Pollster problems in the 2016 US presidential election: vote intention, vote prediction

Natalie Jackson[1,*], Michael S. Lewis-Beck[2], Charles Tien[3]

[1] *Public Religion Research Institute, Washington, DC, USA*
[2] *Department of Political Science, University of Iowa, USA*
[3] *Department of Political Science, Hunter College & The Graduate Center, CUNY, USA*
*Corresponding author. E-mail: nmjb09@gmail.com

Abstract. In recent US presidential elections, there has been considerable focus on how well public opinion can forecast the outcome, and 2016 proved no exception. Pollsters and poll aggregators regularly offered numbers on the horse-race, usually pointing to a Clinton victory, which failed to occur. We argue that these polling assessments of support were misleading for at least two reasons. First, Trump voters were sorely underestimated, especially at the state level of polling. Second, and more broadly, we suggest that excessive reliance on non-probability sampling was at work. Here we present evidence to support our contention, ending with a plea for consideration of other methods of election forecasting that are not based on vote intention polls.

Keywords. Forecasting, polling, research methods, elections, voting.

INTRODUCTION

To understand voter choice in American presidential elections, we have come to rely heavily on public opinion surveys, whose questions help explain the electoral outcome. In recent elections, horserace polls – those which measure vote intention, the declaration that you will vote for the Democrat or the Republican, or perhaps a third party – have been explicitly used to predict the outcome of the election in advance in media forecast models, exacerbating the reliance on them for election prognostication. In 2016, national and state-level polls suggested rather strongly that Hillary Clinton would defeat Donald Trump to become the next president of the United States. When it became clear that Trump would instead win the Electoral College, a debate sparked: Why were such forecasts, based on a mountain of polls, incorrect? Was this a fundamental failure of polling, or an irresponsible over-reliance on them by forecasters and the media-punditry complex? Either way, since the media forecasts rely mostly on polls, any widespread polling error should generate considerable concern.

How serious were these apparent errors? Here we review the performance of the 2016 vote intention polls for president, looking at the national level, where polls performed reasonably well, before turning to the states,

where the 2016 errors seem particularly grave. We offer a theoretical explanation for this error rather than the commonly-cited sources of polling error, which focus on poll mode or bias. Our contention is that pre-election polls suffer from a more critical problem: they are trying to poll a population – voters in an upcoming election – which does not exist at the time of the poll. This assertion means that the polls are not representative of the population they are interpreted to measure even under the best circumstances, making it unsurprising that they sometimes fail spectacularly as prediction tools. Many pollsters have made this exact argument: Polls are a snapshot of what *could* happen at the time they are taken. We extend it further by adding the theoretical underpinnings of how polls fail to satisfy representative sample requirements.

We offer theoretical and practical support for this hypothesis and argue that because of the inability to sample from the population of actual voters, and the inability to quantify the error that stems from that problem, polls should not be relied upon as prediction tools. In fact, there is evidence that this type of prediction can be harmful to natural election processes by impacting turnout. By way of conclusion, we suggest prediction alternatives, turning the focus to modelling the Electoral College result with aggregate (national and state) structural forecasting models and survey-based citizen forecasting.

ERROR IN THE 2016 NATIONAL PRESIDENTIAL ELECTION POLLS

In the popular mind the notion that the polls failed to accurately predict the 2016 electoral outcome seems widespread. What did the publicized polls actually show voters? Let us work through an illustration where "civic-minded Jill" follows the news – the lead stories and the polls – to arrive at her own judgment about who is ahead, who is likely to win. She checks *RealClear-Politics* aggregates daily, since the average percentages from available recent polls are readily understood. She observes, across the course of the campaign (June 16 to November 8) that nearly all the 180 observations report a Democratic lead (in the national 4-way daily poll average; the exceptional days are July 29th and July 30th). It looks like a Clinton win to Jill, but she wants more data, knowing that *RealClearPolitics* is just one aggregator, and she knows others use somewhat different aggregation methods. So, she consults a "Custom Chart" put out by Huffington Post (*HuffPost Pollster*, November 1, 2017)[1] that

Tab. 1. Poll Aggregator Predictions of Popular Vote.

	Clinton (prediction & error from 48.1% actual vote)	Trump (prediction & error from 46.2% actual vote)
New York Times Upshot	45.4%, 2.7%	42.3%, 3.9%
FiveThirtyEight	48.5%, -0.4%	44.9%, 1.3%
RealClearPolitics average	45.4, 2.7	42.2%, 4.0%
The Huffington Post	45.8%, 2.3%	40.8%, 5.4%

looks at the five weeks of national polls taken before election week; it shows Clinton at 46.0 percent and Trump at 42.4 percent, for a 3.6 point lead. Then, a few days before the election, she focuses on the news from other aggregators as well, as illustrated in Table 1 with estimates from *Upshot*, *FiveThirtyEight*, *The Huffington Post*, and *Real-ClearPolitics*. These all show Clinton ahead (from 3.1 to 5.0 percentage points) over Trump.

Jill now has more confidence that it will be a Democratic win. However, she realizes that these aggregates can mask big differences, so she turns to individual, final national polls, to get a better feel for the margins. Jill considers all the available ones, eleven national "likely voter" polls administered in November, and reported in *RealClearPolitics* or *HuffPost Pollster*.[2] She observes, as in Table 2, that the Clinton share of the total vote is always estimated to be in the 40s; further, she calculates Clinton's median support registers 44 percentage points.

Jill wants to compare these numbers to those for Trump, so she examines his estimates from the same polls, as in Table 3. She notes that, except for one observation (from Reuters/Ipsos) his scores are also always in the 40s. Now she calculates the median, and finds it equals 43, which disquiets her, since that estimate falls so close to Clinton's median of 44. She seeks reassurance by looking at the margins of error (MoE) at the 95 percent confidence interval, which are reported in the surveys. These numbers tell her that each survey estimate, for Clinton or Trump, is accurate within 3 percentage points above or below the point estimate 95 percent of the time, suggesting that, after all, Clinton might not be in the lead. As an aid to her thinking, she resorts to the poll range for each candidate, finding that for Clinton it is (42 to 47), while for Trump it is (39 to 44). Over-

[1] http://elections.huffingtonpost.com/pollster/2016-general-election-trump-vs-clinton.

