Public Opinion Quarterly (2023) **Vol 87 Special Issue,** 575–601

Evaluating Pre-election Polling Estimates Using a New Measure of Non-ignorable Selection Bias

Brady T. West[1,*], Rebecca R. Andridge[2]

[1]Research Professor, Survey Research Center of the Institute for Social Research, University of Michigan-Ann Arbor, Ann Arbor, MI, US
[2]Associate Professor, Division of Biostatistics in the College of Public Health, The Ohio State University, Columbus, OH, US

Abstract Among the numerous explanations that have been offered for recent errors in pre-election polls, selection bias due to non-ignorable partisan nonresponse bias, where the probability of responding to a poll is a function of the candidate preference that a poll is attempting to measure (even after conditioning on other relevant covariates used for weighting adjustments), has received relatively less focus in the academic literature. Under this type of selection mechanism, estimates of candidate preferences based on individual or aggregated polls may be subject to significant bias, even after standard weighting adjustments. Until recently, methods for measuring and adjusting for this type of non-ignorable selection bias have been unavailable. Fortunately, recent developments in the methodological literature have provided political researchers with easy-to-use measures of non-ignorable selection bias. In this study, we apply a new measure that has been developed specifically for estimated proportions to this challenging problem. We analyze data from 18 different pre-election polls: 9 different telephone polls conducted in 8 different states prior to the US presidential election in 2020, and nine different pre-election polls conducted either online or via telephone in Great Britain prior to the 2015 general election. We rigorously evaluate the ability of this new measure to detect and adjust for selection bias in estimates of the proportion of likely voters that will vote for a specific candidate, using official outcomes from each election as benchmarks and alternative data sources for estimating key characteristics of the likely voter populations in each context.

*Corresponding author: Brady T. West, Survey Research Center, Institute for Social Research, University of Michigan-Ann Arbor, 426 Thompson Street, Ann Arbor, MI 48109, US; email: bwest@umich.edu.

Advance Access publication June 8, 2023
© The Author(s) 2023. Published by Oxford University Press on behalf of American Association for Public Opinion Research.
All rights reserved. For permissions, please email: journals.permissions@oup.com
https://doi.org/10.1093/poq/nfad018

Introduction

The apparent "failures" of pre-election polling to indicate actual winners in recent local and national elections in the United States and Great Britain have been the subject of a significant amount of media scrutiny (e.g., Alcantara et al. 2016; Sturgis et al. 2016) and cast doubts on the ability of pre-election polls to ever succeed at indicating election outcomes in the future (e.g., Duncan 2016). Prior studies have provided evidence of national polls being more or less accurate in estimating the overall proportion of votes that will be cast for a given candidate in the United States (Kennedy et al. 2018; Clinton et al. 2020). However, significant "misses" of estimates based on carefully designed polls in states that have a large impact on overall election results (due to the US electoral college system, which can produce election outcomes that differ from the candidate receiving the most national votes) will tend to receive much more scrutiny (Kennedy et al. 2018; Clinton et al. 2020). Careful studies of the underlying reasons for these polling "misses" in local, state, and national elections, both in the United States and more globally, are still necessary.

Among the numerous explanations that have been offered for these recent polling errors, the role of selection bias that cannot be corrected by standard weighting approaches has received relatively less focus in the academic literature (e.g., Kennedy et al. 2018; Clinton et al. 2020; Clinton, Lapinski, and Trussler 2022). Indeed, the American Association for Public Opinion Research (AAPOR) has called for more research into this issue (Clinton et al. 2020). Such selection bias may arise due to partisan nonresponse (Silver 2016; McAuliffe et al. 2021; Clinton, Lapinski, and Trussler 2022), where the probability of responding to a poll is a function of the candidate preference that a poll is attempting to measure, even after conditioning on other relevant variables that are often used for weighting adjustments. If this type of selection mechanism is operating, estimates of candidate preferences based on individual or aggregated polls may be subject to significant bias, even after standard weighting adjustments. Recent work has suggested other potential sources of this selection bias, including the reluctance of people with antiestablishment views to participate in polls (e.g., Crawford, Levy, and Backus 2022), "shy" voters who support a particular candidate, and "late swings" (e.g., Sturgis et al. 2016).

Importantly, weighting adjustments designed to correct for selection bias make the critical assumption that selection (due to sampling or nonresponse) is occurring at random when conditioning on the auxiliary variables used to compute the weights. When this assumption is true, Little and Rubin (2019) would label individuals who do not participate in a poll as "missing at random" (MAR). This type of selection, which is a type of *ignorable* selection,

means that the probability of participating in a pre-election poll is essentially constant within a subgroup of individuals defined by some combination of values on the variables used to perform the weighting adjustments. If enough "relevant" auxiliary variables correlated with both candidate preference and nonresponse propensity are used in the weighting adjustments, then these adjustments can shift biased estimates of candidate preference in the correct direction (Kalton and Flores-Cervantes 2003; Little and Vartivarian 2005; Särndal and Lundström 2005; Bethlehem 2009; Brick 2013; Silver 2016; Kennedy et al. 2018).

If, however, selection is *not* occurring at random and is a function of the measures that one is trying to collect (e.g., candidate preference), even after conditioning on the variables used for weighting adjustments, this is referred to as *non-ignorable* selection or "missing not at random" (MNAR) in the spirit of Little and Rubin (2019). In these settings, there is an unobserved factor (such as partisanship) that is predicting response to a poll invitation and support for a particular candidate simultaneously (Groves 2006), leading to bias in the resulting estimate, and weighting adjustments will be ineffective due to the lack of an auxiliary measure of the unobserved factor. This type of selection in pre-election polling has been suspected in multiple prior reports (Clinton et al. 2020; McAuliffe et al. 2021; Clinton, Lapinski, and Trussler 2022), and if one is unable to identify observed correlates of candidate preference that could be used in weighting adjustments, adjustment for selection bias may not be possible. See Haziza and Beaumont (2017), Bethlehem (2009), or Kalton and Flores-Cervantes (2003) for more general discussions of alternative weighting approaches that might be used in polling applications.

Weighting adjustments can also increase the margin of error attached to estimates of candidate preferences without correcting for bias, as increased variability in the weights due to the nonresponse adjustments may increase the standard errors of survey estimates. This is especially true if variables that predict only response or only selection propensity are used for the adjustments (Little and Vartivarian 2005; Andridge and Thompson 2015; Heeringa, West, and Berglund 2017). If the auxiliary variables used for weighting adjustments are correlated with the variables of interest, then both the bias and variance in the weighted estimates may be reduced (Little and Vartivarian 2005; Bethlehem 2009). In this context, approaches that can measure and adjust for the bias introduced by MNAR (or non-ignorable) selection mechanisms while at the same time reducing the margin of error may have promise in helping to solve this difficult estimation problem.

Response rates in pre-election polls that use probability samples, which produce unbiased estimates of population preferences in the absence of nonresponse (Horvitz and Thompson 1952), have also been steadily declining in

recent years (AAPOR 2017; Kennedy et al. 2018; Clinton et al. 2020). Some have argued that probability samples with low response rates can no longer be thought of as probability samples (Gelman 2012), meaning that alternatives to traditional design-based estimation approaches may be needed (Heeringa, West, and Berglund 2017). Dutwin and Buskirk (2017) suggest that the rigorous design of probability samples can still produce estimates with reduced bias if response rates are low, relative to non-probability sampling approaches. Bethlehem (2020) arrives at a similar conclusion by deriving the maximum absolute bias (MAB) in an estimated mean for a probability sample with a moderate response rate and a large non-probability sample with a small response rate, showing that the MAB is substantially larger for the non-probability sample. Probability samples therefore seem to have merit, but the traditional practice of weighting on a handful of demographics may no longer be effective at eliminating bias if additional forms of selection (such as non-ignorable nonresponse) are operating. There is therefore a significant need to develop new methods and approaches to estimating this bias and correcting for it.

