Evaluating Pre-Election Polling Estimates Using a New Measure of Non-Ignorable Selection Bias

Brady T. West, Ph.D.

Survey Research Center

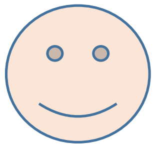
Institute for Social Research

University of Michigan-Ann Arbor

bwest@umich.edu

Acknowledgements

- This work was supported by an R21 grant from NIH (PI: West; NIH Grant No. 1R21HD090366-01A1)
- Thank you to my collaborators in all things SMUBBY: Rebecca Andridge, Phil Boonstra, Rod Little, and Fernanda Alvarado-Leiton
- Thank you to Patrick Sturgis for providing us with access to polling data from the 2015 General Election on Great Britain
- Thank you to Gary Langer and PARC for making national polling data publicly available here in the U.S.
- Thank you to the SRC for encouraging me to take a sabbatical!



Let’s Start with Means of Continuous Variables

- Suppose that we have data on a non-probability sample, including a continuous variable of interest Y and covariates Z
- **Aggregate population information**, via administrative records or some other source (e.g., a large probability sample producing small standard errors), is also available for the covariates Z
- We wish to develop the **best predictor** of Y from Z; for example, this could be the linear predictor of Y from a regression of Y on selected Z
- We call this “best” predictor of Y an *auxiliary proxy* for Y, and denote the auxiliary proxy by X (where X is scaled to have the same variance as Y); $\bar{X}$ is the mean of X for the population
- Assume for sake of generality that other covariates in Z (denoted by U) are orthogonal to X

Approach for Means, cont'd

- Our first proposed index of non-ignorable selection bias is based on maximum likelihood estimates of the parameters for a **normal** pattern-mixture model:

$$(X,Y \mid S=j) \sim N_2\left((\mu_X^{(j)}, \mu_Y^{(j)}), \begin{pmatrix} \sigma_{XX}^{(j)} & \sigma_{XY}^{(j)} \\ \sigma_{XY}^{(j)} & \sigma_{YY}^{(j)} \end{pmatrix} \right)$$

$$\Pr(S=1 \mid X,Y,U) = g\left(U,V\right), \text{ where } V=(1-\phi)X^* + \phi Y,\ U \perp X$$

- Note that the probability of inclusion in the non-probability sample (S = 1) is allowed to depend on both X* (rescaled X) and Y through ϕ; g() arbitrary
- If ϕ = 0, then selection is ignorable, depending on X* (and U) only
- If ϕ = 1, then selection is non-ignorable, depending on Y (and U) only
- There is no information in the data about ϕ, which can be varied in a **sensitivity analysis**

Approach for Continuous Variables, cont'd

- Andridge and Little (2011) show that the ML estimate of the mean of Y, given ϕ, is the following (note that rescaling of X is incorporated):

$$\hat{\mu}_Y(\phi) = \bar{y}^{(1)} + \frac{\phi + (1-\phi) r_{XY}^{(1)}}{\phi r_{XY} + (1-\phi)} \sqrt{\frac{s_{YY}^{(1)}}{s_{XX}^{(1)}}} (\bar{X} - \bar{x}^{(1)})$$

- Note that $r_{XY}^{(1)}$ is the correlation of X and Y in the *non-probability sample*
- Given this result, we propose a **measure of unadjusted bias (MUB)**, which can be rescaled by the observed standard deviation of Y to form a simpler **standardized measure of unadjusted bias (SMUB):**

$$\text{MUB}(\phi) = \bar{y}^{(1)} - \hat{\mu}_Y(\phi) = \frac{\phi + (1-\phi) r_{XY}^{(1)}}{\phi r_{XY}^{(1)} + (1-\phi)} \sqrt{\frac{s_{YY}^{(1)}}{s_{XX}^{(1)}}} (\bar{x}^{(1)} - \bar{X})$$

$$\text{SMUB}(\phi) = \frac{\phi + (1-\phi) r_{XY}^{(1)}}{\phi r_{XY}^{(1)} + 1 - \phi} \frac{(\bar{x}^{(1)} - \bar{X})}{\sqrt{s_{XX}^{(1)}}}$$

Using the Proposed Index

- Note that the proposed index is **simple**: it only depends on ϕ; means, standard deviations, and correlations from the observed non-probability sample; and the population mean for X
- We (Little et al. 2020, JSSAM) propose an intermediate choice of ϕ = 0.5 for computing SMUB, along with an "interval" for the selection bias based on the extreme cases of ϕ:

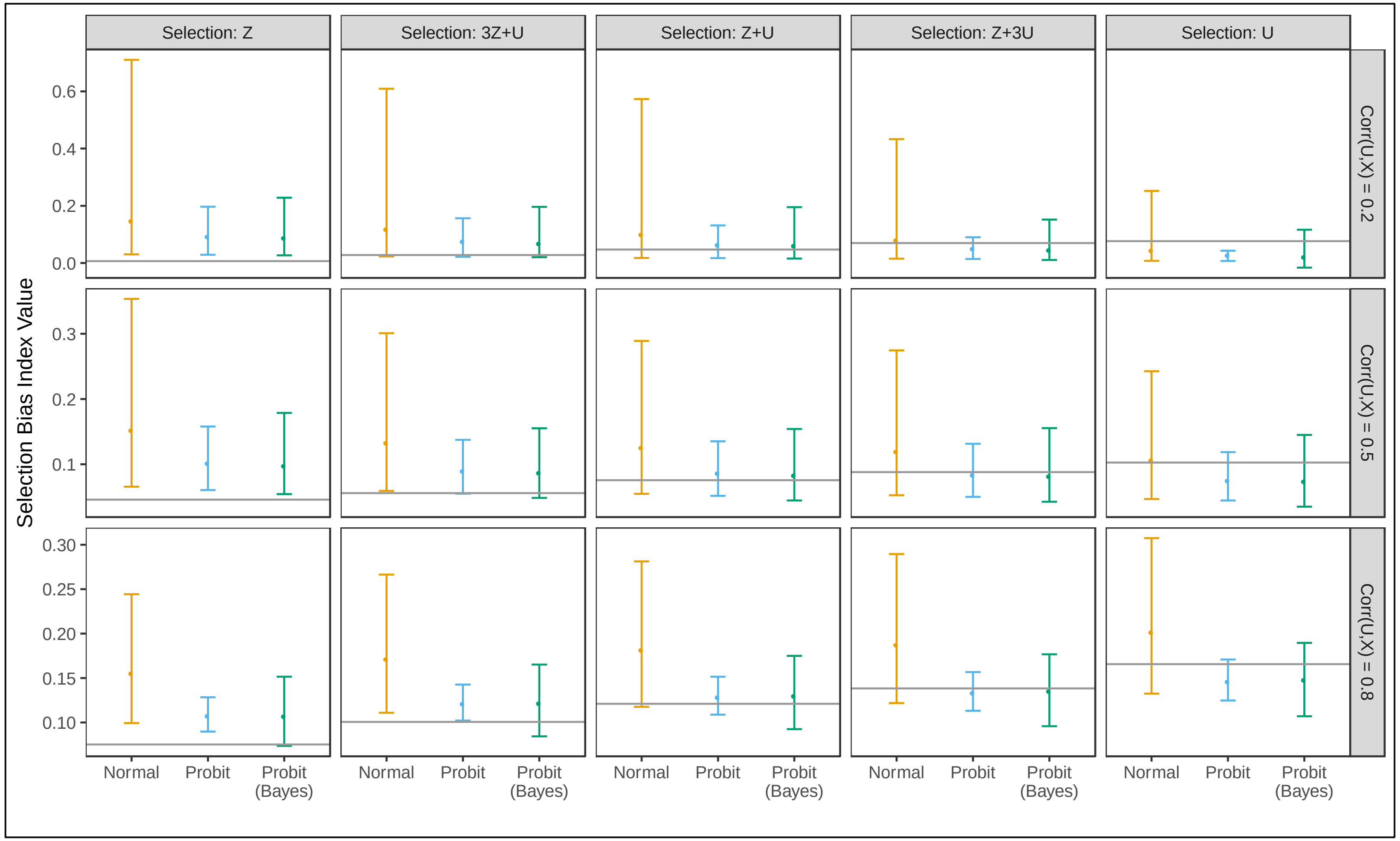
$$\boxed{\mathrm{SMUB}(0.5) = \frac{(\bar{x}^{(1)} - \bar{X})}{\sqrt{s_{XX}^{(1)}}}}$$

$$\mathrm{SMUB}(0) = r_{XY}^{(1)} \frac{(\bar{x}^{(1)} - \bar{X})}{\sqrt{s_{XX}^{(1)}}} \quad \text{and} \quad \mathrm{SMUB}(1) = \frac{1}{r_{XY}^{(1)}} \frac{(\bar{x}^{(1)} - \bar{X})}{\sqrt{s_{XX}^{(1)}}}$$

Now: What About Binary Variables?