[2] http://www.realclearpolitics.com/epolls/2016/president/us/general_election_trump_vs_clinton_vs_johnson_vs_stein-5952.html; http://elections.huffingtonpost.com/pollster/2016-general-election-trump-vs-clinton.

Tab. 2. Final National Polls for Clinton.

Poll	MoE	Clinton poll estimate	Clinton poll error
ABC/Wash Post Tracking	+/-2.5	47	*
FOX News	+/-2.5	48	*
UPI/CVOTER	+/-2.5	49	*
Monmouth	+/-3.6	50	*
CBS News	+/-3.0	45	0.1
Bloomberg	+/-3.5	44	0.6
Rasmussen Reports	+/-2.5	45	0.6
McClatchy/Marist	+/-3.2	44	0.9
NBC News/Wall St. Jrnl	+/-2.7	44	1.4
IBD/TIPP Tracking	+/-3.1	43	2
Reuters/Ipsos	+/-2.3	42	3.8

* = Actual percent of total vote as of 12/2/2016 (48.1 Clinton, 46.2 Trump) is within the poll's margin of error.

Tab. 3. Final Polls for Trump.

Poll	MoE	Trump poll estimate	Trump poll error
FOX News	+/-2.5	44	*
IBD/TIPP Tracking	+/-3.1	45	*
McClatchy/Marist	+/-3.2	43	*
Monmouth	+/-3.6	44	*
UPI/CVOTER	+/-2.5	46	*
ABC/Wash Post Tracking	+/-2.5	43	0.7
Rasmussen Reports	+/-2.5	43	0.7
Bloomberg	+/-3.5	41	1.7
CBS News	+/-3.0	41	2.2
NBC News/Wall St. Jrnl	+/-2.7	40	3.5
Reuters/Ipsos	+/-2.3	39	4.9

* = Actual percent of total vote as of 12/2/2016 (48.1 Clinton, 46.2 Trump) is within the poll's margin of error.

all, this assessment strengthens her belief that Clinton is ahead, but not by as much as she thought.

Jill has studied a good deal of data, but at this point still has uncertainty about which way it is going to go. If she had to bet, she would bet Clinton, but without much conviction. Also, she knows she has not yet really considered polling data from the states. And, she has avoided the sticky problem that even a majority in the national popular vote share, as estimated from the national vote intention polls, does not necessarily make for a presidential winner, since that choice must be made by the Electoral College. So now she takes a serious look at the Electoral College forecasts of the leading media poll aggregators (*NYT, 538, HuffPost, PW, PEC, DK*), as presented by *Upshot* on their *New York Times* website.[3] All these aggregators, which do look at state polls as well, give Clinton a better than 70 percent chance of a majority electoral vote. Moreover, the *Daily Kos* (92 percent), *Huffington Post* (98 percent), and the Princeton Election Consortium (99 percent) all awarded Clinton certainties of victory exceeding 90 percent.[4] As the American Association for Public Opinion Research (AAPOR) sums it up: "However well-intentioned these predictions may have been, they helped crystalize the belief that Clinton was a shoo-in for president." (Kennedy *et al.* 2017, 4).

Jill takes all the foregoing information into account and concludes, like many other American voters, that Clinton will be the next president. As we now know, Clinton received 51.1 percent of the two-party popular vote, compared a 48.9 percent for Trump, for a difference of just 2.1 percentage points. By this metric, the national polls were reasonably accurate. However, she lost the Electoral College, 232 votes to 306 votes, and thus lost the race.

The foregoing pattern of errors and predictions tends to work against the conclusion that these polls, after all, functioned as they should. But, as Sean Trende (November 12, 2016, *RealClearPolitics*) put it: "The story of 2016 is not one of poll failure."[5] That is partly true: national polling error was larger in 2012 than in 2016, showing a very narrow Barack Obama win while he won by nearly four percentage points on Election Day. In 2016, national polls showed Clinton winning by 2-5 points, and she won by two points. Yet because we do not have President Hillary Clinton in office now, the 2016 polls are perceived in a worse light – whereas in 2012 pollsters were taking victory laps.

However, we suggest some qualification to that conclusion, even at the national level. As Martin *et al.* (2005) indicate, accuracy and bias are two important criteria for assessing polling quality. With respect to accuracy, even though national polls were reasonably close to the margin between Clinton and Trump, con-

---

[3] https://www.nytimes.com/interactive/2016/upshot/presidential-polls-forecast.html?_r=0.
[4] http://election.princeton.edu/2016/11/08/final-mode-projections-clinton-323-ev-51-di-senate-seats-gop-house/.

[5] http://www.realclearpolitics.com/articles/2016/11/12/it_wasnt_the_polls_that_missed_it_was_the_pundits_132333.html.

sider the individual estimates from the final national polls (recall Tables 2 and 3), where seven (for Clinton) or six (for Trump) of the eleven poll estimates fall outside the standard margin of error. Further, with respect to bias, almost all these polls (seven for Clinton, eleven for Trump) underestimated the final vote share of the candidates, indicating that third party candidates were overestimated. To say all national polls performed well is to ignore those which came to the right conclusion but with inaccurate estimates. Additionally, final national poll aggregators' estimates all had *A* scores, which measures bias and accuracy (Martin *et al.* 2005), between -0.01 and -0.03, indicating a small but systematic underestimate for Trump – even after accounting for the polls also underestimating Clinton. These patterns, detectable in the national polls, are even more obvious in the state polls, a topic to which we now turn.

ERROR IN THE 2016 STATE PRESIDENTIAL ELECTION POLLS

Our conclusion is not that different from the AAPOR conclusion, which is that despite the 2016 national polls being more accurate than the 2012 national polls, 2016 was marked by inaccurate results at the state level, particularly in a few states that proved critical to Trump's Electoral College victory (Kennedy *et al.* 2017, 2). These state-level errors led poll-based forecasters astray in their Electoral College predictions. The final state polls appear to have had an average positive Clinton bias of about five percentage points. As Linzer (2016) put it, "The Big Question" is "How uncertain should we have been about the polls to make 5 to 10 percentage point errors seem consistent – even minimally – with the data?"

