TRUENORTH
Calibration Tool

WHITE PAPER
November 2024

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

Presented by:

NORC at the University of Chicago

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

Updated November 2024

The survey research landscape seems to encounter new challenges every decade. One ongoing concern is the quality of opt-in samples from online convenience (nonprobability) panels. Long a stalwart for consumer research and election polling, such samples have seen significant degradation, leading to a lack of trust not just in election polling (Cohn 2024), but in consumer research and beyond. It is not uncommon today for over half of "respondents" in such surveys to be discarded as bots, bogus respondents, or simply for a general lack of data quality. Nonprobability samples can miss key segments of the population and substantially over-represent other segments. The problem cannot be solved by demographic weighting or quotas. How can trust be restored to this type of research?

NORC provides a strong suite of solutions, including 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, we know who respondents are, and AmeriSpeak attains survey response rates greater than all other modern research panels.

Sometimes, however, more interviews are needed than AmeriSpeak or any probability solution at a given cost can provide. Alternatively, a budget might not support a solely using a high quality approach like AmeriSpeak. **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.** TrueNorth Calibration reduces the bias of nonprobability samples at not only the topline level but also deep within key demographic groups.[1] Moreover, the approach is tailored to the particular topic of each survey to reduce bias.

How TrueNorth Works

The TrueNorth process solves a number of problems inherent to convenience samples, and in the process creates a sample that is quasi-probabilistic and far less biased than these nonprobability samples alone. Of course, the main way this is achieved is by blending much higher quality and lower bias probability sample with 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, as noted earlier, its use of in-person recruitment, response rates superior to many

[1] Yang et al. 2018; Ganesh et al. 2017; Gupta et al 2019

other probability research designs, a price point already closer to nonprobability samples than other probabilistic research, and the ability to field quickly.

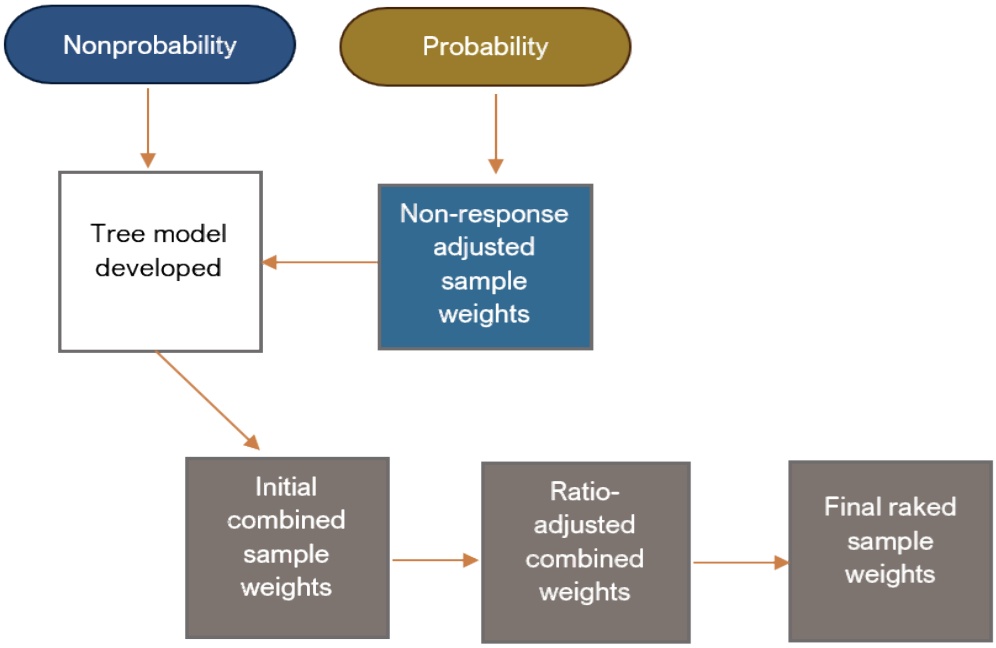
But the real difference is in the sophisticated way in which TrueNorth combines these samples. Nonprobability samples are not randomly selected: Rather, respondents are irregularly invited through a variety of means, driven primarily by convenience (in short, the survey provider has some "easy" means of finding people such as purchasing a list from a company or through advertising on specific websites). Thus, the "types" of people in convenience samples are unknown, and just as concerning, the proportions of these types is unknown. Therefore, any method of weighting (an industry-standard way of making samples more representative) convenience samples needs to be able to effectively typologize respondents into meaningful groups from which to ascertain some type of *post hoc* probability of selection with which to weight.