In this study, we apply and evaluate a new measure for quantifying potentially non-ignorable selection bias in pre-election estimates of the proportions of likely voters who support a particular candidate. This new measure can be applied generally to either low response rate probability samples, which unfortunately describes many recent pre-election polls in the United States (Kennedy et al. 2018), or non-probability samples (which are also widely used for pre-election polling; Chen, Valliant, and Elliott 2019; Clinton et al. 2020). Importantly, we do not propose a solution for the low response rates in pre-election polls; rather, we describe approaches to adjusting estimates for the selection bias in estimated proportions that may result from the low response rates. We first evaluate pre-election estimates of the proportion of likely voters who support President Trump in the 2020 US presidential election, based on nine pre-election telephone polls conducted in eight different US states. Next, we evaluate pre-election estimates of the proportion of likely voters who support the Conservative Party in the 2015 general election in Great Britain, based on nine pre-election polls conducted either online or by telephone (Sturgis et al. 2016). All polls that we evaluate were conducted within approximately two months of the corresponding election, meaning that there is little reason to expect variability in accuracy due to poll timing. We compare standard weighting approaches assuming ignorable selection and our general adjustment approach allowing for non-ignorable selection in terms of their ability to shift the pre-election estimates in the direction of the official outcomes from both elections.

Methods

A New Measure of Selection Bias for Estimated Proportions

Measures that quantify potential non-ignorable selection bias have recently been proposed by our research group for estimates of means (Little et al. 2020), proportions (Andridge et al. 2019), and regression coefficients (West et al. 2021). Since our interest lies in quantifying the selection bias in pre-election estimates of the proportions of likely voters who would support a given candidate, we use the measure of unadjusted bias for a proportion (MUBP) proposed by Andridge et al. (2019), which we briefly describe here in the context of pre-election polling. Readers interested in the technical details underlying this measure can refer to Andridge et al. (2019) or the more mathematical summary of this approach provided in Supplementary Material section 1.

Suppose that we have a target population of likely voters, and we wish to estimate what proportion of this population would vote for a given candidate in a given election. A single pre-election poll gathers data from a sample of these likely voters, resulting in a binary inclusion indicator (whether an individual from the target population was included in the responding sample). We only observe the values of a binary indicator of candidate preference (the survey variable of interest) for individuals included in the poll. Although many pre-election polls are probability samples, they tend to have very low response rates (e.g., Kennedy et al. 2018; Clinton et al. 2020), and researchers designing these polls do not have control over the probability that an individual self-selects to respond to the poll. We therefore treat the pre-election poll as a non-probability sample, where the distribution of the inclusion indicator is not known in advance.

The poll also captures data on additional covariates describing each responding sample unit (e.g., age, education). Using data from the poll respondents, we fit a probit regression model to the binary candidate preference indicator, including all the covariates as predictors and carefully checking the model for acceptable fit and good predictive power (Hosmer, Lemeshow, and Sturdivant 2013). For each respondent, we then compute a linear predictor of the candidate preference indicator based on the estimated regression coefficients in the probit model and the values of the responding unit's covariates. We refer to this as a *proxy* variable, representing the best predictor of candidate preference based on the available (and ideally relevant) covariates; doing so effectively reduces the multidimensional covariate set into a single variable that is most predictive of candidate preference. A probit regression model is used for the binary indicator of interest because this model assumes that the observed indicator arises from an underlying, unobserved latent variable that follows a normal distribution. This leads to

mathematical and computational convenience in computing the MUBP, which is an extension of earlier indicators assuming that the variable of interest is normally distributed (Andridge et al. 2019; Little et al. 2020).

Critical to the estimation of non-ignorable selection bias based on the MUBP is the availability of *aggregate* information for the non-selected likely voter population on the available covariates. This aggregate information takes the form of population means, variances, and covariances for these covariates; for example, if age and education were the two available covariates, we would compute the means and variances of age and education for the target population of likely voters, along with their covariance. Unlike other approaches to adjusting for non-ignorable selection bias based on selection models (e.g., Heckman 1976), computation of the MUBP *does not* require microdata for the non-selected cases. Given this aggregate information, we can compute an estimate of the population mean and variance of the aforementioned proxy variable. Specifically, plugging the mean for each covariate into the fitted probit regression model produces the estimated mean of the proxy, and its variance can also be estimated (see Andridge et al. 2019 for details). In the context of pre-election polls, the size of the respondent sample will be far lower than the size of the likely voter population, and thus aggregate information for the full population of likely voters is effectively the same as for the subpopulation of likely voters not included in a given poll.

The basic idea of the MUBP is that we can measure the degree of selection bias present for the selected sample mean of the computed proxy variable, since we have an estimate of the population-level mean for this variable (based on the aggregate population-level information for the covariates). If the candidate preference indicator is correlated with this proxy variable in the selected sample, then this provides information about the potential selection bias in the proportion estimated based on the respondent sample; similar rationale has been proposed by Bethlehem (2009) and Särndal and Lundström (2005). Andridge et al. (2019) use what is known as a *bivariate normal pattern-mixture model* (PMM) for the joint distribution of the *underlying* normal latent variable of interest and the auxiliary proxy variable to develop the MUBP measure. The PMM specifies separate mean and variance parameters for these two variables, specifically for two subgroups: the likely voters who responded to the poll, and all other likely voters in the population who were *not* included in the poll. The proportion of *all* likely voters who would vote for a given candidate can then be estimated as a weighted average of the estimated proportions for each of the two subgroups.

There is no information in the available data that we can use to estimate the parameters of the distribution of the underlying normal latent variable for the likely voters who are not included in the poll, and thus additional assumptions are necessary to produce estimates. Specifically, we assume that

the sample inclusion indicator is an *unspecified* function of a known linear combination of the auxiliary proxy variable and the underlying normal latent variable for the candidate preference indicator. The exact form of the function (e.g., logit or probit) does not need to be specified for inferences based on the PMM to be valid (Andridge and Little 2011). This function includes a parameter, denoted by ϕ, that describes how much inclusion in the respondent sample depends on the (unobserved) latent variable for candidate preference versus the observed proxy variable. This parameter ϕ is an unknown sensitivity parameter—meaning it is inestimable from the data—that takes values between 0 and 1. The possibility of considering alternative values for this sensitivity parameter (and thus the extent that selection is not occurring at random) is a key distinguishing feature of the MUBP approach relative to other adjustment approaches in the literature.

Prior work (Andridge et al. 2019; Andridge and Little 2020) has shown that for a specified value of the sensitivity parameter ϕ, we can use the PMM to estimate the proportion of interest. Importantly, the choice of ϕ corresponds to a particular selection mechanism. If $\phi = 0$, then the distribution of the sample inclusion indicator depends on only the proxy variable, which itself is based on observed covariates, corresponding to an MAR (or *ignorable*) selection mechanism. This assumption is implicitly made when using weighting adjustments based on covariates to adjust for selection bias. If $\phi > 0$, then selection depends at least partially on the latent variable underlying the binary variable of interest (candidate preference), making the selection mechanism MNAR (or *non-ignorable*). For a chosen ϕ, the proportion of all likely voters who would vote for a specific candidate is then estimated as a weighted average of the estimated proportion in the selected respondent sample (the pre-election poll respondents) and the estimated proportion based on the PMM for the non-selected cases. The MUBP is then the difference between the selected sample proportion and the estimated overall proportion.