Take a closer look at polling accuracy in the states. There were five states in which Clinton held poll leads but lost on Election Day: Florida, North Carolina, Michigan, Pennsylvania, and Wisconsin. We begin with the first two. Polls in Florida and North Carolina showed the race closing in the final week. In Florida, Trump narrowly led by 0.2% according to *RealClearPolitics,* Clinton was up 1.8 percent according to HuffPost Pollster and 0.6 percent according to *FiveThirtyEight*. *RealClearPolitics* also had Trump leading by 1 point in North Carolina*,* while Clinton was up 1.6 percent according to *HuffPost Pollster*, and 0.7 percent according to *FiveThirtyEight*. Trump won Florida by 1.2 points, and North Carolina by 3.7 points.

The bigger shocks were in the Rust Belt states of Michigan, Pennsylvania, and Wisconsin – states that

Obama had won handily in 2008 and 2012 and which were often referred to as Clinton's "blue wall" in the Midwest. That narrative was driven in part by relatively strong Clinton polls. For example, not a single poll taken in Wisconsin ever showed Trump ahead in the state; the modal poll had Clinton up by 6-8 points in the final weeks of the campaign. In Michigan, in the final week most polls showed Clinton up by 1-5 points. One survey from the Trafalgar Group showed Trump up by two points, but it seemed to be a conservative-leaning outlier from a Republican-affiliated landline-only automated pollster. Since landline-only polls skew toward older, more conservative respondents, it was rational to think that a Republican poll conducted this way might be doubly skewed to the right. In Pennsylvania, Clinton was up by about 2-4 points in most late campaign polls; the only poll to show Trump ahead was again from Trafalgar Group.

But the story of state-level polling error does not end with the five states that went in the opposite direction from what was expected. Trump's vote share was underestimated in more than 35 states, and in many cases by more than ten points. The figures below show how polling aggregates performed relative to actual outcomes, calculated by subtracting the actual result margin between Clinton and Trump from the poll's margin between Clinton and Trump: Poll (Clinton% - Trump%) – Actual (Clinton% - Trump%). Figure 1 (originally pub-
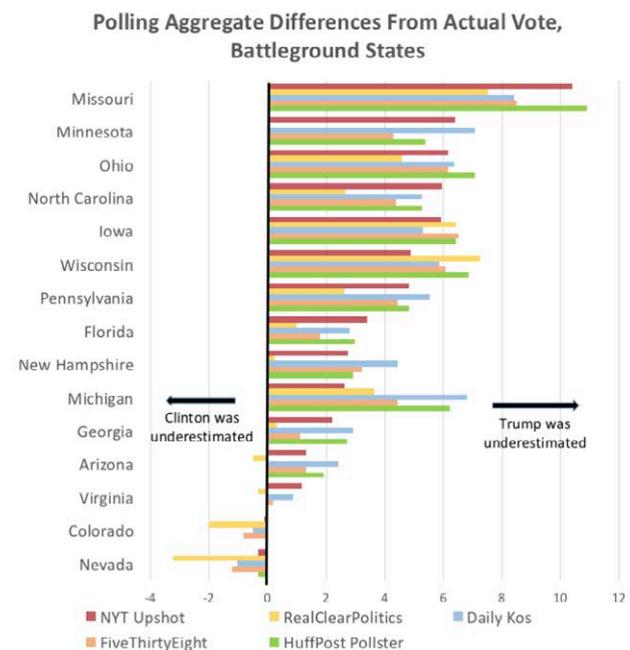

Fig. 1. Polling Aggregate Differences From Actual Vote, Battleground States.
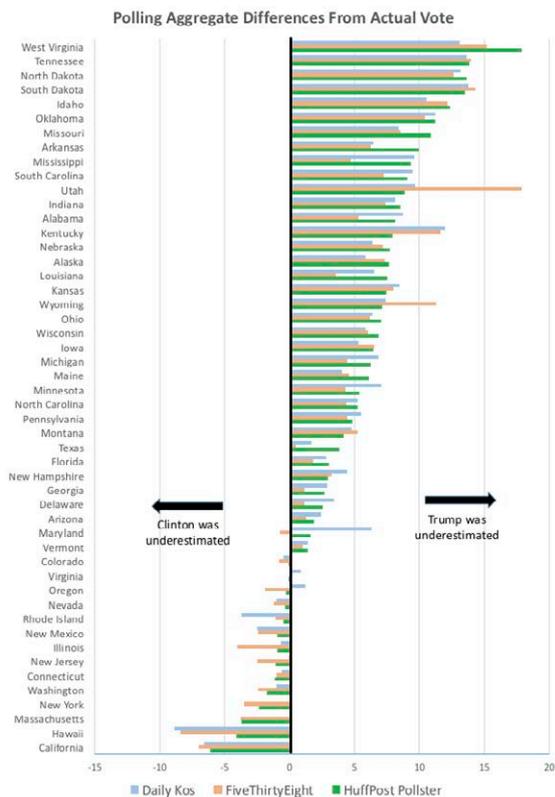
Fig. 2. Polling Aggregate Differences From Actual Vote, All States.

lished in Jackson 2016) shows the 15 most competitive states, where there were five aggregators active. Across the board – and including *RealClearPolitics*, whose national averages were nearly spot-on – Trump was systematically underestimated in 12 of the 15 states. The visual is even more striking among the aggregators who had all 50 states available (Figure 2, also originally published in Jackson 2016). The distribution is very lop-sided; Trump was underestimated in 35 states, while Clinton was underestimated in fewer than a dozen states. Average *A* scores (Martin *et al.* 2005) across all states for these three aggregators hovered around -0.04, again, demonstrating the consistent, lopsided bias in poll estimates.

The nature of the 2016 state-level polling errors – the vast majority of polls underestimated Trump regardless of any particular poll's characteristics – makes assessing the reasons for the misses difficult. The two most commonly-cited reasons for the 2016 polling misses are late shifts among voters, and overestimating college graduates, a weighting problem that was often not corrected (Kennedy *et al.*, 2017, p.3).

