run a small area model (SAM) for the analytical variables identified in step 1. These models generate probability-based benchmarks for deep calibration on the nonprobability sample within each domain. This approach is unique to TrueNorth.
4. Finally, the combined data is calibrated to these SAM benchmarks as well as to standard Census demographic benchmarks.

The graphic below illustrates the process.



In multiple assessments of this technique, we find that bias is reduced significantly, beyond what is attained via simple demographic benchmark weighting and simple combinations of the two samples. The effectiveness of the procedure will depend in a number of characteristics of the data in the first place, such as finding variables that strongly show variance between the probability and nonprobability estimates and are correlated with many other variables in the data. But commonly, the bias after TrueNorth Calibration has been found to be insignificant from benchmarks in most metrics across a half-dozen test surveys, with the remaining estimates having their bias reduced by at least half. And given the small-domain approach, this bias reduction runs deep into subgroups and subsamples in the data, not just for overall point estimates.

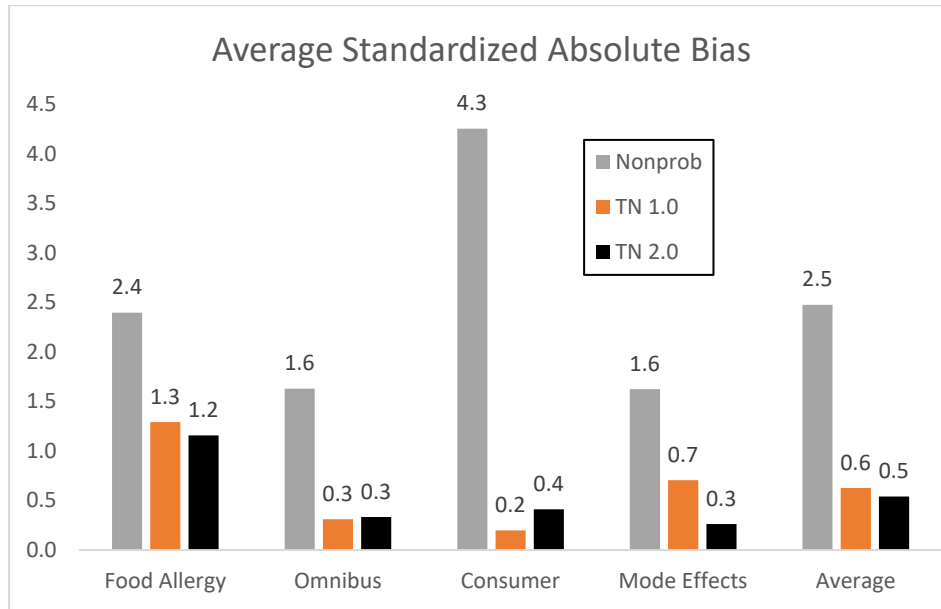
TrueNorth offers a range of applications, yes, allowing for cost-effective surveys for researchers without significant budgets, but as well to provide options for greater sample sizes overall or within low-incidence populations or limited geographies.

### **Recent Advances in the Method: TrueNorth 2.0**

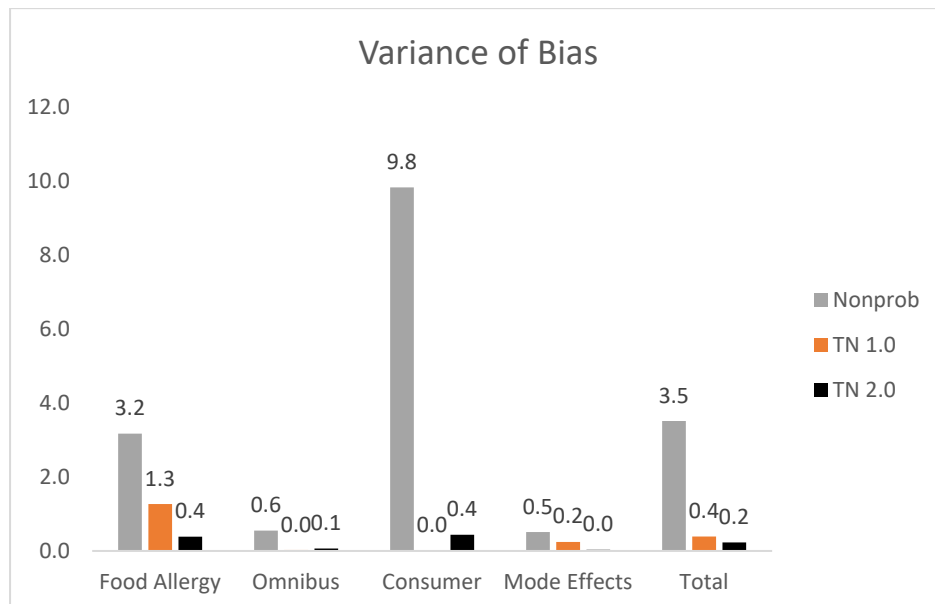
NORC continues to refine its TrueNorth approach to even further reduce bias in the nonprobability sample. Through extensive testing in both simulations and case study research, we have developed key refinements. In the original instantiation, the probability and nonprobability samples underwent simple calibration (raking) to population benchmarks prior to entering the machine learning process and small area modelling illustrated above. NORC now utilizes a more complex sample-matching procedure that “borrows” the probability weights on a case-by-case basis and applies them to the nonprobability cases.