Of course, at its heart, this is what all weighting does. For example, nearly all samples are put into "types" by age group, gender, race/ethnicity, etc., and we can attain the correct proportions of each type via U.S. Census data. Raking or some other typical procedure will then create weights to ensure proper representation of each type of respondent: if a sample only 40% male, once weighted it will in effect be a proper 48% (the current percent of males in the U.S.). Unfortunately, multiple studies document in detail that weighting solely by demographics is a necessary, but quite insufficient, method to reduce bias in nonprobability samples (though highly effective in probability samples). So while TrueNorth, like most nonprobability weighting schemes, does weight to these important demographic parameters, more needs to be done: Again, new groupings of people need to be defined and the proportions of each type need to be set.

TrueNorth does this by using a tree-based non-parametric supervised learning algorithm to classify respondents into types based on their actual survey responses. It in effect produces quasi-probabilities of selection for a sample that *a priori* has none. TrueNorth first leverages the fact that it has a companion probability sample that, properly weighted, is assumed to be generally unbiased. The TrueNorth algorithm classifies sample into types/response probabilities based on how respondents best cluster by their responses to survey data. It thus solves both problems for the convenience sample: It first creates respondent types that the tree-based algorithm classifies into distinct leaves (types), and second, leverages the weighted probability sample to provide the estimated weighted proportion of each leaf in the overall tree.

Notably, it is often typical that some leaves end up without any nonprobability sample cases. This in effect represents the fact that nonprobability sample does not actually cover all types of people (for example, those that do not have Internet access, as well as people who could not be reached because they do not visit the websites in which the survey was advertised, or do not belong to the lists used by the nonprobability provider). For such leaves, the final weights of the cases are unchanged. For leaves with both probability and nonprobability cases, a ratio adjustment that resembles a poststratification adjustment forces the total weight in the leaf to match the total design weight across probability sample units in that leaf.

The graphic below illustrates the process.