- Per Andridge and Little (2009, 2018), suppose that a binary variable Y arises from a latent variable U that follows a normal distribution; we form X from a probit regression of Y on Z, and use the **biserial correlation** of X and Y
- Then, following a similar approach based on the pattern mixture model for U and X, we can form indices of selection bias based on the observed respondent proportion and the ML estimate of the mean of Y
- We can then define MUBP (the **measure of unadjusted bias for a proportion**; no need for standardization; Andridge et al. 2019, JRSS-C), along with MUBP(0), MUBP(0.5), and MUBP(1) indices for forming intervals
- We can also apply a **fully Bayesian approach** for forming credible intervals for the MUBP (given sufficient statistics on Z for the non-selected cases)

Simulation Results: Binary Case



- The gray horizontal line represents actual bias
- Note that the proposed intervals for the bias based on SMUB (orange, the normal model) are **substantially wider**
- The intervals based on the probit model (MUBP) are much narrower, and tend to cover the bias equally well (especially when using the Bayesian approach)

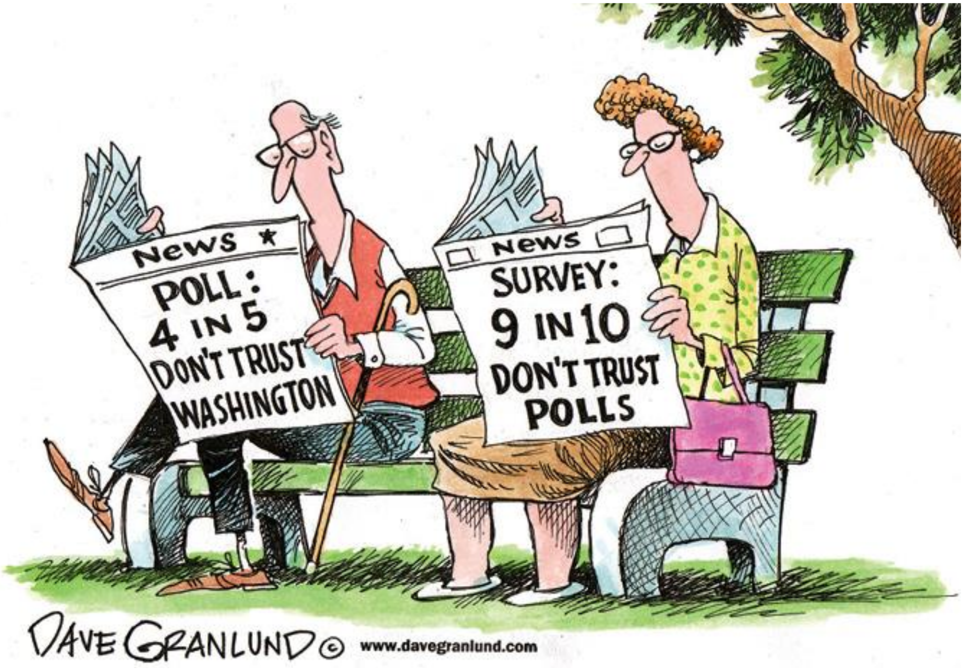
An Example: NSFG Smartphone Users

- We treat data from 16 quarters (2012-2016) of the National Survey of Family Growth (NSFG) as a hypothetical population
- We then consider smartphone users in this “population” as our hypothetical non-probability sample, simulating the selection process
- Our Y variables were variables of interest to NSFG data users
  - We considered both continuous and binary Y variables, to assess how robust the SMUB measures were to assumptions about normality
- Our Z variables included those where pop. aggregates may be available:
  - age, race/ethnicity, marital status, education, household income, region of the U.S. (based on definitions from the U.S. Census Bureau), current employment status, and presence of children under the age of 16 in the household
- We regressed Y on Z for smartphone users to form our linear predictor X

Applying the Binary Approach

- When computing the MUBP indices and their corresponding intervals for the **16 proportions** based on the binary variables in this motivating illustration, the following results emerged:
  - The proposed intervals were **significantly less wide** than the SMUB intervals (regardless of the correlation), reflecting the sensitivity of the MUBP index (derived from the probit model) to the discrete nature of the binary variables
  - 10 of the 16 estimated bias values were covered by the proposed intervals, representing an improvement over the SMUB approach (only 8 out of 16)

So About Polling…



"Should I put **genuinely** undecided or **damned if I'm telling you** undecided?"

Today's Talk: Assessing Non-Ignorable Selection Bias in Pre-Election Polling Estimates

- Polling "misses" at the national and state level receive a great deal of scrutiny, and a great deal of work has been conducted trying to understand the causes of these errors (Kennedy et al. 2018; Clinton et al. 2020)
- **Partisan nonresponse bias** is one possible explanation (Clinton et al. 2022)
- We sought to apply the MUBP measure to pre-election polling data from 18 different polls in the U.S. and Great Britain (**election outcomes known!**)
- Selected data from U.S. pre-election polls conducted in 2020 are now publicly available, via PARC and Roper; we focused on nine polls
- Patrick Sturgis provided us with selected data from nine additional polls conducted in Great Britain for the 2015 General Election