The MUBP intuitively works as follows. If there is little selection bias evident for the auxiliary proxy and this variable is strongly correlated with the binary indicator of candidate preference, then there is likely to be little selection bias in the selected sample proportion, even with a non-ignorable selection mechanism. If these two variables are highly correlated but there *is* evidence of selection bias for the mean of the auxiliary proxy, then the MUBP is able to estimate the bias in the estimated proportion accordingly, based on the selection bias for the auxiliary proxy and the relationship between the proxy and the binary indicator of interest. If, however, there is *not* a strong correlation between these two variables, then regardless of the selection bias evident in the auxiliary proxy, there is not much evidence for or against bias in the estimated proportion based on the selected sample. In this scenario there will be much more uncertainty in the MUBP, reflecting the

limited information available about possible non-ignorable selection bias due to the weak predictive power of the covariates.

Andridge et al. (2019) describe two estimation approaches for the MUBP. In this study, we use their Bayesian approach, which accounts for uncertainty in the creation of the auxiliary proxy based on the probit model fitted to the polling data (see Gill 2014 for a general introduction to Bayesian methods in the social sciences). Using the steps described in Andridge and Little (2020, section 3.2), this Bayesian procedure simulates many draws of the MUBP from a posterior distribution for the MUBP based on the PMM described above, and then generates a 95 percent credible interval for the MUBP that effectively averages over all possible ϕ values, that is, over varying degrees of MNAR/non-ignorable selection mechanisms (including those arising from random sampling).[1] Once we have the 95 percent credible interval for the MUBP for a particular pre-election poll, we can use this interval to adjust the estimated proportion based on the selected sample. The MUBP-adjusted proportion is estimated by subtracting the 50th percentile of the MUBP draws from the selected sample proportion; subtracting the 2.5th and 97.5th percentiles of the MUBP provides lower and upper bounds for the MUBP-adjusted proportion. Importantly, the selected sample proportion used in these adjustments is the *unweighted* proportion from the pre-election poll.

The ABC/Washington Post Pre-election Polls (available via PARC)

Overview

Microdata for the seven most recent (i.e., closest to the election) 2020 pre-election polls conducted by ABC and the *Washington Post* in key battleground states (Michigan, Wisconsin, Pennsylvania, Minnesota, Florida, Arizona, and North Carolina) were obtained from the publicly accessible PARC ABC/*Washington Post* (ABC/WP) polling archive.[2] These data were collected by Abt Associates in September and October 2020, meaning that in theory they may produce more accurate estimates than earlier polls. Similar

1. In this study, this procedure begins with random draws of the ϕ parameter from a Uniform(0,1) prior distribution, effectively allowing for draws of the MUBP measure corresponding to all possible selection mechanisms. Alternative prior distributions for this parameter are certainly possible. For example, West et al. (2021) consider a prior that assigns equal probability to three possible values for the ϕ parameter: 0, 0.5, or 1.0. This prior places higher probability on an entirely missing-at-random (or ignorable) selection mechanism when computing the MUBP values. Prior simulations have indicated that this type of prior produces more posterior variance in the MUBP values, so we use the Uniform prior in these applications. See Supplementary Material section 1 for more details. Future replications of the work presented here could certainly consider alternative priors.

2. https://2020electionpolls.parc.us.com/client/index.html#/.

ABC/WP polling data are not currently available for any other states. Briefly, these pre-election polls were all conducted using dual-frame probability sampling of landline and cellular telephone numbers, and interviews were conducted using computer-assisted telephone interviewing (CATI).[3] Weights for poll respondents were computed based on probabilities of selection that accounted for ownership of landline and cellular phones, and these weights were adjusted to known population control totals on various socio-demographic variables from the National Health Interview Survey (NHIS) and the US Census Bureau.[4] Counts of respondents for these polls ranged from 777 to 1,043, and AAPOR RR4 *combined* response rates (designed for these types of dual-frame telephone number samples) ranged from 4.5 percent to 6.5 percent.

Variables

We used the data from these seven polls to estimate the proportion of likely voters in each state that will vote for Donald Trump in the 2020 presidential election. Additional variables measured in these polls were used as covariates for calculating MUBP-adjusted estimates if they met two key criteria: (1) they were *relevant* predictor variables in the probit regression models for the Trump vote indicators (e.g., Kennedy et al. 2018), and (2) they would be readily available in other large probability samples of likely voters, for the purpose of computing the required aggregate means, variances, and covariances for this target population. Table 1 details the variables used in each of the seven ABC/WP data sets. Only data from respondents who were classified as likely voters (as described in table 1) were used for analyses.

Alternative data sources for the target population of likely voters

A critically important component of the new MUBP measure is having a high-quality external source of population information for estimation of population means, variances, and covariances for the exact same covariates used to form the auxiliary proxy (table 1). Given our specific application, we needed to compute these means, variances, and covariances for a target population of *likely voters*. To this end, we reviewed publicly available micro-data from several large probability-based surveys that also included measures of our covariates of interest, including the November 2020 Current Population Survey (CPS) voter supplement,[5] the 2020 American National

3. We only analyze data from US pre-election polls conducted over the telephone in this study. Online pre-election polls are certainly conducted in the United States; more than 60 percent of pre-election polls were done online in 2020 (Clinton, et al. 2020). The only *publicly available* pre-election polling data at the time of this writing were from telephone polls.

4. https://abcnews.go.com/US/PollVault/abc-news-polling-methodology-standards/story?id=145373.

5. https://cps.ipums.org/cps/.

Table 1. Variables computed in the seven ABC/WP data sets.

Variable	Possible values
Likely voter indicator	1 = Likely voter in 2020; 0 = Not likely voter in 2020 (Likely = registered to vote/will register to vote by voting day AND absolutely certain to vote/will probably vote/chances 50-50 will vote/already voted)
Trump voting indicator	1 = Already voted for Trump/intends to vote for Trump/likely to vote for Trump; 0 = Intends to vote for different candidate; Missing = Not registered to vote/not planning to vote/don't know/no opinion
Male indicator	1 = Male; 0 = Female
Age category	Binary (1,0) indicator variables for: 18–24, 25–29, 30–39, 40–49, 50–64, and 65+ (99 = missing)
Education category	Binary (1,0) indicator variables for: HS or less, Some college, College, and Post-college (don't know/refused = missing)
Race/ethnicity	Binary (1,0) indicator variables for: White, Black, Hispanic, and Other (don't know/refused = missing)
Ideology	Binary (1,0) indicator variables based on 5-point ideology scale for: Liberal ideology (very liberal/somewhat liberal), Moderate ideology (moderate), and Conservative ideology (very conservative/somewhat conservative) (don't know/no opinion = missing)

(continued)

Table 1. Continued.

Variable	Possible values
Political party identification	Binary (1,0) indicator variables for: Democrat/leans Democratic, Republican/leans Republican, and Independent/Other (don't know/refused = missing)

Election Studies (ANES) pre-election survey,[6] and the AP/NORC VoteCast 2020 data.[7]

Ideally, one of these data sources would contain all covariates in the poll data that were highly predictive of candidate preference (table 1) and would be available at the time the pre-election polls were conducted, as this would enable the MUBP to be used as a potential *a priori* indicator of selection bias. None of the data sources were ideal in this sense; we therefore conducted our analyses using each population source separately and compared their performance. The AP/NORC VoteCast 2020 data contained likely voter indicators and all predictor variables of interest for large samples from each state, but this data source was not entirely based on a probability sample, and it would not have been available at the time of computing *pre-election* estimates in 2020. The November 2020 CPS supplement available prior to the election did not include the same highly relevant measures of ideology and party preference that were available in the poll data (table 1) and also has other known limitations in terms of population representation (Ansolabehere, Fraga, and Schaffner 2021). The ANES did include the political measures, but both the CPS and ANES data sets had smaller samples available for estimation of the aggregate population features from each state than the AP/NORC data set. By repeating our analyses using each population source, we can assess whether having aggregate population measures available for more *relevant* political measures serves to improve the performance of the MUBP measure. We revisit issues with finding good sources of population information more generally in the Discussion section.

