

But What's My Margin of Error?

ESTIMATING SAMPLING VARIANCE WHEN BLENDING
NONPROBABILITY AND PROBABILITY SAMPLES

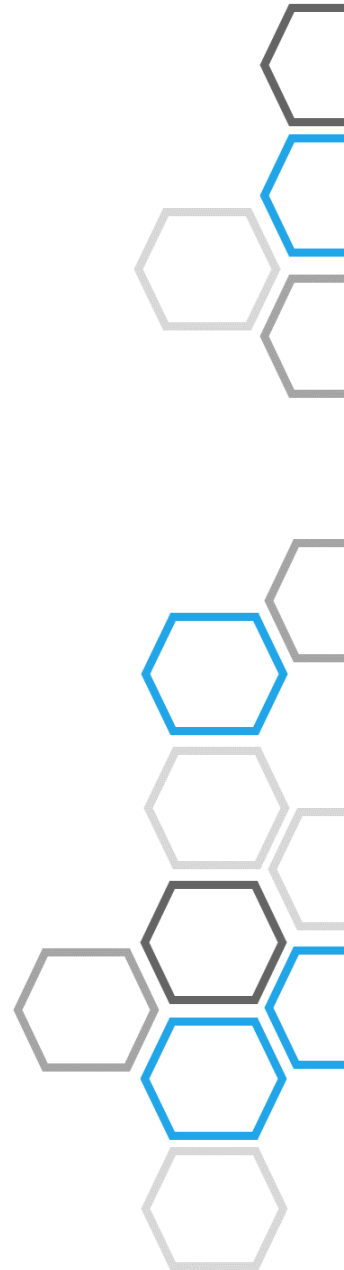
Mickey Jackson | Director of Data Science Innovation

ssrs
research. refined.

AAPOR
Annual Conference
2022
Chicago, IL
May 13, 2022

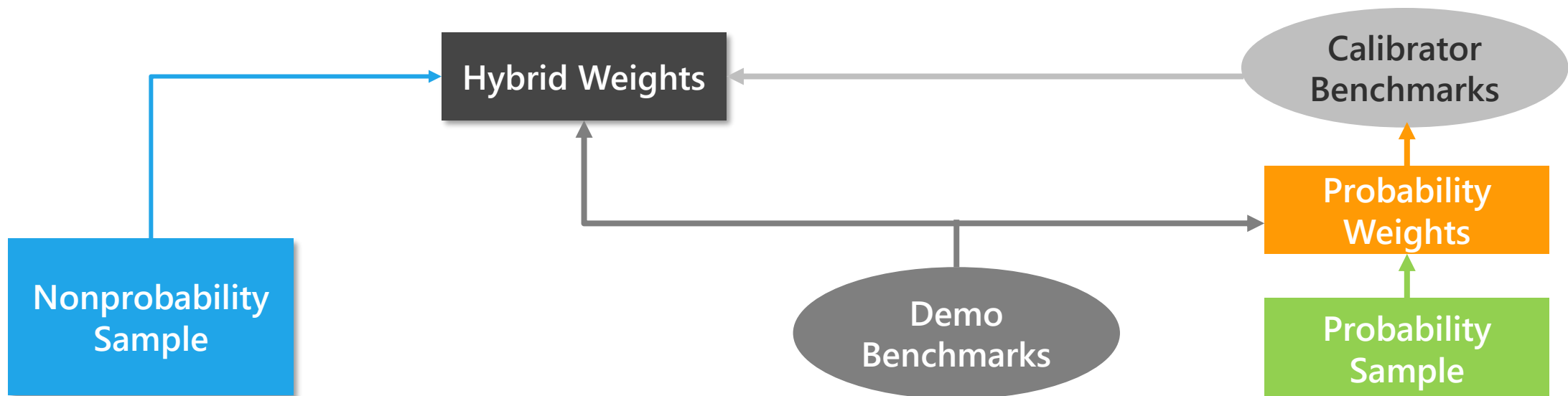
Why Hybrid Sampling?