Data Sources

- **Survey Data:**
  - Seven polls in swing states, conducted right before the 2020 U.S. Presidential Election by ABC and the Washington Post
    - All telephone (dual-frame RDD sampling; n ranges from 777 to 1,043; RR4 = 4.5 - 6.5%)
  - Two polls in Arizona and Alaska, conducted right before the 2020 Presidential Election by Siena College and the New York Times
    - Both telephone (dual-frame RDD; n = 653 and 423; RR4 = 4.7 - 9.1%)
  - Nine polls conducted prior to the 2015 General Election in Great Britain
    - Six opt-in web panels
    - One mix of opt-in web panel and dual-frame RDD sample
    - Two dual-frame RDD samples
    - Northern Ireland excluded
    - Sturgis et al. (2016) conducted an extensive commissioned study of these data

Data Sources, cont'd

- **Survey Variables (U.S. Polls):**
  - An indicator of intention to vote for Trump (DV)
  - An indicator of being likely to vote in the 2020 Presidential Election
  - Covariates: male, age, education, race/ethnicity, ideology, party ID
  - Ideology not available in the Siena/NYT data sets
  - Siena/NYT weights incorporates predicted likelihood of voting, and were calibrated to ACS and CPS distributions for age, region, gender, and education; ABC/WP weights were simply calibrated in a similar fashion
- **Survey Variables (Great Britain Polls):**
  - An indicator of intention to vote for the conservative party candidate (DV)
  - Covariates: male, age, SGO region, party choice in 2010 election
  - Weights incorporated predicted likelihood of voting

Data Sources, cont'd

- Finding a good source of population data on likely voters, <u>with predictors of voting for a given candidate that are also available in the poll data</u>, **was the hardest part of this research!**
- For the U.S. polls, we considered (all publicly available):

1. The November 2020 CPS Voter Supplement (…no measures of ideology / party preference)
2. The 2020 ANES Pre-Election Survey (…smallest samples from each state, and Alaska sample not sufficient)
3. The AP/NORC VoteCast 2020 Data (…most data from each state, but not entirely probability-based)

Data Sources, cont’d

- For the Great Britain polls, we used the publicly-available 2015 British Election Study data, per recommendations of Sturgis et al. (2016)
- For the true outcomes in each U.S. state, we used publicly-available data from the MIT Election Data and Science Lab for 2020
- For the true outcome in the U.K., we used information in Sturgis et al. (2016); 37.7% ultimately voted for the conservative party candidate

Analytic Approach

- Compute the **unweighted** estimates of the proportion of likely voters that would vote for each candidate (in addition to SRS SEs and CIs)
- Compute the **weighted** estimates of the proportion of likely voters that would vote for each candidate (standard design-based approach)
- Compute the MUBP measure, **the adjusted proportion based on the MUBP measure**, and an adjusted 95% credible interval, in two ways:

1. Using demographics only in the probit model (is this better than standard weighting, since it allows for non-ignorable selection?)
2. Using all covariates in the probit model (improvements?)

Analytic Approach, cont’d

Quality Measures for the Competing Estimates:

1. Visual assessment of bias, including 95% CIs / credible intervals
2. The proportion of bias removed (PBR) by an adjustment:

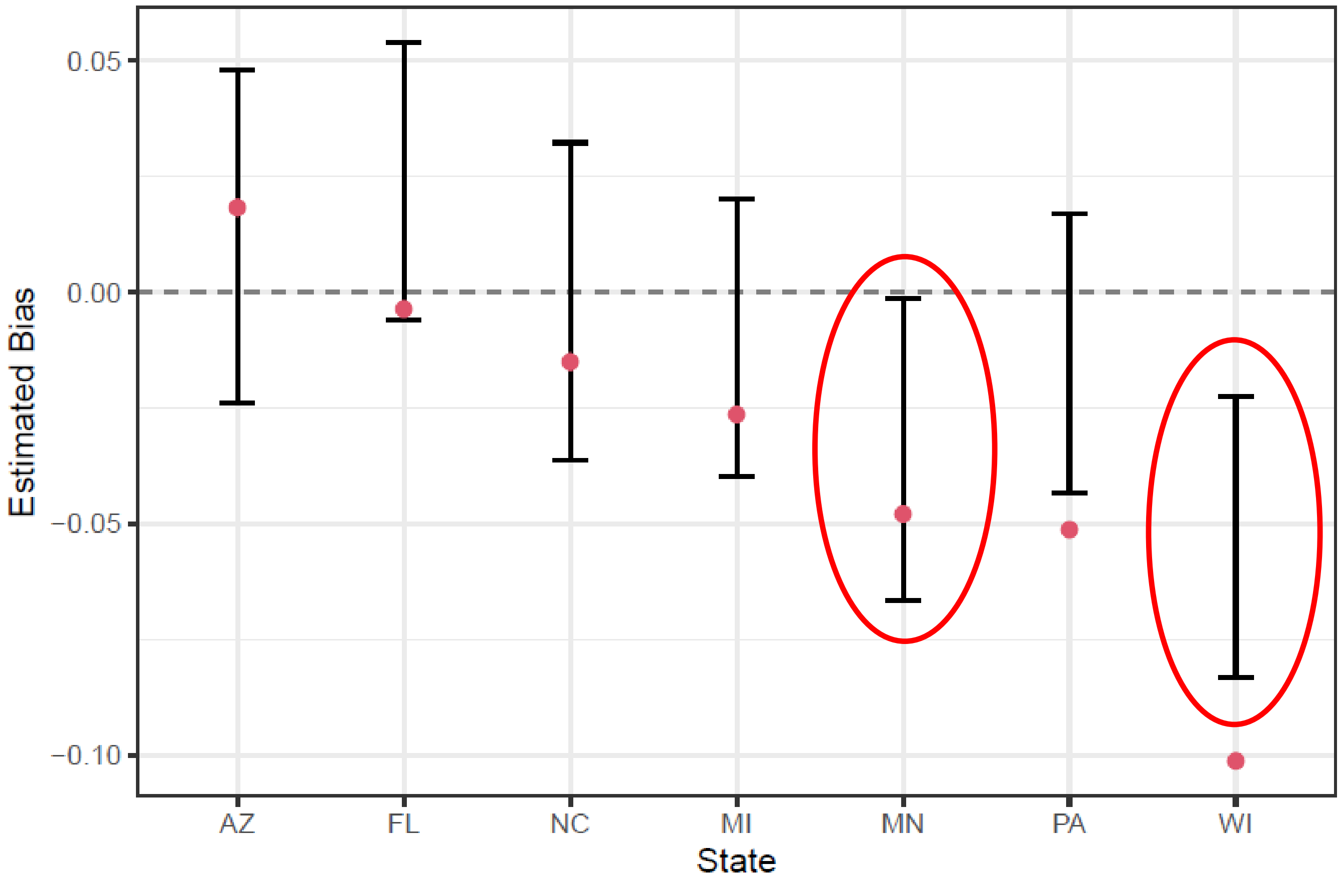
   PBR = (adj. est. - unweighted est.) / (true proportion - unweighted est.)

   - The PBR indicates whether an adjustment exacerbates the bias (% bias removed < 0), removes some or all of the bias (% bias removed between 0% and 100%), or overshoots the bias removal (% bias removed > 100%).
   - Unweighted estimates that seem unbiased can severely inflate the PBR
3. A pseudo-RMSE for each estimate, based on the estimated bias of the estimate *and* the estimated SE of the estimate