There is considerable support for last-minute shifts in vote intentions aiding Trump's side. According to

national exit polls, 13 percent of voters decided whom to vote for within the last week before Election Day (November 6), and 26 percent of voters decided in the last month. Those voters deciding in the last week broke 45-42 for Trump nationally, and voters deciding in the last month broke 48-40 for Trump nationally. Such late decisions might have been decisive in the three critical states of Michigan, Pennsylvania, and Wisconsin: in Michigan, those who decided in October broke 55-35 for Trump, in Pennsylvania the last-week deciders went 54-37 for Trump, and in Wisconsin it was 59-30 in Trump's favor among those who decided in the final week. Many of the last polls in Michigan, Pennsylvania, and Wisconsin were conducted a week or more prior to Election Day and could not possibly be expected to capture late deciders. But the polling industry cannot do anything about late deciders, except poll as close to Election Day as possible, and then communicate very clearly the risk and uncertainty that late deciders infuse into the estimates.

The second issue, the question of weighting to overcome bias, is closer to the root of the problem with pre-election polls, but only focusing on one weight – in this case, education – is only a small piece of the much larger issue, one which might also be responsible for making last-minute shifts seem substantial: All pre-election preference polls are attempting to sample from a population that does not yet exist. It is our contention that this missing population problem is at the root of pre-election polling inaccuracy. Pollsters simply cannot weight their way out of it, even under the best of circumstances.

THEORETICAL FAILURES OF PRE-ELECTION POLLS

In sampling theory, the population that an election poll wants to survey is people who voted in an election *which hasn't happened yet*. However, the fundamental admonition remains: sampling must be carried out, to the extent possible, following the scientific, mathematical methods of probability sampling laid down most fully by Kish (1965) and his disciples (Groves 1989; Weisberg 2005). Briefly stated, respondent selection must be made randomly (at every point where a selection is to be made), from a proper sampling frame, one targeting the relevant voting population. Following these principles has become expensive, and the problem of low response rates has not gone away. Indeed, it is our argument that it is impossible to get a representative sample of likely voters for a pre-election poll given the inability to get a sampling frame of actual voters before the election.

A true, probability pre-election poll, as defined by Kish's (1965) requirements, would have all of those who vote in the future election as the sampling frame. That sampling frame simply does not exist, forcing pollsters to substitute the frame of all Americans or registered voters. Thus, contrary to the long-standing assumption that Random Digit Dial (RDD) telephone polls are probability-based, we put them in the non-probability category, because there is no way to get a scientific random sample of Americans who will vote in the election prior to that election. That means a fundamental source of error in all the 2016 pre-presidential election polls stems from the fact that they employed *non-probability samples* rather than true probability sampling of future voters (Ansolabehere and Schaffner 2014; Brüggen, Van Den Brakel, Krosnick 2016; Shino and Martinez 2017).

We can see evidence of the inability to sample from the true population illustrated in the 2016 state polls. While the lop-sided nature of the poll underestimates is the first thing to stand out in Figure 2, equally important is the states with the largest errors. Polls underestimated Clinton most in California and Hawaii. Trump's largest underestimates were in West Virginia and Tennessee. The outcomes were never in question in any of those states, but the polling errors are very large. This points to an issue that has gone overlooked in election polling for decades: Polls that get the answer right, but still have considerable error, are considered "okay." Polls with small amounts of error that miss the result are considered bad. Not scrutinizing these errors in the right direction has cost us knowledge about polling errors. Pollsters estimate "likely voters," but often do not say how or why, or offer any discussion of how likely voter estimates are quite different from having a true probability sample of the correct population.

How Mode and Sampling Further Complicate Election Polling

The issue of sampling from the correct population is further exacerbated by mode and sampling problems that affect all polls. Pollsters in 2016 conducted both telephone and online polls. Consider telephone polls first, where the two main types are computer-assisted interviews (CATI) or those that are computer driven with no live interviewing – "robopolls." (The robopoll is quite inexpensive; however, it is illegal in the U.S. to use them on cell phones, so most pollsters using this method are either missing a substantial part of the population or use web-based methods to supplement the phone calls.) Historically, telephone samples come from random digit dialing (RDD), employing a computer algorithm for randomly selecting phone numbers that appear valid. Effectively, this defines the target population as all those who have (access to) a usable phone, so generating an obviously less than perfect list of voters. Moreover, the response rates with RDD have become perilously low, under ten percent of the numbers called (Keeter *et al.* 2017). Weights are used to account for nonresponse and make the survey representative of the U.S. adult population where it is not – although in well-designed samples these weights should be small. However, in this situation the sample's lack of representativeness of actual future voters is obvious: Not everyone reached by random selection will vote in an election, and there is no information beyond the respondents' own words to help inform whether they will vote. That survey respondents overestimate their likelihood to vote is a well-documented issue, even in the very high-quality and expensive American National Election Study (Jackman and Spahn 2019).

In an attempt to solve this inference problem, some pollsters turn to registered voter lists matched to phone numbers in order to generate their samples, making registered voters rather than American adults the population. These samples are closer to random samples of voters – where election pollsters want to be – than RDD samples and contain valuable information about registrants' past vote history. Nevertheless, these registered voter lists suffer from the exclusion of new registrants not on the rolls yet, and that not all sampled registered voters will cast a ballot. Some pollsters supplement these lists with additional sampling to address the issue of new registrants, but that brings back the issue of whether the respondent correctly indicates their likelihood to vote.

In contrast to telephone polling, online polling has found increasing use because of its low cost. Usually, the respondents are members of a panel, which serves as the database for subsequent surveys. The initial difficulty exists in recruiting the panel members, since an email list of all eligible voters does not exist. Most commonly, these web panels are made of respondents who have volunteered to participate in surveys via online advertisements. This is a means of self-selection whereby one learns of the panel, wants to be a member and can be, provided they satisfy the email invitation. While this volunteer method makes it easier to fill the panel, it still must wrestle (perhaps even more seriously) with the problem of opt-ins, who are not likely to be representative of the eligible voting population. The panel provider uses quotas and modelling to make any given survey appear representative of the U.S. adult population, and again the problem of the quality of the list, and the subsequent lack of represent-

ativeness of the sample drawn, surfaces. In a few cases, other means of recruitment are employed, such as RDD or address-based sampling using the United States Postal Service address file, which may lead to selection of a person in the household who answers the phone or responds to a mail request to join a panel. By this telephone method, a panel can be formed, and from it respondents randomly selected to participate in an election survey. Of course, even if the respondents are randomly selected from the panel, that does not mean they are representative of the population of future voters – again, the population does yet not exist.

