

# Assessing the 2020 Pennsylvania Election Polls

Berwood Yost



February 2021

## Table of Contents

Introduction .....	3
The Polls in Pennsylvania Performed Reasonably Well.....	3
Accuracy and Bias.....	3
Representational and Measurement Errors .....	4
Thinking Beyond the Horse Race.....	6
Poll Findings about Election Context in 2016 and 2020 .....	7
Getting More from Pre-Election Polling .....	7
Poll Results for Pennsylvania Polling Organizations.....	10
Bias Estimates for 2020 Pennsylvania Polling Organizations .....	10
Summarized Poll Measures for Selected Polling Organizations.....	11
Poll Accuracy and Election Proximity .....	12
References.....	13
Notes .....	14

## Introduction

Pennsylvania's 2020 election, which included millions of mail-in ballots, vast partisan differences in mail-in voting, and a delayed vote count, resulted in an outpouring of snap judgments that the Pennsylvania polls were off the mark in 2020. This headline the day after the election, *An embarrassing failure for election pollsters*, expressed what many were thinking and feeling in that moment.<sup>i</sup> These initial judgments about the 2020 polling "failure" remain mostly uncorrected despite a final outcome that now seems to align with the expectations created by the pre-election polls, at least in Pennsylvania. With the vote counting complete, what can we learn about the performance of the state polls in the 2020 election cycle?

## The Polls in Pennsylvania Performed Reasonably Well

### Accuracy and Bias

Comparing the final Pennsylvania poll estimates to the election results for polls conducted within three weeks of election day shows some positive results for the state's polling accuracy, and some areas of concern. Nearly half the polls produced in the final three weeks of the campaign produced biased estimates, meaning the poll systematically over- or under-estimated one party's share of the vote beyond its margin of error.<sup>ii</sup> Despite these biases, which for technical reasons may be overstated, four out of five polls accurately predicted a Biden victory in the state and the average candidate error was only about 1.5 points.<sup>iii</sup> The absolute error of the state's polls was about four points and the signed error was nearly three points, indicating that President Trump's final margins were underestimated by about three points in the state.

Comparatively, the 2020 Pennsylvania polls performed better than in the 2016 election, but worse than the 2008 and 2012 elections. Every state poll in 2008 correctly predicted an Obama win as did almost every poll in 2012. Every other measure of bias and error was lower in 2008 and 2012 than in 2016 or 2020. Perhaps the most interesting pattern evident in these polls is the consistent underestimate for the winners' shares of the vote (see Table 1). If the polls are viewed in terms of their aggregated predicted margin and the accompanying aggregated margin of error, the historical performance of the polls in the state is reasonable--in no race was the difference between the predicted poll margins and the actual election margins outside the average margin of sample error reported by the polls.<sup>iv</sup>

Table 1: Poll Performance in Presidential Elections in Pennsylvania, 2008 - 2020

Election Year	% Biased Candidate Error	Correct Winner	Absolute Error	Signed Error
2008	23%	100%	2.62	-1.18
2012	25%	94%	2.17	-1.18
2016	66%	7%	4.67	4.50
2020	48%	79%	3.95	2.97

A visual presentation of the polling errors by election captures the essential feature of the polling errors in the state over the past four election cycles: the 2008 and 2012 election polls had a slight bias in favor of the Republican candidates while the 2016 and 2020 polls had a larger bias in favor of the Democratic candidates. A curve centered over the vertical dashed line at zero in Figure 1 indicates no bias in the polling, while positive values suggest the polls favored Republicans and negative values suggest the polls favored Democrats.