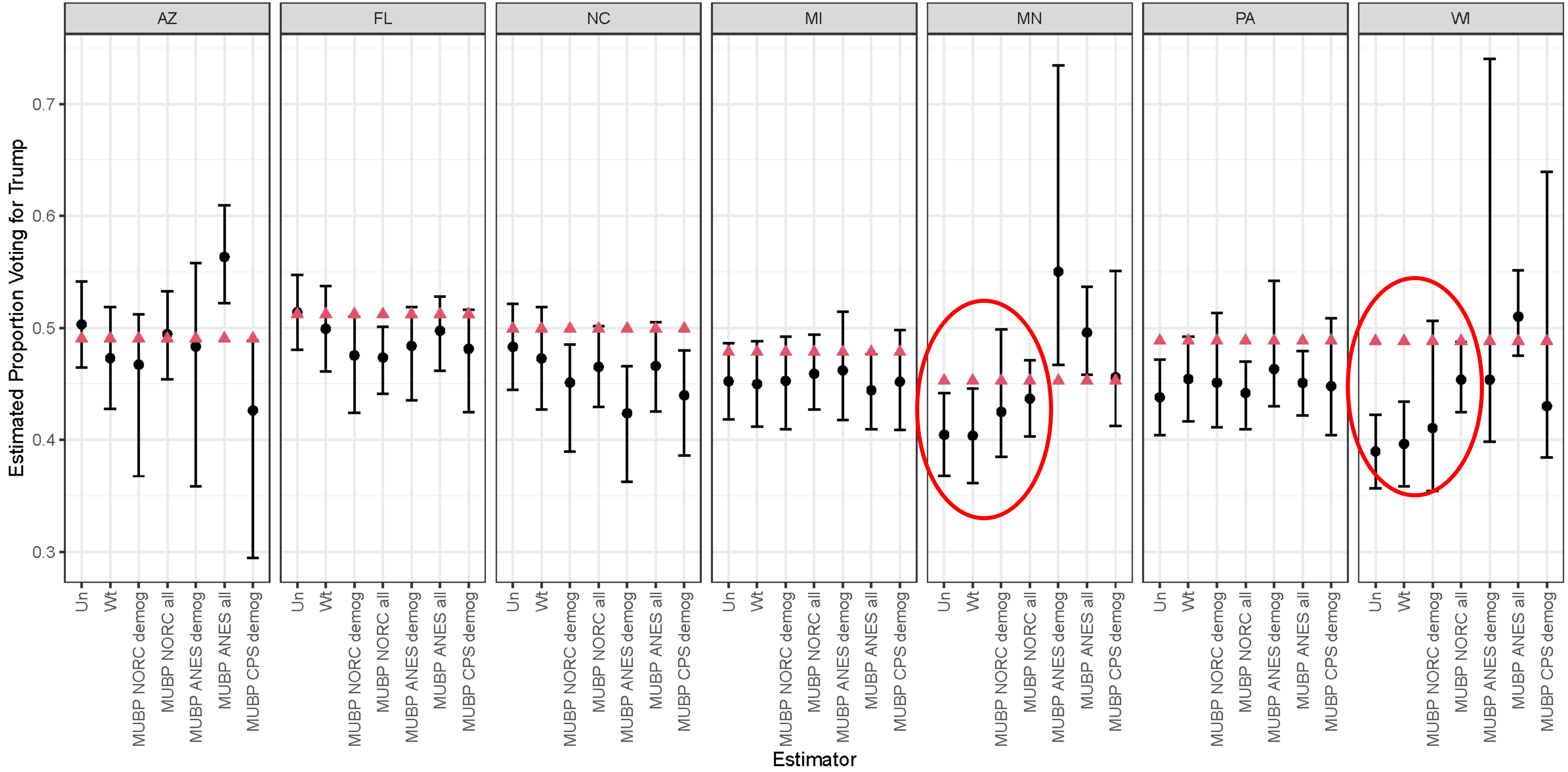
We Aren’t Making This Up…

- You can walk through / replicate all of our analyses in Rstudio: https://github.com/bradytwest/IndicesOfNISB

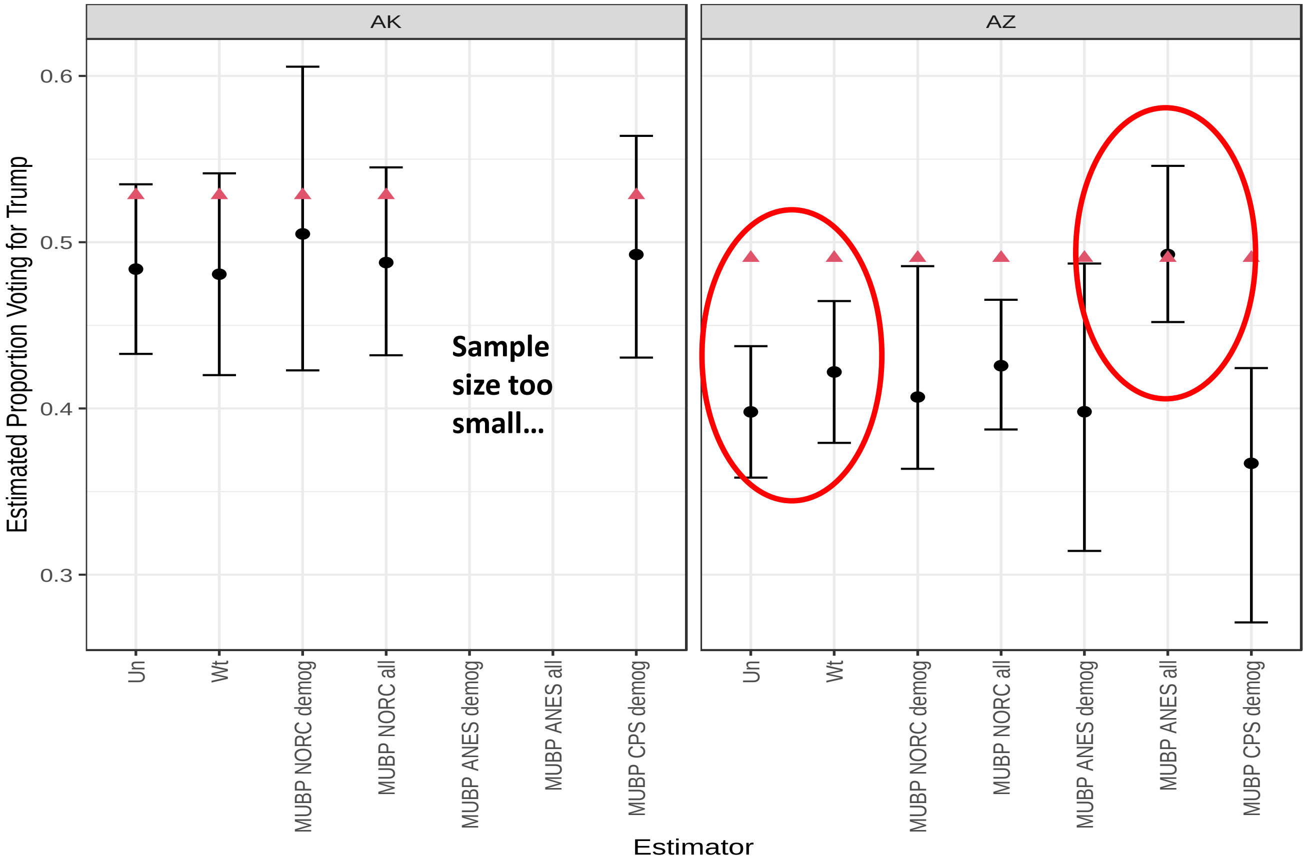
Results: True Bias and MUBP Bayes Intervals