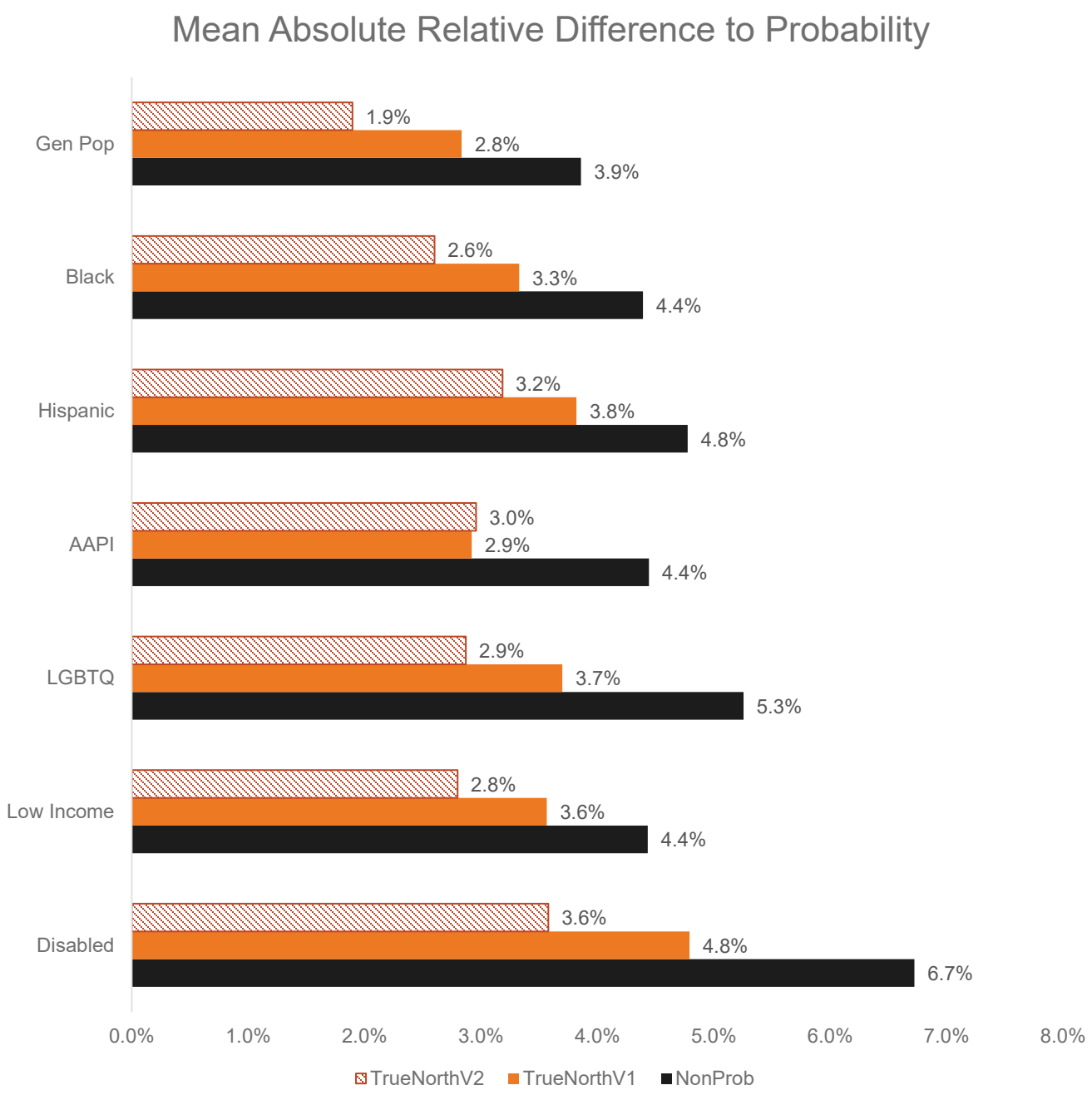


Example Application

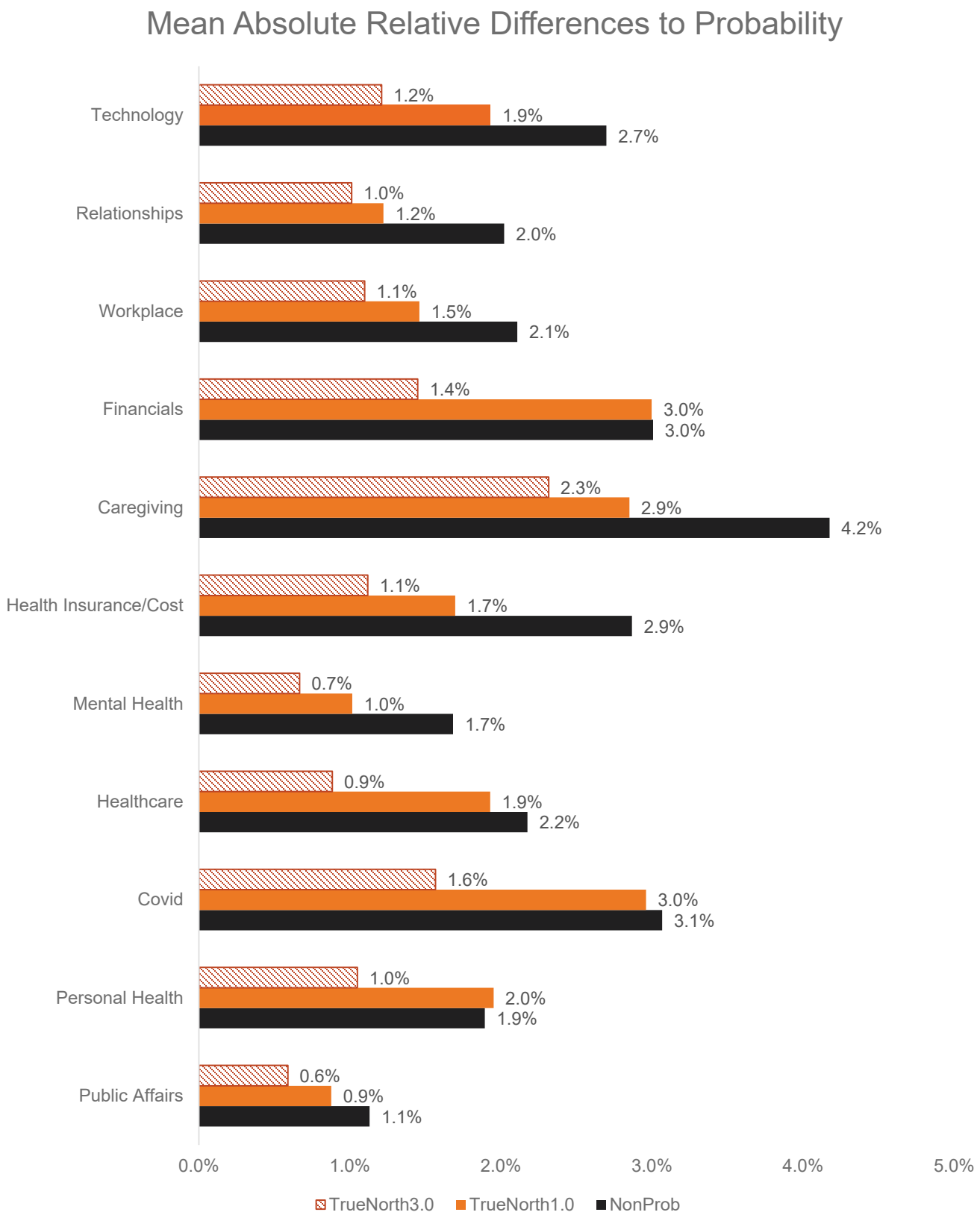
NORC has run hundreds of simulations and has fielded over 100 surveys to confirm the effectiveness of TrueNorth, a process which has evolved continuously since its inception in 2019. One major study fielded by NORC for the purpose of evaluating TrueNorth is our Proof of Concept survey, which fielded 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.

Table 1 below documents TrueNorth effectiveness overall, as well as by subgroup. Notably, we include the comparison results from the first version of TrueNorth to document the degree to which our methods have further improved bias reduction. Overall, our first version of TrueNorth was able to reduce bias by a little over one-third, but the current version reduces bias over one half.

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



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


Mean Absolute Relative Differences to Probability
Technology
1.2%
1.9%
2.7%
Relationships
1.0%
1.2%
2.0%
Workplace
1.1%
1.5%
2.1%
Financials
1.4%
3.0%
3.0%
Caregiving
2.3%
2.9%
4.2%
Health Insurance/Cost
1.1%
1.7%
2.9%
Mental Health
0.7%
1.0%
1.7%
Healthcare
0.9%
1.9%
2.2%
Covid
1.6%
3.0%
3.1%
Personal Health
1.0%
2.0%
1.9%
Public Affairs
0.6%
0.9%
1.1%
0.0%
1.0%
2.0%
3.0%
4.0%
5.0%
TrueNorth3.0
TrueNorth1.0
NonProb