We obtained the official 2020 presidential election results for each state from the Massachusetts Institute of Technology (MIT) Election Data and Science Lab.[8] We extracted the proportion of officially counted votes in each state that were cast for President Trump and saved these proportions for our analyses as the benchmark truth in each state.

6. https://electionstudies.org/data-center/2020-time-series-study/.
7. https://apnorc.org/projects/ap-votecast/.
8. https://doi.org/10.7910/DVN/42MVDX.

The New York Times/Siena College Pre-election Polls (available via Roper)

Overview

Microdata for 2020 pre-election polls designed by the *New York Times* were available for only two states from the Roper Institute archive:[9] Arizona ($n = 653$; 5.4 percent landline response rate, 4.7 percent cell response rate) and Alaska ($n = 423$; 8.1 percent landline response rate, 9.1 percent cell response rate). These two dual-frame telephone polls were conducted in September (Arizona) and October (Alaska) 2020 by Siena College Research Institute, ReconMR, the Institute for Policy and Opinion Research at Roanoke College, and M. Davis and Company. For each poll, a likely-to-vote probability was computed for each respondent based on their stated likelihood to vote and by virtue of the imputation of a turnout probability score based on past voting behavior applied to their specific voting history. Final weights for the poll respondents were computed based on this probability to vote weight along with calibration adjustments that used age, region, education, and gender distributions from the American Community Survey and the Current Population Survey.

Variables

Nearly all the same variables analyzed in the ABC/WP polls were also analyzed in the two NYT/Siena data sets, including the Trump indicator, the likely voter indicator, gender, age, race, and education (see the specific variables used to compute these measures in Supplementary Material table S1). Party identification was available but without evidence of "leaning" toward a particular party. The only variable not available was political ideology, which was not measured in the NYT/Siena polls. The probit regression models used for the NYT/Siena data sets therefore did not benefit from quite as much "relevant" information.

Alternative data sources for the target population of likely voters

All the variables available in the NYT/Siena data sets could also be computed in both the AP/NORC VoteCast data and the ANES for purposes of estimating the means, variances, and covariances of these variables for the likely voter population. However, the sample size from the ANES was too small for Alaska to be useable ($n = 9$); for Arizona, the sample size was also small ($n = 158$), but these data were still used. As with the analysis of the ABC/WP polls, the CPS only enabled estimation of population aggregates for sociodemographics and provided smaller samples than the AP/NORC

9. https://ropercenter.cornell.edu/ipoll.

VoteCast data for both states. The MIT data were once again used to define the official benchmark election results.

The 2015 Great Britain Pre-election Polls

Overview

The inability of pre-election polls in Great Britain to predict the actual election outcome in the 2015 general election (a rather convincing victory for the Conservative Party) was the subject of an extensive government study (Sturgis et al. 2016). We were provided access to data from nine pre-election polls (from that study) that were conducted between the months of March and May 2015 by nine different polling firms. Six polling firms conducted these polls using opt-in web panels, one firm used a combination of an opt-in web panel and a dual-frame telephone sample, and two firms used dual-frame telephone samples exclusively (see the Appendices in Sturgis et al. 2016 for details, including response rates). Similar to the NYT/Siena polls, the weighting approach used in these polls adjusted for a predicted likelihood of voting in 2015, along with calibration adjustments based on known population distributions on age and gender; see Appendix 4 of Sturgis et al. (2016) for details. We only consider respondents in these polls with weights greater than zero for our analyses.

Variables

Our focus was the proportion of likely voters who intended to vote for the Conservative Party candidate. Covariates for use in estimating the MUBP were limited to a set of harmonized variables that were collected similarly across the nine polls, including age, gender, Standard Government Office Region (geographical location within the United Kingdom), and a set of indicators for which party the respondent reported voting for in the 2010 general election (Conservative, Labour, Liberal Democrat, or Other; "did not vote" was coded as missing). We only analyzed data collected from Great Britain (excluding Northern Ireland) per recommendations of the authors of the 2015 United Kingdom (UK) polling report (Sturgis et al. 2016), meaning that these results only apply to Great Britain.

Data source for the target population of likely voters

Per recommendations of the authors of the 2015 UK polling report, we used the 2015 British Election Study (BES) as the source of population information on likely voters in the general election. This publicly available data set[10] provides variables that measure the exact same categories of age, gender,

10. http://www.britishelectionstudy.com/data-objects/cross-sectional-data/.

region, and 2010 vote that were available in the nine pre-election polls (Fieldhouse et al. 2015). This data set also provides an indicator of whether the respondent voted in the 2015 election, and we used this indicator to define the likely voter population. The final BES weights were used to compute the population aggregates necessary for the MUBP computations.

We used 37.7 percent as the official benchmark percentage of the Great Britain voters that voted for the Conservative Party in 2015 (Parliament House of Commons, 2015) for evaluating the performance of the standard weighted estimates and the MUBP adjustments.

Statistical Evaluation of Selection Bias and Competing Adjustments

For each of the 18 polls, we began by computing the *unweighted* estimate of the proportion of likely voters (or respondents with a nonzero weight, in the case of the pre-election polls in Great Britain) that would vote for a given candidate (Trump in the United States, or the Conservative Party in Great Britain), along with a 95 percent confidence interval assuming simple random sampling. We then computed the corresponding *weighted* estimate of this proportion (using the appropriate final weight variable in each data set) as one type of adjustment approach, along with an appropriate design-based 95 percent confidence interval using Taylor Series Linearization to compute standard errors (Heeringa, West, and Berglund 2017). Notably, we did not have access to the control totals used for the calibration adjustments in the polling data sets, which likely would have reduced the standard errors of the weighted estimates.

Next, we applied the Bayesian approach described earlier to compute a posterior median of the MUBP measure of selection bias, a 95 percent credible interval[11] for the selection bias, an adjusted estimate of the proportion of interest based on the MUBP measure, and an adjusted 95 percent credible interval for the proportion of interest. To evaluate the performance of the MUBP measure in a variety of scenarios distinguished by the richness of the available population information, we considered these alternative approaches for each poll:

1. Using *sociodemographic variables only* in the probit regression model and from the relevant population data source; or

11. In Bayesian analysis, where parameters of interest are random variables, a 95 percent credible interval describes a range of values that will include 95 percent of the values that the random parameter will take, given prior assumptions about the distribution of the parameter and the observed data. This is in contrast to a 95 percent confidence interval in frequentist analysis, which can be interpreted as covering the true parameter of interest 95 percent of the time in repeated samples of the same size, given that the interval is constructed the same way for each sample.

2. Using *all available variables* (including measures of political party preference and ideology for the US polls, and 2010 vote history for Great Britain polls) in the probit regression model and from each population data source.

We repeated the two analyses above for each alternative population data source (AP/NORC VoteCast, ANES, CPS) when analyzing the ABC/WP and NYT/Siena polls. We remind readers that when using the CPS, only demographic aggregates were available, meaning that only the first approach was possible when using the CPS supplement.