As shown below, our revised approach has two major benefits. First, it further reduces bias compared to the original approach. Second, it reduces variance of that bias reduction. This is in fact the most significant improvement: In short, whereas the original TrueNorth approach reduced bias on average, it would reduce bias more for some survey questions than others. The new approach not only reduces bias, on average, more than the original method, but does so far more evenly across all survey questions.

In our four case studies, our revised procedure reduced bias across all studies by a factor of 4.5, compared to 4.0 for the original TrueNorth approach. Notably bias was reduced most significantly in the two case studies where TrueNorth had the least substantial initial reduction properties.



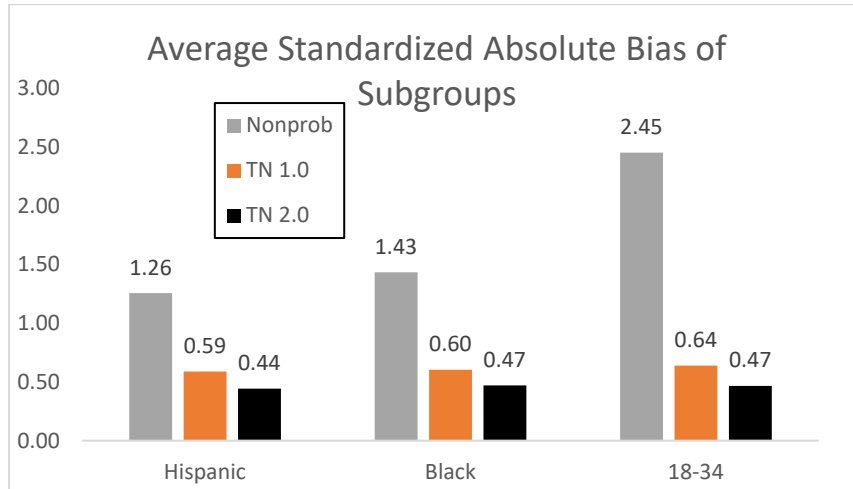
But notably, while this reduction in bias is a small but significant improvement from the standard TrueNorth method, the revised procedure makes substantial improvements in reducing the variance of the bias. While the original TrueNorth procedure effectively reduced the variance of bias significantly, in fact by a factor of 9, the new procedure nearly doubles this reduction, by over 1500%. In short, TrueNorth 2.0 gives much great confidence not only in reducing bias overall, but consistently across all point estimates generated in a given survey.



There is of course concern that any statistical procedure will increase variance in that the “harder” weighting and calibration techniques work to correct for bias, typically, the larger the variance of weights, which inflates margins of error and therefore reduces the ability to make significant claims from the data. However, we find that TrueNorth 2.0 does not significantly increase the design effects of the data: On average, design effects for a simple nonprobability weighting routine in our case studies was 1.3, compared to 1.5 for TrueNorth 1.0 and 1.6

for TrueNorth 2.0. These are very modest changes given the significant reduction in bias, on average and within individual estimates.

Finally, we find that TrueNorth 2.0 does particularly well in reducing bias in subpopulations. Across our four case studies, bias was reduced an additional 25% from the degree to which TrueNorth 1.0 reduced bias from the nonprobability sample (which was already near 200%).



NORC is constantly seeking to improve survey science. The latest evolution in TrueNorth is a result of deep research and development of our statistics group, striving to make combined probability-nonprobability samples nearly unbiased. Yes, TrueNorth is incomparably more accurate than simple nonprobability samples. When you need to get it right, reach out to NORC’s TrueNorth Calibration tool.