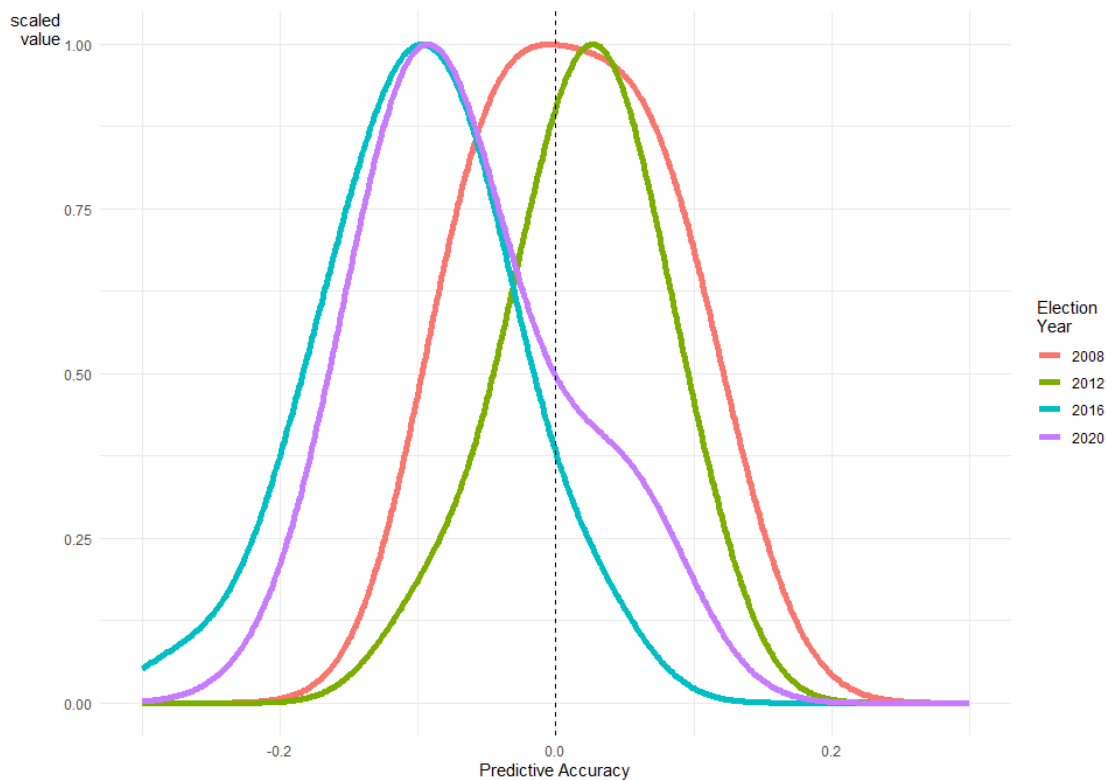


Figure 1. Distribution of Polling Errors in Pennsylvania Presidential Polls, 2008 - 2020

## Representational and Measurement Errors

Clearly, there have been more biased polls and larger polling misses in the state during the last two election cycles than in the preceding two election cycles. There are many possible causes for these biases, but the reasons largely fall into two broad categories: *representational errors* created by a sample that fails to look like the people who voted and *measurement errors* that arise from how questions are asked, understood, and answered.

Pollsters face significant issues recruiting survey participants, which requires them to adjust their samples (i.e, apply weighting procedures) so the final sample represents all voters. These weighting procedures introduce design effects that reduce a survey's effective sample size and increase its variability. Surveys that fail to incorporate design effects into their estimated variability (as reported by the margin of sampling error) are overstating their sample precision. Only eight of the 29 polls reported in the final three weeks of the campaign (28%) accounted for design effects in their estimated margins of sampling error.<sup>v</sup> This reporting failure means that sample precision is overstated. These

poor reporting practices compound the known biases associated with election polling (Shirani-Mehr et al. 2018).

The task of measuring poll accuracy is also complicated when voters' attitudes and preferences change--and the 2020 polls show that voters' intentions did change. While Joe Biden's monthly polling averages in Pennsylvania were seemingly stable, President Trump's monthly poll averages revealed a slow rise for the President: 50% to 43% (July), 49% to 44% (August), 50% to 44% (September), 50% to 45% (October), and 50% to 46% (November), Biden to Trump, respectively. The October Franklin & Marshall College Poll, completed nine days prior to the election, showed Biden leading Trump 50% to 44% among likely voters, which was similar to the polling average at the time.<sup>vi</sup> But this monthly stability masks how voters' preferences ebbed and flowed through the course of the campaign (see Figure 2). The monthly polling average and the general trend in the polls themselves suggests that President Trump was slowly closing Joe Biden's pre-election lead. It seems clear on its face that voters' attitudes changed throughout the campaign, making it likely that some voters would be undecided about their final choice until election day. The polling data in Figure 2 suggests that about five percent of voters were undecided on election day.