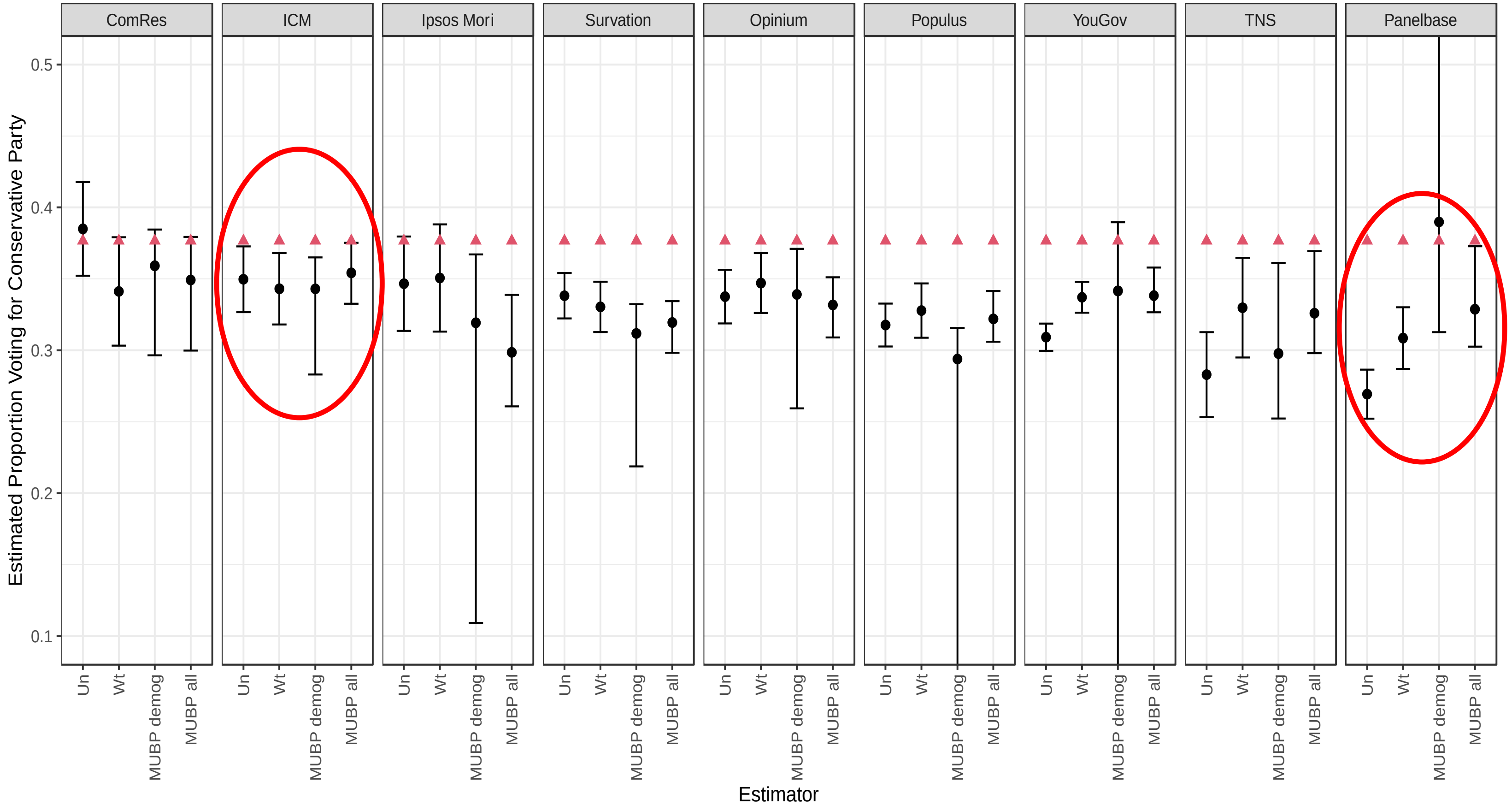
Results: ABC/WP Polls (U.S.)



Results: Siena/NYT Polls (U.S.)