Quality measures

We compared the MUBP-based adjustments and the more standard weighting adjustments in terms of their ability to shift estimates closer to the true election outcomes. In evaluating the quality of a given estimate, we used three approaches:

1. A visual assessment of bias, via comparison of the point estimate to the official election result, including 95 percent confidence or credible intervals.
2. Calculation of a measure capturing the *proportion of bias removed* (PBR) by an adjustment: PBR = (adjusted est. – unweighted est.)/(true proportion – unweighted est.). This PBR measure provides an indication of whether an adjustment exacerbates the bias (percent bias removed < 0), removes some or all of the bias (percent bias removed between 0 percent and 100 percent), or overshoots the bias removal (percent bias removed > 100 percent).
3. Calculation of a *pseudo-RMSE* (root mean squared error) measure, capturing both bias and variance. This measure first computed the bias in a given point estimate as the difference between the point estimate and the official election result. Next, the variance of a given estimate was estimated as either the linearized variance estimate (for weighted estimates) or the half-width of the 95 percent credible interval divided by 1.96 and squared (for the MUBP-adjusted estimates), and added to the squared bias estimate. The resulting sum is a standardized overall quality measure that can be used to compare performance across the alternative estimation approaches.

Results

Figure 1 presents the point estimates for the ABC/WP polls along with 95 percent confidence or credible intervals (depending on the estimation method

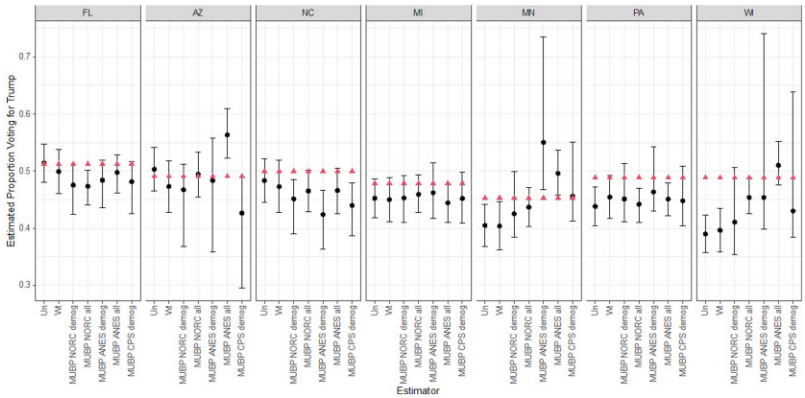

Figure 1. ABC/WP estimates and 95 percent confidence or credible intervals (triangles = official outcomes in each state).

used). Here we see general support for the use of the MUBP adjustments based on *all available covariates* when considering the official election outcomes in each state. The bias correction is particularly noteworthy in Minnesota and Wisconsin, where the unweighted estimates had notable bias and weighting adjustments assuming an MAR selection mechanism ("Wt" in figure 1) did not repair this bias. We note that in several cases, MUBP adjustments based on demographics only (e.g., "MUBP ANES demog," "MUBP NORC demog") have significantly wider credible intervals than MUBP adjustments that include additional relevant predictors (Party ID and political ideology, in this specific context). This demonstrates that the inclusion of as many relevant predictors of the outcome variable as possible will help reduce uncertainty in the measures of selection bias and corresponding adjusted estimates.

Figure 2 presents the estimates based on the two NYT/Siena polls, and patterns are similar to those observed in figure 1 for the ABC/WP polls. In Arizona, the MUBP adjustment based on all available covariates from the ANES had the best performance, repairing the selection bias that could not be corrected by the weighting adjustment. In Alaska, there was not much difference among the estimates, but all adjusted estimates based on the MUBP were slightly closer to the truth than the weighted estimator (which actually was more biased than the unweighted estimate). For both states, adjustments based on demographics only once again tended to have more uncertainty.

Figure 3 presents the estimates for the nine pre-election polls conducted in Great Britain prior to the 2015 general election. We once again note the increased uncertainty in MUBP adjustments based on demographics only. The MUBP adjustments based on all covariates (including the 2010 vote) and the

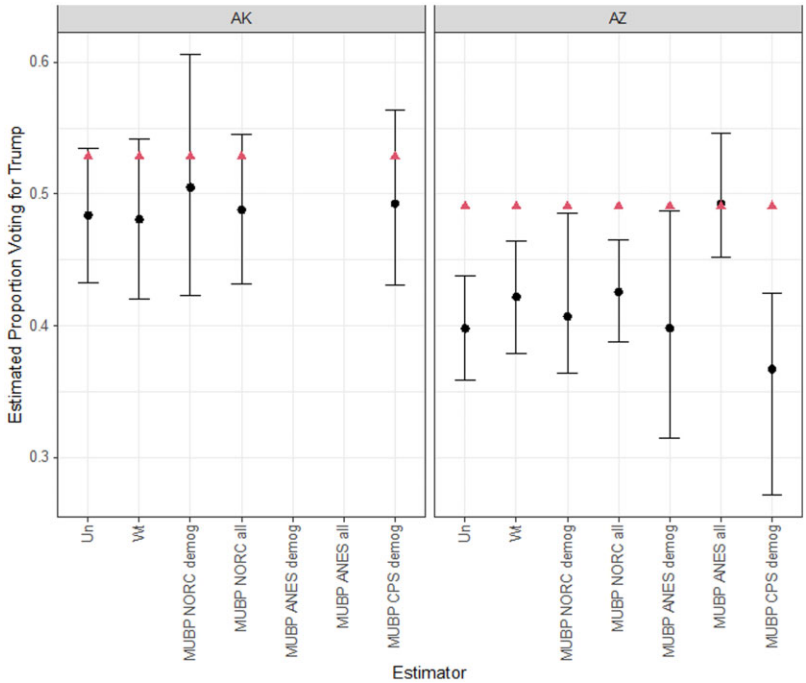

Figure 2. NYT/Siena estimates and 95 percent confidence or credible intervals (triangles = official outcomes in each state).

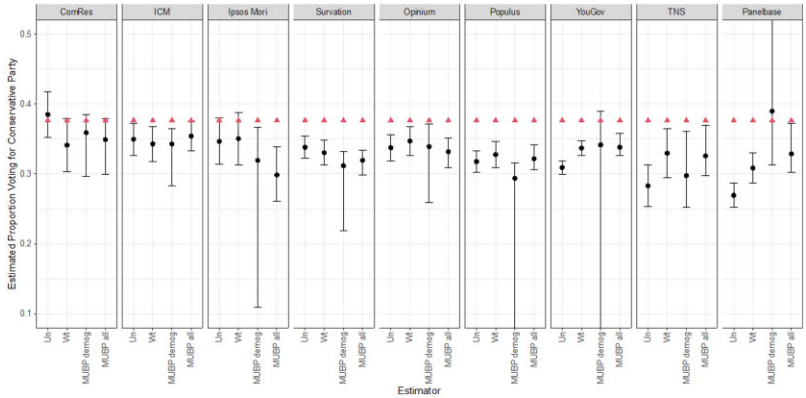

Figure 3. Visual comparison of estimates from the nine polls in Great Britain, with 95 percent confidence or credible intervals (triangles = official 2015 general election outcome; intervals that go outside of the figure have lower limits of 0 or upper limits of 1).

weighting adjustments generally had similar performance, with no method having consistently better performance in all polls. The MUBP approach was able to correct some of the bias that weighting could not in two of the polls (ICM and Panelbase), but no adjustments were able to correct the bias in the majority of the polls. This may reflect the lack of relevant covariates in the harmonized data set to which we had access for this study, or it may be an indication that non-ignorable selection bias was not the primary driver of the polling misses.