In sum, most telephone and online election polls are based on a form of quotas, whether using them at the sampling stage or de facto forcing the data to fit quotas in the weighting stage. The respondents selected (even if eventually weighted), may not be truly representative of the socio-demographic sectors from which they were chosen, and almost certainly will not be representative of the yet-unknown population of actual voters. As Kalton (1983, 92) put it succinctly, regarding such methods "the chief consideration is to form groups that are internally homogenous in the availability of their members for interview," which makes them different from others in the category who were not sampled.

International polling experience is instructive here as well. In the 2015 United Kingdom General Election, the leading polls all showed a Labour-Conservative race too close to call, despite the final 6.5 percentage point lead of the Conservatives. A blue-ribbon committee appointed to investigate these discrepancies concluded that these erroneous results were the product of methods – essentially quota-style sampling – that rendered the surveys unrepresentative of the voting population (Sturgis *et al.* 2016).

British pollsters, in the run-up to the 2015 United Kingdom general election, all used quota sampling, applying weights known from population demographics (Sturgis *et al.* 2016). Following tradition, then, the UK quotas were fixed at the beginning of the sampling process. In contrast, common practice in the US basically fixes quotas at the end of the sampling process as weights, although some nonprobability online poll vendors do considerably more modelling and careful control of the sample than others. The essential disadvantage of either approach in nonprobability samples is that valid population parameter estimates, along with their probable error, are quite difficult to obtain (Freedman 2004). Therein lies the rub, as quota sampling, no matter how carefully designed and modelled, does not require that the respondents be selected randomly, and it certainly cannot select only those who will vote in the future.

WHAT CAN POLLSTERS DO?

The best solution is for pollsters to continue to refine their craft and adhere to the highest standards. That means leaning on probability samples wherever possible, and particularly encouraging more investment in high-quality polling at the state level – a solution also suggested in the AAPOR report (Kennedy *et al.* 2017). Still, estimating the voting population will remain a significant issue. There is no theoretically-sound substitute for sampling using the correct population and sampling frame that would satisfy Kish's (1965) requirements for probability sampling.

Pollsters, attempt to resolve the problem by using "likely voter" selection or modelling based on a respondent's self-reported propensity to vote and/or their voting history as available on voter registration lists. As demonstrated by polling misses, these methods are insufficient to fix the problem. In one high-profile case, Gallup, one of the oldest and most revered pollsters, mis-called the 2012 election in part due to their likely voter models underestimating the likelihood that voters who favored President Barack Obama would vote (Gallup 2013). After an investigation into the issues, one of the giants of the industry, which was among the first to conduct pre-election polling, decided to no longer release pre-election polling horserace numbers. While Gallup's decision is unusual, most pollsters have faced similar challenges in determining which of their respondents will vote. As Nate Cohn demonstrated in *The New York Times* Upshot (Cohn 2016), and a Pew Research report shows (Keeter and Igielnik 2016), the act of trying to predict who will vote has considerable impact on the poll's final numbers. Cohn showed how different assumptions lead to completely different outcomes in a 2016 Florida poll.

Transparency on likely voter selection should be demanded, and perhaps multiple numbers presented to demonstrate the uncertainty of those likely voter estimates. By presenting only one set of "likely voter" numbers, pollsters lean dangerously close to indicating that these numbers are predictions of the vote, rather than simple snapshots of one potential electorate. Reporting the survey's margin of error helps, but this figure is typically buried in fine print below much larger numbers championing the point estimates. And, even with margin of error, there are many other sources of potential polling error that are unaccounted for in this simple figure – in particular the error of misestimating who will vote, but also coverage error and measurement error.

Additionally, increasing response rates offer a source of hope for pollsters seeking to improve their performance. Public polls generally do not release response

rates, but a study conducted by Pew Research revealed their RDD response rates to be in the mid-single digits (Kennedy and Hartig 2019). Assuming that most polls show similar response rates, a few examples of higher response rate polls are instructive. In one case, the British Election Study (BES) and the British Social Attitudes Survey (BSA), results were better than the public polls. The BES and BSA employed classic multi-stage stratified probability sampling in their investigations of the 2015 general election, achieving response rates of 56% and 51% (AAPOR Response Rate 1), respectively; furthermore, the actual Conservative vote lead over Labour (of 6.5 percentage points), was estimated by these surveys almost exactly, with BES at seven points and BSA at six points, so offering a telling contrast to the gross errors made in the commercial polling exercises (Sturgis, *et al.*, 2016).

The American National Election Study (ANES) is one of the few surveys conducted face-to-face (with an online component) using address-based sampling, and also shows signs of being more accurate than public polls. The response rates (AAPOR Response Rate 1) were 44% and 50% for pre-election waves, and 84% and 90% for post-election waves.[6] With respect to the reported vote shares, it was 48.5% for Clinton and 44.3% for Trump, yielding an estimated difference of 4.2 points, not perilously far from the actual difference of 1.9 points (48.1% for Clinton - 46.2% for Trump), similar to estimates from other pre-election polls (see Tables 2 and 3) but without the systematic underestimates for one or both candidates from which several of those polls suffered. Of course, this accuracy was achieved at relatively great expense, and ANES still overestimates the proportion of Americans who will vote.

IMPLICATIONS FOR FORECASTERS AND COMMENTATORS

Most importantly, however, polls should not be used as the sole basis for election forecasts or assertions about who will win an election. Pollsters, to their credit, often remark that vote intention polls are snapshots of opinion now, not on election day. In other words, they are measures of conditions at a moment in time, not meant to be used as forecasts of the final electoral event. Nevertheless, political scientists, data journalists, and interested voters routinely turn to vote intention polls to make an educated guess about who will win. To quote the recent AAPOR report: "they attempt to predict a future event.

As the 2016 election proved, that can be a fraught exercise." (Kennedy 2017, 4).

Given the fact that polls will always have accuracy problems due to the absence of a population and sampling frame from which to draw a true probability sample, it is simply not advisable to use polls as the sole input in a forecast. Turnout changes in every election, and there is no way to predict the exact patterns beforehand, which means the error in the polls due to population mis-specification for any one election cannot be quantified. Polls, and poll-based forecasts will always suffer occasional failures. Political commentators should also heed these warnings. Even those who understand the possible errors in election polls and forecasts often seem to lean heavily on those results to fill airtime on television and to produce splashy content online.