- Nonprobability samples are cheap, but often biased
- Key challenges with nonprobability samples:
 - Exact selection mechanism (e.g., original recruitment source for a given complete) is unobservable to the user
 - Inclusion probabilities are unobserved and therefore must be modeled (implicitly or explicitly)
 - Empirically, selection is often related to non-demographic characteristics that lack external weighting benchmarks
- Enter hybrids: blending completes from probability and nonprobability sources



Weighting Hybrid Samples

Type of Weighting Variable	Used to Weight Probability Sample?	Used to Weight Hybrid Sample?	Source of Weighting Benchmarks
Demographics	X	X	<i>External (CPS, ACS, etc.)</i>
Non-demographic Calibrators		X	<i>Internal (Weighted Probability Sample)</i>

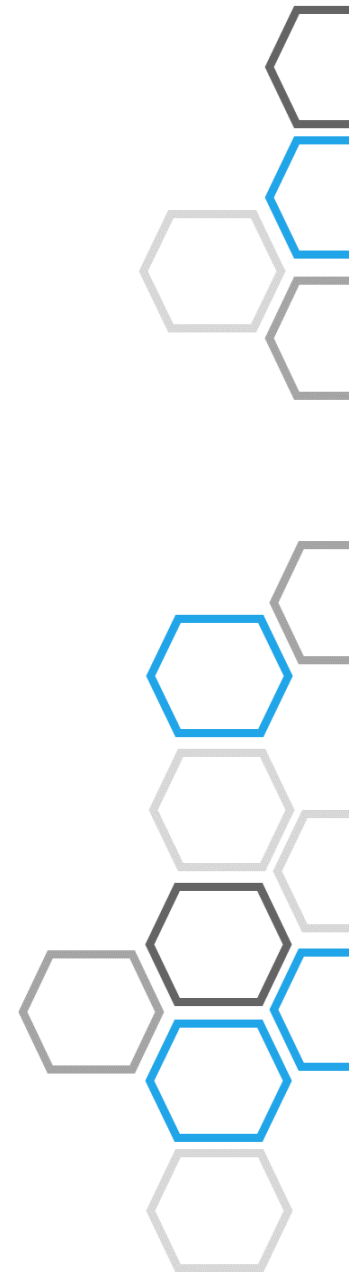


But What's My Margin of Error?

- Assessments of hybrid calibration have largely focused on bias (understandably)
- But we are usually also interested in estimates of sampling variance
 - Clients want to see a margin of error, standard error, etc.
 - Implicit motivation for hybrid sampling: obtain a larger sample (lower variance) than we could afford from probability sources alone
- Typical approach in practice: Apply familiar variance formulas from the probability world

Challenge: Internal Benchmarks

- Some weighting benchmarks come from within the same sample
 - Non-demographic calibrators use benchmarks obtained from the probability-based portion of the hybrid sample
 - By definition, the probability sample is smaller than the hybrid sample
- So, weighting benchmarks themselves are subject to non-negligible sampling variance
- How can we account for this in our variance estimates, and what are the consequences if we do not?



Simulation Setup

"POPULATION" FILE:
AP VoteCast 2020 Public-use File

OUTCOME OF INTEREST:
Trump Vote Share

**INTERNAL CALIBRATORS
(CORRELATION WITH
OUTCOME):**
Identifies as Republican (0.80)
Identifies as Conservative (0.67)



Simulation Procedure

- Assign assumed true inclusion probabilities to all “population” units
 - Probability sample: logit function of demographics only (**Missing At Random**)
 - Nonprobability sample: logit function of demographics, calibrators, and outcomes (**Missing Not at Random**)
- Select 1,000 replicate samples of each of the following:
 - Probability (n = 250, 1000, and 2000)
 - Nonprobability (n = 1000)
- Weight each replicate hybrid sample to:
 - Demographic benchmarks from synthetic population file (assume fixed)
 - Party and ideology benchmarks estimated from probability portion of sample (vary across samples)

Variance Estimation Methods – Non-Resampling

UWE APPROXIMATION

Simple random sample variance multiplied by Kish unequal weighting effect (UWE):