Next, table 2 presents the results of our comparative analysis of the quality metrics (PBR and pseudo-RMSE). In terms of the PBR metric, the MUBP produced the "best" bias removal for 11 of the 18 polls. In two polls (Florida in the United States and ComRes in Great Britain), the unweighted estimate had very little bias, resulting in severely inflated PBR values. In two additional polls (North Carolina in the United States and Survation in Great Britain), all of the adjustments (including weighting) moved the estimates in the wrong direction. If we set aside these four polls, the MUBP had the "best" bias removal for 10 of the remaining 14 polls. When also considering the second-best adjustment for each poll, we find general evidence of the MUBP adjustment based on all covariates (not just demographics) tending to have consistently strong performance. In the US context, there was slight evidence of MUBP adjustments based on the ANES tending to have the best performance, although each population source emerged as producing the best adjustment for at least one poll. Notably, in 7 of the 18 polls (prior to excluding the four unusual polls) and 6 of the 14 polls (after the exclusions), the best PBR results were for MUBP adjustments based on demographics only.

These percent bias removed results should be interpreted in the context of the increased uncertainty in the adjustments noted in figures 1–3. Our pseudo-RMSE quality metric also incorporates the uncertainty of the adjustment estimates and provides additional support for the performance of the MUBP adjustment. For 10 of the 18 polls, the MUBP adjustment had the lowest pseudo-RMSE, with the MUBP adjustment based on all covariates (not just demographics) having the lowest pseudo-RMSE in 8 of these 10 polls. Interestingly, all four of the polls for which the standard weighting adjustment produced the lowest pseudo-RMSE quality metric were in Great Britain, with three of these four polls from opt-in web panels. However, the MUBP adjustments based on all available covariates had similar pseudo-RMSE values in the majority of these cases, suggesting that allowing for an MNAR mechanism may not harm the overall quality of the adjusted estimates significantly. Overall, given the similar performance of the MUBP adjustments in cases where weighting adjustments were found to work "best" (e.g., TNS, Populus), we find general support for the use of the MUBP adjustments in table 2.

Table 2. Comparisons of quality metrics (boldfaced measures indicate the "best" results for a given poll).

			Percent bias removed			Pseudo-RMSE			
Poll (dates, mode)	Bias in unweighted est. %	Population source	Weighted %	MUBP – Demos only %	MUBP – All vars. %	Unweighted	Weighted	MUBP – Demos only	MUBP – All vars.
ABC/WP—Florida[a]	0.18	AP/NORC	**827.1**	2,155.9	2,264.9	**0.017**	0.023	0.043	0.042
(10/24/20–10/29/20, phone)		ANES		1681.1	924.4			0.035	0.022
		CPS		1829.7	[c]			0.039	[c]
ABC/WP—Arizona	1.25	AP/NORC	241.3	287.2	**69.1**	0.023	0.029	0.044	**0.020**
(9/15/20–9/20/20, phone)		ANES		157.1	−482.2			0.051	0.076
		CPS		614.0	[c]			0.081	[c]
ABC/WP—North Car.[b]	−1.62	AP/NORC	**-64.0**	−197.4	−110.6	**0.025**	0.035	0.054	0.039
(10/12/20–10/17/20, phone)		ANES		−366.1	−105.6			0.080	0.039
		CPS		−267.1	[c]			0.064	[c]
ABC/WP—Michigan	−2.61	AP/NORC	−9.2	1.5	25.8	0.031	0.035	0.033	**0.026**
(10/20/20–10/25/20, phone)		ANES		**37.3**	−30.6			0.030	0.038
		CPS		−1.3	[c]			0.035	[c]
ABC/WP—Minnesota	−4.82	AP/NORC	−1.9	42.2	66.7	0.052	0.054	0.040	**0.024**
(9/8/20–9/13/20, phone)		ANES		302.2	189.3			0.119	0.047
		CPS		**106.3**	[c]			0.035	[c]
ABC/WP—Pennsylvania	−5.04	AP/NORC	32.4	25.8	7.5	0.053	0.039	0.046	0.049
(10/24/20–10/29/20, phone)		ANES		**50.0**	25.5			**0.038**	0.040
		CPS		19.7	[c]			0.049	[c]

(continued)

Table 2. Continued.

Poll (dates, mode)	Bias in unweighted est. %	Population source	Percent bias removed: Weighted %	Percent bias removed: MUBP – Demos only %	Percent bias removed: MUBP – All vars. %	Pseudo-RMSE: Unweighted	Pseudo-RMSE: Weighted	Pseudo-RMSE: MUBP – Demos only	Pseudo-RMSE: MUBP – All vars.
ABC/WP—Wisconsin	−9.86	AP/NORC	6.9	21.3	65.1	0.100	0.094	0.087	0.038
(10/20/20–10/25/20, phone)		ANES		65.0	**122.2**			0.094	**0.029**
		CPS		41.0	[c]			0.087	[c]
NYT/Siena—Alaska	−4.45	AP/NORC	−6.9	**47.6**	8.8	0.052	0.057	0.052	0.050
(10/9/20–10/14/20, phone)		ANES		[c]	[c]			[c]	[c]
		CPS		19.6	[c]			**0.049**	[c]
NYT/Siena—Arizona	−9.26	AP/NORC	25.9	9.7	30.0	0.095	0.072	0.089	0.068
(9/15/20–9/20/20, phone)		ANES		0.1	**102.3**			0.102	**0.024**
		CPS		−33.4	[c]			0.130	[c]
ComRes[a]	0.84	BES 2015	523.1	**308.7**	427.9	**0.019**	0.040	0.028	0.034
(3/28/15–5/6/15, phone)									
ICM	−2.69	BES 2015	−24.6	−24.8	**16.7**	0.029	0.036	0.040	**0.025**
(4/10/15–5/6/15, phone)									
Ipsos Mori	−3.00	BES 2015	**13.3**	−91.3	−160.2	0.034	**0.032**	0.087	0.081
(4/12/15–5/6/15, phone)									
Survation[b]	−3.84	BES 2015	**−20.4**	−68.9	−48.9	**0.039**	0.047	0.071	0.058
(4/2/15–5/6/15, web/phone)									

(continued)

Table 2. Continued.

Poll (dates, mode)	Bias in unweighted est. %	Population source	Percent bias removed: Weighted %	Percent bias removed: MUBP – Demos only %	Percent bias removed: MUBP – All vars. %	Pseudo-RMSE: Unweighted	Pseudo-RMSE: Weighted	Pseudo-RMSE: MUBP – Demos only	Pseudo-RMSE: MUBP – All vars.
Opinium (4/2/15–5/5/15, web)	−3.91	BES 2015	**24.3**	4.0	−14.9	0.040	**0.031**	0.047	0.046
Populus (3/31/15–5/7/15, web)	−5.89	BES 2015	**17.1**	−40.5	7.3	0.059	**0.050**	0.113	0.055
YouGov (3/29/15–5/6/15, web)	−6.74	BES 2015	41.4	**48.0**	43.1	0.068	0.040	0.103	**0.039**
TNS (3/26/15–5/4/15, web)	−9.36	BES 2015	**50.0**	15.8	45.9	0.095	**0.050**	0.084	0.054
Panelbase (3/31/15–5/6/15, web)	−10.73	BES 2015	36.6	**112.3**	55.4	0.108	0.069	0.176	**0.051**
Number of "best" results			**7**	**7**	**4**	**4**	**4**	**2**	**8**
Number of "best" results, excluding Florida, North Carolina, ComRes, Survation (14 polls)			**4**	**6**	**4**	**0**	**4**	**2**	**8**

Note: Boldface indicates the "best" results for a given poll, that is, percent bias removed closest to 100 percent or smallest pseudo-RMSE.

[a]Indicates polls where the unweighted estimate was close to unbiased, and the percent bias removed metric ends up being extreme due to division by a number close to zero.

[b]Indicates polls where all estimators adjusted in the incorrect direction.