Several countries go so far as to ban polls in a certain time period before the election, ranging from one day in France to as much as 15 days in Italy. This is due to the belief that these polls could change opinions or influence turnout, and some include campaigning blackouts as well. The U.S. has not taken this step, but the question of how polls impact vote choices has been heavily researched, concluding that there are some connections between polls and voting behaviour (e.g., Moy and Rinke 2012). One would imagine that forecasts have an even more substantial effect. Indeed, research has shown that both forecasters and commentators pushing the message that Clinton was winning handily in 2016 could have depressed turnout (Westwood *et al.* 2020). Any exercise which has the capacity to impact voter turnout is one that should be very carefully considered for its public benefit before proceeding with widespread attention. Media poll-based forecasts are certainly in this category, and we strongly urge caution in creating, using, or interpreting such forecasts.

VOTE INTENTION AS PREDICTION: FORECASTING ALTERNATIVES

Because of the challenges that polling, as a tool for forecasting elections, seems to increasingly face, we would like to conclude with some alternative strategies for election prediction, away from the dilemmas of vote intention polling. We turn explicitly to other scientific methods of election forecasting, namely structural models and citizen forecasting (see, respectively, the examples of Lewis-Beck and Tien, 2016b; and Lewis-Beck and Tien, 1999). The target of our exercise ends with a correct prediction of the Electoral College outcome. As we observed early on, "A common measure, share of popu-

---

[6] http://www.electionstudies.org/studypages/anes_timeseries_2016/anes_timeseries_2016_userguidecodebook.pdf.

lar vote, is rejected in favor of the tally that ultimately matters, the Electoral College vote share. Success or failure in that body, then, becomes the object of prediction, or forecasting." (Lewis-Beck and Rice 1992, 21). These two alternative forecast methods have traditionally focused on the popular vote, but if applied at the state level could be applied to the Electoral College.

In the election forecasting literature, structural models are a long-standing tradition. Typically, a single equation, specified according to well-established theories of voting behavior, finds application in prediction of the overall election outcome. Data are collected over a long time-series, with single forecasts made months before the election. Most of these models rely on some combination of objective economic indicators, survey data of presidential approval, and incumbent advantage (for examples see Abramowitz 2016, Lewis-Beck and Tien 2016a, Lockerbie 2016, Norpoth 2016). In 2016, these models generally performed very well, making forecasts within 2.5 percentage points of the popular vote outcome, at least 74 days before election day (see Campbell 2016 for a summary). These structural-model forecasts performed comparatively better than the likely voter polls taken in November where 13 of twenty-two November polls for Clinton and Trump were off by more than 2.5 percentage points (see Table 2 and Table 3 again). Nine of the eleven models correctly forecasted Clinton's popular vote win. They did not model the Electoral College.

Our parsimonious Political Economy model, with just two predictors (economic growth and presidential popularity) virtually hit the 2016 popular vote election outcome on the head, forecasting Clinton with 51.0 percent of the two-party vote (Lewis-Beck and Tien 2016b). One well-placed critique of this model, and other national structural models, comes from the fact that they do not directly estimate the Electoral College outcome. However, in practice, the two-party national popular vote, which the model forecasts, actually predicts the Electoral College voter share quite well, as a general rule. In Figure 3 we see the scatterplot, with the regression line of electoral vote on popular vote. Note that the 18 elections fall very close to the line, and the linear fit of the model is quite snug, at R-squared = .93. It correctly forecast all the ultimate winners of all but two of these presidential elections – 2000 and 2016. While not a bad track record in general (16 of 18), its miss in 2016 persuades us it is worth considering further the state level of analysis, where the decisions are made (Berry and Bickers 2012; Campbell, 1992; Holbrook and DeSart, 2003; Klarner 2012; Jerôme and Jerôme-Speziari 2016).

Last, but not least, we want to offer the alternative of citizen forecasting of US presidential elections. Look-

ing first at the national level, we have shown that citizens can be very good at predicting who will win U.S. presidential elections (Lewis-Beck and Tien 1999). When asked before the election who they thought would win, a majority of ANES respondents correctly predicted the outcome in nine of eleven elections between 1956 and 1996, missing only the close elections of 1960 and 1980. In an update of this citizen voter model, Murr, Stegmaier, and Lewis-Beck (2016) forecast that for 2016, Clinton would win 51.4 percent of the two-party vote, based on the opinion of those who had decided to vote. This result was extremely close to the 51.1 percent of the two-party vote that she received. Of course, citizens will use polls as part of the calculus for their forecast, but they will also consider an unknown number of other factors that polls alone do not include, such as economic conditions, what undecided or third party voters might actually do, and how late-breaking events might change the outcome. [Murr, Stegmaier, and Lewis-Beck (2020), have recently published a citizen forecasting paper for British general elections, showing the clearly superior performance of vote expectations over vote intentions, 1950-2017.]

Murr (2015) has applied the citizen forecasting idea to respondents in each state, to good effect. Taking the ANES data (through 2012), he broke out respondents by state, and examined their answers to the question: "Which candidate for President do you think will carry this state?" Murr (2015) assigned the winner of each state (as judged by the Republican or Democrat who received the most "will carry" predictions) its electoral votes, summing them in order to arrive at the overall Electoral College winner. In eight of the nine elections, voter expectations by state matched the real winner overall. Note that this approach seems especially promising as a survey method, one that works at the state
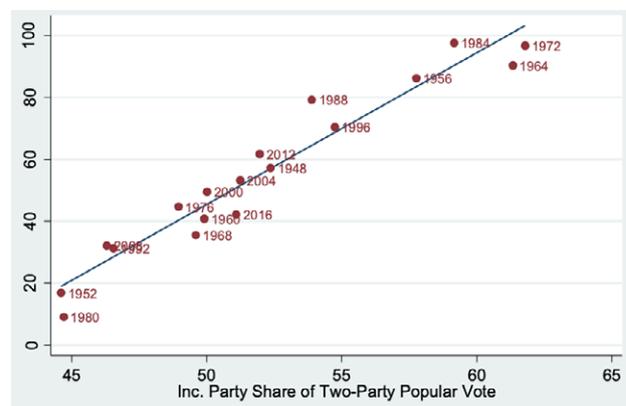

Fig. 3. Electoral College and Popular Votes, 1948-2016.

level. Finally, and importantly, the state subsets were not drawn to represent the states (rather, they were part of a very high-quality national random sample) but managed to work, drawing in practice on the "wisdom of the crowds" and in theory on Condorcet's Jury Theorem (Murr 2015). Clearly more work is needed to determine whether this method works at the state level in individual state polls of lesser quality than the ANES, but this analysis shows promising results. [It should be mentioned that Murr (2016) also applied the citizen forecasting strategy successfully to constituency results in the 2015 United Kingdom General Election.]