Results: Great Britain Polls



No method was consistently best. Was the lack of covariates a problem?

Also, note the huge CIs when only demographics are used…

Results: Quality Measures

- PBR: Number of "Best" Results (remember: only focuses on bias!)
  - Weighted Estimate: 7 polls
  - MUBP, Demographics Only: 7 polls
  - MUBP, All Variables: 4 polls
  - **For 11/18 polls, the MUBP approach resulted in the best adjustment**
  - If we set aside two polls where the unweighted estimate was seemingly unbiased (Florida in the U.S. and ComRes in Great Britain), and two polls where <u>every adjustment</u> moved in the wrong direction, **the MUBP resulted in the "best" bias reduction for 10 of the 14 remaining polls**
  - Slight evidence of the ANES producing the best MUBP performance (as a population source for likely voters)

Results: Quality Measures

- Pseudo-RMSE: Number of “Best” Results
  - Unweighted Estimate: 4 polls
  - Weighted Estimate: 4 polls
  - **MUBP, Demographics Only: 2 polls**
  - **MUBP, All Variables: 8 polls**
  - For 10/18 polls, the MUBP approach resulted in the best adjustment
  - When weighted adjustments were “best”, the MUBP approach still produced similar results, suggesting minimal harm from using the MUBP approach as a general tool

Summary of Results

- Some polls produced unweighted estimates that were not biased; when estimates were biased, the MUBP approach seemed to help more
- MUBP adjustments based on *demographics only* have a tendency to have much wider credible intervals; relevant correlates of the measure of interest really seem to improve performance!
- The MUBP approach produces tighter credible intervals than the weighted approach, and when the weighted approach was “best” or the unweighted estimate was close, the MUBP results were similar
  - No harm in using this approach
  - The MUBP approach can do better in cases with more extreme bias

Discussion Points

- The MUBP measure offers a key advantage over standard weighting approaches, in that it allows for non-ignorable selection mechanisms
- Standard weighting approaches can still work well in some cases, but adjusted estimates based on the MUBP approach (and their credible intervals) were still quite similar
- There is clearly still room for improvement: Better covariates? Better sources of population data?
  - Important implications for measuring relevant covariates in large benchmark surveys like the ACS and the CPS
  - Correlations of the auxiliary proxies with the outcome variables were higher in the U.S. than in Great Britain, possibly due to better covariates

Discussion Points, cont’d

- Finding a high-quality probability sample measuring the same relevant covariates was a challenge; may not be as difficult in other settings outside of political polling
- Population means, variances, and covariances of the covariates estimated from these probability samples (necessary for the Bayesian MUBP approach) may themselves be biased, and are still estimates with associated uncertainty: how can we account for this?
- Additional applications of the MUBP approach in other settings are necessary, and we would love to hear about them!

Thank You! / Questions?

The paper containing all details of this study is currently under second review at POQ.

Please direct any and all inquiries to Brady West (bwest@umich.edu)!

References

- Andridge R.R. and Little, R.J.A. (2009). Extensions of proxy pattern-mixture analysis for survey nonresponse [conference paper]. *Proceedings of the 2009 Joint Statistical Meetings*, Section on Survey Research Methods, 2468-2482.
- Andridge, R.R. and Little, R.J. (2011). Proxy pattern-mixture analysis for survey nonresponse. *Journal of Official Statistics*, 27, 2, 153-180.
- Andridge, R.R., & Little, R.J. (2018). Proxy pattern-mixture analysis for a binary variable subject to nonresponse. *Journal of Official Statistics*, 36(3), 703-728.
- Andridge, R.R., West, B.T., Little, R.J.A., Boonstra, P., and Alvarado-Leiton, F. (2019). Indices of non-ignorable selection bias for proportions estimated from non-probability samples. *Journal of the Royal Statistical Society (Series C)*, 68(5), 1465-1483.
- Clinton, Joshua et al. 2020. "Task Force on 2020 Pre-Election Polling: An Evaluation of the 2020 General Election Polls." AAPOR. Available at https://www.aapor.org/Education-Resources/Reports/2020-Pre-Election-Polling-An-Evaluation-of-the-202.aspx (November 23, 2021).
- Clinton, Joshua, John S. Lapinski, and Marc J. Trussler. 2022. "Reluctant Republicans, Eager Democrats? Partisan Nonresponse and the Accuracy of 2020 Presidential Pre-Election Telephone Polls." Public Opinion Quarterly, DOI: https://doi.org/10.1093/poq/nfac011.
- Kennedy, Courtney et al. 2018. "An Evaluation of the 2016 Election Polls in the United States." Public Opinion Quarterly 82: 1–33.
- Little, R.J.A., West, B.T., Boonstra, P., and Hu, J. (2020). Measures of the degree of departure from ignorable sample selection. *Journal of Survey Statistics and Methodology*, 8, 932-964.