[c]Indicates not applicable (the November CPS supplement did not include measures of Party ID or political ideology; ANES sample too small for use with the Alaska poll).

Discussion

In this study, we evaluated the ability of the new and easy-to-compute MUBP measure to adjust for the selection bias in estimated proportions of likely voters who would vote for a particular candidate in 18 different polls from the United States and Great Britain. The new MUBP measure offers a key advantage over more commonly used weighting adjustments in that it has the potential to correct for selection bias in estimates that may be *non-ignorable*. Comparing the ability of this new measure to correct for selection bias to a standard weighting approach, we find evidence of improved inference in the majority of the polls analyzed. Importantly, in the few instances where standard weighting adjustments were the most effective (e.g., the TNS poll in Great Britain), at least one MUBP-adjusted estimate was very similar and had only slightly worse performance. Collectively, these results suggest that the MUBP approach has merit as a general adjustment technique in polling applications. We have provided annotated and easy-to-use R code on GitHub, along with examples from this study, for other researchers interested in using this measure (https://github.com/bradytwest/IndicesOfNISB).

While the improved adjustments of selected pre-election polling estimates based on this new measure generally seem promising, there is clearly still room for improvement in this approach. First, are there additional *relevant* predictor variables that we could include in the probit regression models used to compute our auxiliary proxies? We remind readers that we would need to be able to compute means, variances, and covariances for these predictor variables in the population of likely voters for the MUBP approach, meaning that these variables would also need to be available from a large probability sample of likely voters. We found that the inclusion of additional relevant correlates of voting intention in our models had a general tendency to improve the quality of the adjusted estimates, especially in terms of their variance, relative to estimates based only on the demographics that are typically used in weighting adjustments. This is certainly a direction for future work in this area that would benefit from additional expertise with respect to likely voter models and candidate preferences. This also speaks to the importance of measuring predictor variables related to political and voting preferences in large ongoing national studies based on probability samples. While the measurement of these "relevant" variables in large national samples would likely also improve weighting adjustments, we remind readers that the MUBP approach also allows for non-ignorable selection mechanisms.

In this application, both the ABC/WP and NYT/Siena polls measured good predictors of support for President Trump that served to improve the MUBP adjustments. The biserial correlations between the binary candidate preference indicator and the proxy constructed from available covariates, a key component of the MUBP estimation, ranged from 0.84 to 0.90 for the

ABC/WP polls and from 0.81 to 0.90 for the NYT/Siena polls when including all covariates, enabling relatively precise estimates even when assuming a non-ignorable selection mechanism. For the Great Britain polls, these correlations were lower even when including the 2010 vote variable, ranging from 0.66 to 0.80. Having access to strong auxiliary predictors of voting intention is important for the precise estimation of selection bias in context. This is immediately obvious when looking at the MUBP results that use sociodemographics only (no party identification, political ideology, or past voting history). The biserial correlations are much lower when excluding these relevant predictors (ABC/WP: 0.38–0.63; NYT/Siena: 0.40–0.73; Great Britain: 0.29–0.40), and this in turn makes the 95 percent credible intervals for the MUBP measure wider.

Second, the identification of a high-quality source of auxiliary data collected from the likely voter population that includes the same measures used to predict the response of interest in the selected sample is a nontrivial task. We considered data from the CPS, the ANES, the AP/NORC VoteCast study, and the British Election Study. For the US polls, the VoteCast study provided the largest samples for likely voters from each state, but was not entirely based on a probability sample. The CPS and ANES had smaller sample sizes from each state, and the CPS did not measure party preference or ideology. Other researchers applying the MUBP adjustment will need to spend time identifying the right source of auxiliary information[12] that will enable computation of accurate estimates of the population-level means, variances, and covariances for all of the common predictor variables that are needed for the MUBP approach.

An additional limitation of the MUBP measure is that the population aggregates calculated from probability-based population sources may themselves suffer from one of two types of bias: (1) potential (non-ignorable) selection bias, and (2) measurement error. The MUBP measure assumes that neither of these problems is significant, but any measurement error in the variables used to compute the population estimates or non-ignorable selection bias in the computed aggregates may have a significant effect on the MUBP calculations. In theory, all three population sources we used for the US polls provided summaries of the *same population*, and ideally the aggregate estimates (means, variables) from the different sources would be the

12. We also note that the ANES (with a response rate of about 61 percent) and the BES (with a response rate of about 56 percent) may not be viewed as ideal sources of auxiliary information for computing accurate estimates of the population parameters needed to apply the MUBP measure. Users of the MUBP measure should critically evaluate the large probability sample(s) (or source(s) of population information used to compute estimates of the population parameters for potential selection bias and other quality issues before computing the parameters. The MUBP adjustments will only be effective if the population parameters accurately reflect the population of interest.

same, producing identical MUBP-adjusted estimates regardless of the population source. However, the MUBP-adjusted estimates were sometimes quite different depending on the population source (e.g., Minnesota), which is an indication that these population sources may indeed suffer from bias, and we have no way to know which is the "right" population summary. We also note that the current implementation of the MUBP measure does not account for uncertainty in the population estimates of the means, variances, and covariances of the common auxiliary variables, treating the point estimates of these quantities as fixed. Given the differences among estimates using the different US population sources, accounting for this uncertainty would be a worthwhile direction for future methodological work.

Finally, we note that this study only presents 18 applications of the new MUBP measure to the problem of adjusting for potentially non-ignorable selection bias in estimates from pre-election polls. While these polls were from different countries and different states and used different modes, additional replications of this work would help further our understanding of how to optimize the ability of this promising MUBP measure to both quantify selection bias in pre-election polls and adjust estimates for that bias.

Supplementary Material

Supplementary Material may be found in the online version of this article: https://doi.org/10.1093/poq/nfad018.

Funding

This work was partially supported by the National Institutes of Health [1R21HD090366-01A1 to B.T.W.].

Acknowledgements

We thank the editors and the referees for extremely constructive suggestions that helped improve the manuscript significantly. We also thank Gary Langer, PARC, and the Roper Institute for making pre-election polling data from the United States available to the public, and Patrick Sturgis for providing us with access to the pre-election polling data from the 2015 general election in Great Britain. Finally, we thank the Survey Research Center for providing Brady West with generous sabbatical time that enabled this project to move forward.

Data Availability

Replication data and documentation are available at https://doi.org/10.17605/OSF.IO/PDBXH.

References

American Association for Public Opinion Research (AAPOR). 2017. "An Evaluation of 2016 Election Polls in the U.S." https://aapor.org/wp-content/uploads/2022/11/AAPOR-2016-Election-Polling-Report.pdf. Date accessed 1 July 2022.

Alcantara, Chris, Aaron Steckelberg, Darla Cameron, and Tim Meko. 2016. "Where Polling Underestimated Trump's Chances." *Washington Post*. https://www.washingtonpost.com/graphics/politics/2016-election/where-the-polls-got-it-wrong/. Date accessed 9 November 2016.

Andridge, Rebecca R., and Roderick J. A. Little. 2011. "Proxy Pattern-Mixture Analysis for Survey Nonresponse." *Journal of Official Statistics* 27:153–80. https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/proxy-pattern-mixture-analysis-for-survey-nonresponse.pdf.

———. 2020. "Proxy Pattern-Mixture Analysis for a Binary Variable Subject to Nonresponse." *Journal of Official Statistics* 36:703–28. https://doi.org/10.2478/jos-2020-0035.

Andridge, Rebecca R., and Katherine Jenny Thompson. 2015. "Using the Fraction of Missing Information to Identify Auxiliary Variables for Imputation Procedures via Proxy Pattern-Mixture Models." *International Statistical Review* 83:472–92. https://doi.org/10.1111/insr.12091.