The relative success of these alternative methods of election forecasting at the national level, particularly in 2016, indicates that applying them to the state level and estimating Electoral College outcomes could be a substantial improvement over polls-only state-level forecasts. Indeed, using *only* vote intention polls to predict elections is an especially fraught exercise – one bordering on malpractice – given that there are other political and social factors that we know affect election outcomes. [An additional difficulty with the sole use of vote intention polls to forecast is deciding on the optimal lead time (Jennings, Lewis-Beck, and Wlezien, 2020).] Vote intention polls cannot possibly capture everything due to the unknown future population that pollsters are not able to sample. The result is that these polls often do not match outcomes, errors that become unnecessarily amplified in the context of vote intention polls-only forecasting. By combining these methods with more high-quality polls at the state level, we would gain much more insight into the possible Electoral College outcomes of a given presidential election.

CONCLUSION AND RECOMMENDATIONS

In sum, while the problem of trying to survey a population that does not yet exist offers some intractable complications for pre-election horserace polls, we do see a few reasonable approaches to improving polls and forecasts based on lessons learned from 2016 as well as research on other forecast methods. Pollsters and organizations sponsoring polls should primarily focus on obtaining the highest-quality samples possible, especially at the state level, even when that means investing more money into the process. There is no guarantee that high-quality polls will be completely accurate all the time – in fact, it is almost guaranteed that they will not be correct on some occasions – but high-quality data are preferred and much more likely to be correct than low-quality data. Additionally, more information could be

gleaned from polls, again, especially those at the state level, by adding a short question asking which candidate respondents expect to win the election. While survey time costs considerable money, this question would be very short and relatively cheap. This would bring considerable additional media attention to the poll and the pollster, particularly in battleground states, and therefore be a worthwhile addition.

Forecasters should be extremely wary of relying on polling data alone. Given the unsolvable problem of not having the correct population, relying on polls – or even incorporating other information but weighting it heavily toward the polls – is a misuse of polling data. Instead, structural forecasting models should be developed that move beyond the popular vote to estimating the Electoral College, and citizen forecasts (using the above-mentioned survey question on who will win the election) should be expanded to do the same. Since two of the last five presidential elections (2000 and 2016) have ended in a split between the popular vote and the Electoral College, it is critical to model the Electoral College if the goal is to accurately predict who will take office. These two additional techniques could then be combined with polling data – but notably weighted equally with the polls – to produce an estimate of which candidate might win the Electoral College. [Another possibility involves combination of vote intention with structural models, in an effort to produce 'synthetic' models that help control for the omitted variable problem (Dassonneville and Lewis-Beck, 2015).]

The ultimate lesson from 2016 extends beyond pollsters and forecasters, however, to commentators and any 'Jill' who consumes election polling and forecast information: be aware of the limitations of these data, and do not become overconfident in any outcome until the votes are counted. For everyone producing data and estimates, think carefully about the public good of the messages going out or any impact – intended or not – that your data might have on whether someone votes, who they vote for, and how they experience democracy.

REFERENCES

Abramowitz, Alan. 2016. "Will Time for Change Mean Time for Trump?" *PS: Political Science & Politics* *49*(4): 659-660. doi: 10.1017/S1049096516001268.

Ansolabehere, Stephen and Brian F. Schaffner. 2014. "Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison." *Political Analysis* *22*(3): 285-303.

Berry, Michael J. and Kenneth N. Bickers. "Forecasting the 2012 Presidential Election with State-Level Eco-

nomic Indicators." *PS: Political Science & Politics* 45: 669-674. doi: 10.1017/S1049096512000984.

Brüggen, E., Van Den Brakel, J., Krosnick, H. 2016. "Establishing the accuracy of online panels for survey research." Report 2016/4, *Statistics Netherlands*, The Hague, The Netherlands.

Campbell, James E. 1992. "Forecasting the Presidential Vote in the States." *American Journal of Political Science* 36: 386-407.

Campbell, James E. 2016. "Introduction." *PS: Political Science & Politics* 49(4), 649-654. doi:10.1017/S1049096516001591.

Campbell, James E., and Kenneth A. Wink. 1990. "Trial Heat Forecasts of the Presidential Vote." *American Politics Quarterly* 18: 251-269.

Cohn, Nate. 2016. "We Gave Four Good Pollsters the Same Raw Data. They Had Four Different Results." *The New York Times Upshot.* September 20, 2016. Accessed November 24, 2017. https://www.nytimes.com/interactive/2016/09/20/upshot/the-error-the-polling-world-rarely-talks-about.html.

Dassonneville, Ruth and Michael S. Lewis-Beck. 2015. "Comparative Election Forecasting: Further Insights from Synthetic Models." *Electoral Studies* 39: 275-283.

Freedman, David A. 2004. "Sampling" in *The Sage Encyclopedia of Social Science Research Methods*, Vol. 3, eds. Michael S. Lewis-Beck, Alan Bryman Tim Futing Liao. P. 986-990.

Gallup. 2013. "Gallup 2012 Presidential Election Polling Review." Accessed November 24, 2017. http://news.gallup.com/poll/162887/gallup-2012-presidential-election-polling-review.aspx

Groves, Robert M. 1989. *Survey Errors and Survey Costs.* New York: Wiley.

Holbrook, Thomas, and Jay DeSart. 2003. "State-wide Trial-Heat Polls and the 2000 Presidential Election: A Forecasting Model." *Social Science Quarterly* 84: 561-573.