$$1 + CV(\text{weight})^2$$

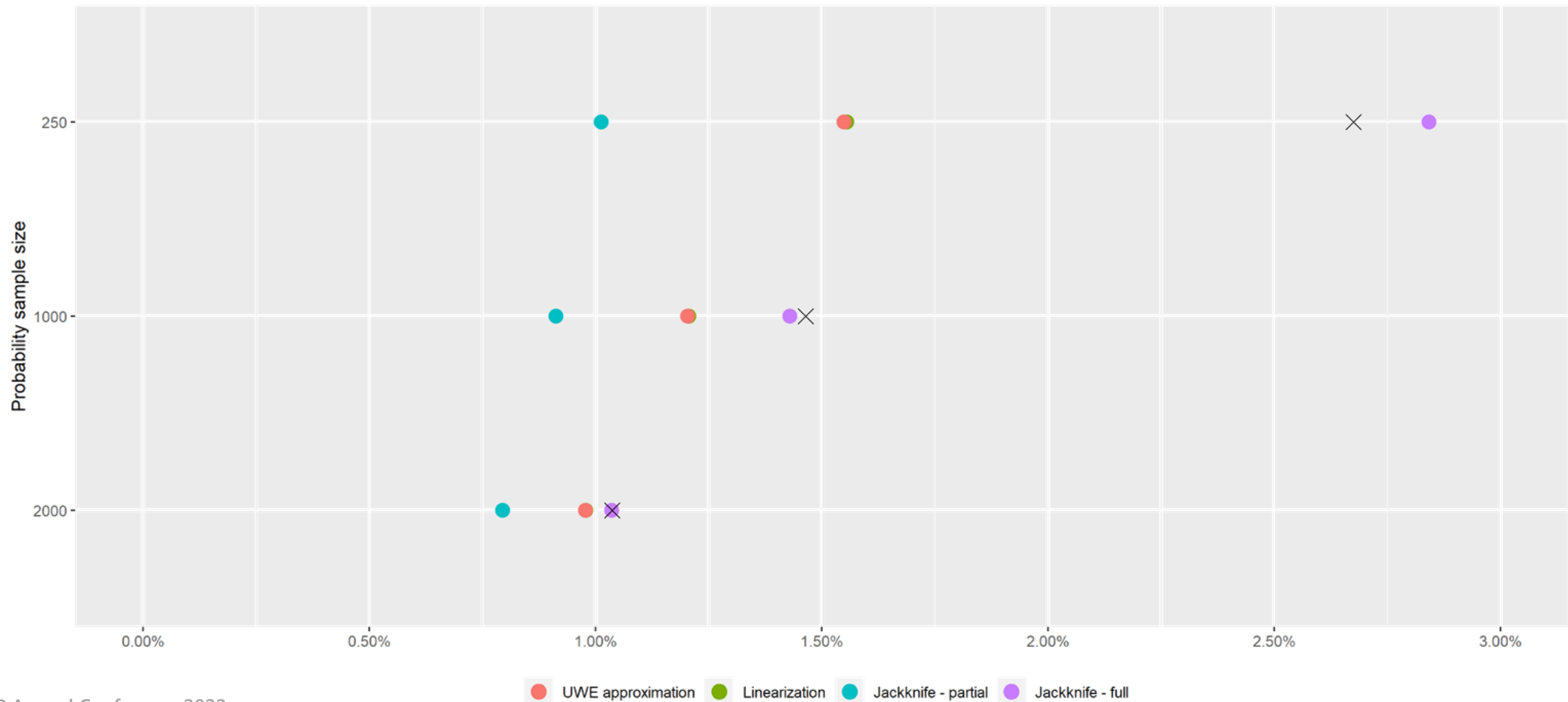
LINEARIZATION

Taylor Series linearization (default in most statistical packages)

Variance Estimation Methods – Resampling

- Two variants of delete-a-group jackknife—create and re-calibrate 50 random subsamples from each hybrid sample
- Jackknife – partial: re-calibrate all subsamples to same internal calibration benchmarks estimated from full probability sample (i.e., treat internal benchmarks as fixed)
- Jackknife – full: re-estimate internal calibration benchmarks from probability cases within each jackknife subsample (to account for variability in internal benchmarks)

Estimated vs. True Standard Errors, Trump Vote, Constant Nonprobability Mechanism



Conclusions

- Full jackknife (or other replication method), including re-calculation of internal benchmarks, is needed to obtain adequate variance estimates for internally calibrated hybrid samples
- Partial jackknife (assuming fixed internal calibration benchmarks across jackknife subsamples) consistently performs the worst
- If not using resampling methods: consider reporting variance based on both hybrid and probability-based sample sizes (to provide a conservative upper bound)
- UWE approximation is not a conservative method with internally calibrated hybrid samples

Other Results (from Paper)

- Simpler approximations (including Kish UWE) are adequate when calibration variables are weakly correlated with outcomes (i.e., when calibration is ineffective)
- Full jackknife is somewhat robust to “hidden variability” in nonprobability selection mechanism (but only with strong calibration variables)

Thank You, AAPOR

MICKEY JACKSON | MJACKSON@SSRS.COM

DIRECTOR OF DATA SCIENCE INNOVATION, SSRS



AAPOR
Annual Conference
2022
Chicago, IL
May 13, 2022