Andridge, Rebecca R., Brady T. West, Roderick J. A. Little, Philip S. Boonstra, and Fernanda Alvarado-Leiton. 2019. "Indices of Non-Ignorable Selection Bias for Proportions Estimated from Non-Probability Samples." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 68:1465–83. https://doi.org/10.1111/rssc.12371.

Ansolabehere, Stephen, Bernard L. Fraga, and Brian F. Schaffner. 2021. "The CPS Voting and Registration Supplement Overstates Minority Turnout." *Journal of Politics* 84:1850–55. https://doi.org/10.1086/717260.

Bethlehem, Jelke. 2009. *Applied Survey Methods–A Statistical Approach*. Hoboken, NJ: John Wiley & Sons.

———. 2020. "Working with Response Probabilities." *Journal of Official Statistics* 36:647–74. https://doi.org/10.2478/jos-2020-0033.

Brick, J. Michael. 2013. "Unit Nonresponse and Weighting Adjustments: A Critical Review." *Journal of Official Statistics* 29:329–53. https://doi.org/10.2478/jos-2013-0026.

Chen, Jack K. T., Richard Valliant, and Michael R. Elliott. 2019. "Calibrating Non-Probability Surveys to Estimated Control Totals Using LASSO, with an Application to Political Polling." *Journal of the Royal Statistical Society Series C: Applied Statistics* 68:657–81. https://doi.org/10.1111/rssc.12327.

Clinton, Joshua, Jennifer Agiesta, Megan Brenan, Camille Burge, Marjorie Connelly, Ariel Edwards-Levy, Bernard Fraga, Emily Guskin, D. Sunshine Hillygus, Chris Jackson, Jeff Jones, Scott Keeter, Kabir Khanna, John Lapinski, Lydia Saad, Daron Shaw, Andrew Smith, David Wilson, and Christopher Wlezien. 2020. "Task Force on 2020 Pre-Election Polling: An Evaluation of the 2020 General Election Polls." AAPOR. https://www.aapor.org/Education-Resources/Reports/2020-Pre-Election-Polling-An-Evaluation-of-the-202.aspx. Date accessed 23 November 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* 86:247–69. https://doi.org/10.1093/poq/nfac011.

Crawford, Meghann, Don Levy, and April Backus. 2022. "The Anti-Establishment Voter." Paper presented at the 2022 Annual Conference of the American Association for Public Opinion Research (AAPOR), Chicago, IL, May 12.

Duncan, Pamela. 2016. "How the Pollsters Got It Wrong on the EU Referendum." *The Guardian*. https://www.theguardian.com/politics/2016/jun/24/how-eu-referendum-pollsters-wrong-opinion-predict-close. Date accessed 24 June 2016.

Dutwin, David, and Trent D. Buskirk. 2017. "Apples to Oranges or Gala Versus Golden Delicious? Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples." *Public Opinion Quarterly* 81:213–39. https://doi.org/10.1093/poq/nfw061.

Fieldhouse, E., J. Green, G. Evans, H. Schmitt, C. van der Eijk, J. Mellon, and C. Prosser. 2015. British Election Study, 2015: Face-to-Face Survey [computer file], November. https://www.britishelectionstudy.com/data-objects/cross-sectional-data/. Date accessed 1 July 2022.

Gelman, Andrew. 2012. "Statistics in a World Where Nothing Is Random." December 12. http://andrewgelman.com/2012/12/17/statistics-in-a-world-where-nothing-is-random/.

Gill, Jeff. 2014. *Bayesian Methods: A Social and Behavior Sciences Approach*, 3rd ed. Boca Raton, FL: Chapman and Hall/CRC Press.

Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70:646–75. https://doi.org/10.1093/poq/nfl033.

Haziza, David, and Jean-François Beaumont. 2017. "Construction of Weights in Surveys: A Review." *Statistical Science* 32:206–26. https://doi.org/10.1214/16-STS608.

Heckman, James J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5:475–92. https://www.nber.org/system/files/chapters/c10491/c10491.pdf.

Heeringa, Steven G., Brady T. West, and Patricia A. Berglund. 2017. *Applied Survey Data Analysis*, 2nd ed. Boca Raton, FL: Chapman and Hall/CRC Press.

Horvitz, Daniel G., and Donovan J. Thompson. 1952. "A Generalization of Sampling Without Replacement from a Finite Universe." *Journal of the American Statistical Association* 47: 663–85. https://doi.org/10.1080/01621459.1952.10483446.

Hosmer, David W., Stanley Lemeshow, and Rodney X. Sturdivant. 2013. *Applied Logistic Regression*, 3rd ed. Hoboken, NJ: John Wiley & Sons.

Kalton, Graham, and Ismael Flores-Cervantes. 2003. "Weighting Methods." *Journal of Official Statistics* 19:81–97. https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/weighting-methods.pdf.

Kennedy, Courtney, Mark Blumenthal, Scott Clement, Joshua D. Clinton, Claire Durand, Charles Franklin, Kyley McGeeney, Lee Miringoff, Kristen Olson, Douglas Rivers, Lydia Saad, G. Evans Witt, and Christopher Wlezien. 2018. "An Evaluation of the 2016 Election Polls in the United States." *Public Opinion Quarterly* 82:1–33. https://doi.org/10.1093/poq/nfx047.

Little, Roderick J. A., and Donald B. Rubin. 2019. *Statistical Analysis with Missing Data*, 3rd ed. Hoboken, NJ: John Wiley and Sons.

Little, Roderick J., and Sonya Vartivarian. 2005. "Does Weighting for Nonresponse Increase the Variance of Survey Means?" *Survey Methodology* 31:161–68. https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2005002/article/9046-eng.pdf.

Little, Roderick J., Brady T. West, Philip S. Boonstra, and Jingwei Hu. 2020. "Measures of the Degree of Departure from Ignorable Sample Selection." *Journal of Survey Statistics and Methodology* 8:932–64. https://doi.org/10.1093/jssam/smz023.

McAuliffe, Colin, Johannes Fischer, Charlotte Swasey, Jason Katz-Brown, Jason Ganz, Gustavo Sanchez, Sean McElwee. 2021. "Memo: 2020 Polling Retrospective." Data for Progress. https://www.dataforprogress.org/memos/2020-polling-retrospective. Date accessed 16 November 2021.

Parliament. House of Commons. 2015. "General Election 2015" (Number CBP7186). https://researchbriefings.files.parliament.uk/documents/CBP-7186/CBP-7186.pdf. Date accessed 8 July 2022.

Särndal, Carl-Erik, and Sixten Lundström. 2005. *Estimation in Surveys with Nonresponse*. Chichester, UK: John Wiley & Sons.

Silver, Nate. 2016. "Pollsters Probably Didn't Talk to Enough White Voters without College Degrees." *FiveThirtyEight*. https://fivethirtyeight.com/features/pollsters-probably-didnt-talk-to-enough-white-voters-without-college-degrees/. Date accessed 1 Deember 2016.

Sturgis, Patrick, Nick Baker, Mario Callegaro, Stephen Fisher, Jane Green, Will Jennings, Jouni Kuha, Ben Lauderdale, and Patten Smith. 2016. *Report of the Inquiry into the 2015 British General Election Opinion Polls*. London: Market Research Society and British Polling Council.

West, Brady T., Roderick J. Little, Rebecca R. Andridge, Philip S. Boonstra, Erin B. Ware, Anita Pandit, and Fernanda Alvarado-Leiton. 2021. "Assessing Selection Bias in Regression Coefficients Estimated from Nonprobability Samples with Applications to Genetics and Demographic Surveys." *The Annals of Applied Statistics* 15:1556–81. http://doi.org/10.1214/21-AOAS1453