Jackman, Simon and Bradley Spahn. 2019. "Why Does the American National Election Study Overestimate Voter Turnout?" *Political Analysis* 27: 193-207.

Jackson, Natalie. 2016. "Election Polls Underestimated Donald Trump In More Than 30 States." *Huffington Post.* December 23, 2016. Accessed October 18, 2017. https://www.huffingtonpost.com/entry/election-polls-donald-trump_us_585d40d7e4b0de3a08f504bd.

Jennings, Will, Michael S. Lewis-Beck, and Christopher Wlezien, 2020. "Election Forecasting: Too Far Out?" *International Journal of Forecasting*, 36(3) (2020P), pp. 949-962.

Jerôme, Bruno and Veronique Jerôme-Speziari. 2016. "State-Level Forecasts for the 2016 US Presidential Elections: Political Economy Model Predicts Hillary Clinton Victory." *PS: Political Science and Politics* 49(4), 680-686.

Kalton, Graham. 1983. *Introduction to Survey Sampling,* Chapter 11. Sage Publications, Beverly Hills, CA.

Keeter, Scott, and Nick Hatley, Courtney Kennedy, Arnold Lau. 2017. "What Low Response Rates Mean for Telephone Surveys." *PEW Research Center.* May 15, 2017. Accessed October 17, 2017. http://www.pewresearch.org/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/.

Keeter, Scott and Ruth Igielnik. 2016. "Can Likely Voter Models Be Improved?" *Pew Research Center.* January 7, 2016. Accessed November 24, 2017. http://www.pewresearch.org/2016/01/07/can-likely-voter-models-be-improved/.

Kennedy, Courtney, Mark Blumenthal, Scott Clement, Joshua D. Clinton, Claire Durand, Charles Franklin, Kyley McGeeney, Lee Miringoff, Kristen Olson, Doug Rivers, Lydia Saad, Evans Witt, and Christopher Wlezien. (Ad Hoc Committee on 2016 Election Polling). 2017. An Evaluation of the 2016 Elections in the United States. AAPOR.

Kennedy, Courtney, and Hannah Hartig. 2019. "Response Rates in Telephone Surveys Have Resumed their Decline." *Pew Research Center*, February 17, 2019. Accessed March 7, 2020. https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/

Kish, Leslie. 1965. *Survey Sampling*, New York, Wiley.

Klarner, Carl E. 2012. "State-Level Forecasts of the 2012 Federal and Gubernatorial Elections. " *PS: Political Science and Politics* 45 (October), 655-662.

Lewis-Beck, Michael S., and Tom W. Rice. 1992. *Forecasting Elections.* Washington, D.C: CQ Press.

Lewis-Beck, Michael S., and Mary Stegmaier. 2014. "US Presidential Election Forecasting." *PS: Political Science and Politics* 47(2): 284 – 288.

Lewis-Beck, Michael S., and Charles Tien. 2016a. "Election Forecasting: The Long View." *Oxford Handbooks Online.* Retrieved March 15, 2017. DOI: 10.1093/oxfordhb/9780199935307.013.92.

Lewis-Beck, Michael S., and Charles Tien. 2016b. "The Political Economy Model: 2016 US Election Forecasts." *PS: Political Science and Politics* 49: 661-663.

Lewis-Beck, Michael S., and Charles Tien. 2012. "Election Forecasting for Turbulent Times." *PS: Political Science and Politics* 45: 625-629.

Lewis-Beck, Michael S., and Charles Tien. 1999. "Voters as Forecasters: A Micromodel of Election Prediction." *International Journal of Forecasting* 15:175-184.

Linzer, Drew. 2016. "Poll Aggregation in the 2016 U.S. Presidential Election: Accuracy and Uncertainty."

Slides, presented December 6, 2016. Personal Communication.

Lockerbie, Brad. 2016. "Economic Pessimism and Political Punishment." *PS: Political Science & Politics, 49*(4), 673-676. doi:10.1017/S104909651600130X.

Martin, Elizabeth A., Michael W. Traugott, and Courtney. Kennedy. 2005. "A Review and Proposal for a New Measure of Poll Accuracy." *Public Opinion Quarterly* 69: 342-369.

Moy, Patricia, and Eike Mark Rinke. 2012. "Attitudinal and Behavioral Consequences of Published Opinion Polls." In *Opinion Polls and the Media: Reflecting and Shaping Public Opinion*, edited by Jesper Strömbäck and Christina Holtz-Bacha, 225–45. Basingstoke, UK: Palgrave Macmillan.

Murr, Andreas. 2015. "The Wisdom of Crowds: Applying Condorcet's Jury Theorem to Forecasting US Presidential Elections." *International Journal of Forecasting* 31: 916-929.

Murr, Andreas. 2016. "The Wisdom of the Crowds: What do Citizens Forecast for the 2015 British General Election?" *Electoral Studies* 41: 283-288.

Murr, Andreas, Mary Stegmaier, and Michael S. Lewis-Beck. 2016. "Using citizen forecasts we predict that with 362 electoral votes, Hillary Clinton will be the next president." In *The LSE US Centre's Daily Blog on American Politics and Policy*. http://bit.ly/2fgjRrW (November 3, 2016).

Murr, Andreas, Mary Stegmaier, and Michael S. Lewis-Beck. 2020, forthcoming. "Vote Expectations Versus Vote Intentions: Rival Forecasting Strategies," *British Journal of Political Science,* https://doi.org/10.1017/S007123419000061.

Norpoth, Helmut. 2016. "Primary Model Predicts Trump Victory." *PS: Political Science & Politics, 49*(4), 655-658. doi:10.1017/S1049096516001323.

Shino, Enrijeta and Martinez, Michael. 2017. "The Impact of Survey Model on Model Estimates of Voting Behavior." Paper presented at The Southern Political Science Association Meetings, New Orleans, LA, January, 2017.

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

Trende, Sean. 2016. "It Wasn't the Polls That Missed, It Was the Pundits." RealClearPolitics.com. November 12, 2016, *RealClearPolitics.*

Weisberg, Herbert F. 2005. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: University of Chicago Press.

Westwood, Sean Jeremy, Solomon Messing, and Yphtach Lelkes. Forthcoming 2020. "Projecting Confidence: How the Probabilistic Horse Race Confuses and Demobilizes the Public." *Journal of Politics.*