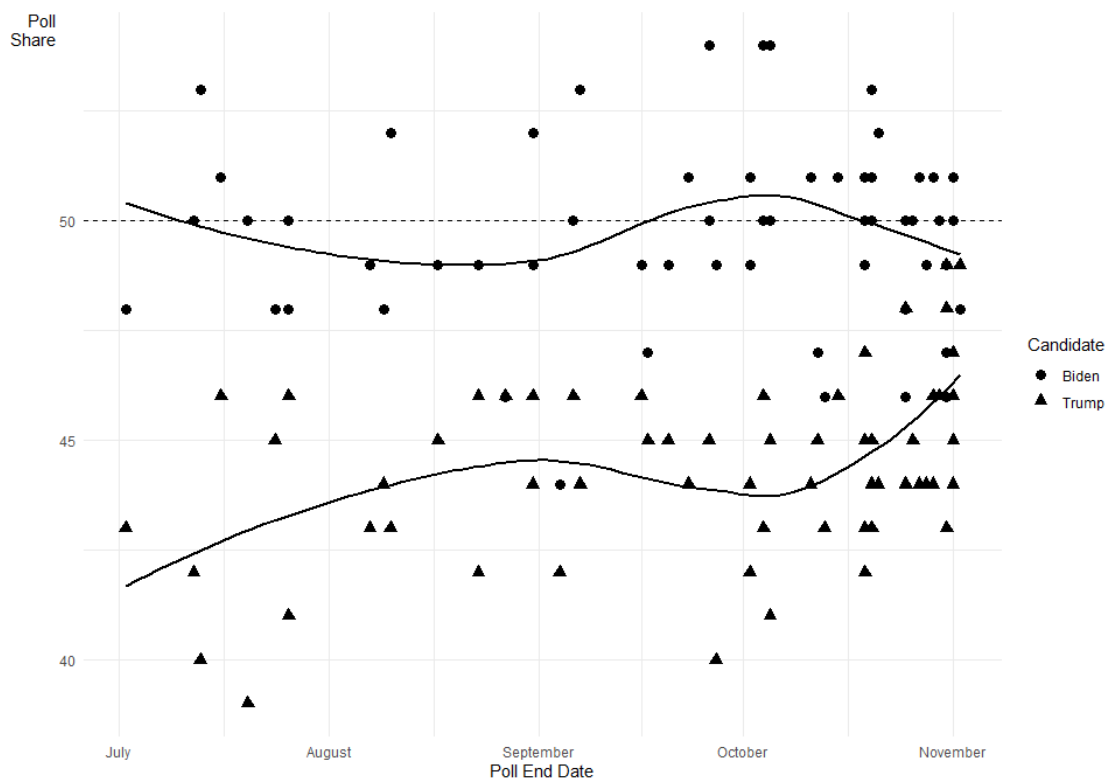


Figure 2. Pennsylvania Presidential Polls, 2020 (Created by author using data downloaded from [https://www.realclearpolitics.com/epolls/2020/president/pa/pennsylvania\\_trump\\_vs\\_biden-6861.html](https://www.realclearpolitics.com/epolls/2020/president/pa/pennsylvania_trump_vs_biden-6861.html) (accessed 1/5/2021).)

The 2020 race followed the same general pattern as the 2016 race in that the Democratic candidate had a relatively consistent advantage; the 2016 polling averages in Pennsylvania suggested that Hillary Clinton led Donald Trump throughout the entire fall campaign. Her monthly average lead was four points in July, eight points in August, five points in September, seven points in October, and three points in November. But unlike 2020, individual polls rarely showed either candidate with support from a majority of voters; the averages in the final week showed Clinton's expected vote share at 47% and Trump's expected vote share at 44%. Polling outcomes in 2016 also differed from those in 2020 by showing a sizable number of undecided and other voters that, given the tightening of the race over the final weeks, suggested movement toward Trump. The undecided and third-party candidates totaled nearly 20 points in July, 19 in August and September, 14 in October, and 11 in November. The rate of undecided and third party voters was much higher in 2016 than in 2012. Polls released in the final month of the 2012 campaign found only 5% of voters were undecided or planning to vote for a third-party candidate, as was the case in 2020.

Election poll post-mortems showed there were likely two main issues associated with the 2016 polling errors, one of which was a disproportionate swing in voters who changed their presidential preference. According to C. Kennedy et al. (2018), callback studies conducted in 2000, 2004, 2008, and 2012 showed that voters who changed their minds divided evenly between the Republican candidate and the Democratic candidate, but that in 2016, those who changed their minds voted for Trump by a 16-point margin.<sup>vii</sup>

Vote switching and late-deciding voters undoubtedly affected the poll outcomes in Pennsylvania in 2016. Post-election re-interviewing in Pennsylvania found that Donald Trump benefited from having a significant advantage among late-deciding voters. According to Yost, Redman, and Thompson (2017), more voters who supported Clinton pre-election moved away from her than moved away from Trump, and more voters who preferred neither candidate prior to the election voted for Trump than for Clinton.

This same pattern existed in 2020 according to post-election interviews, although the shifts were not as large as in 2016 because there were fewer late deciding or undecided voters this year.<sup>viii</sup> This data suggests that most of those who had chosen Trump (45% of all voters) or Biden (49% of all voters) before the election stayed consistent in their choice, but that President Trump had a slight advantage among the roughly five percent who were planning to vote for a third-party candidate or who were undecided when they were interviewed.

## Thinking Beyond the Horse Race

Trying to assess the polls' performance by looking at a single indicator is like judging the quality of a car by its paint color--easy to judge but meaningless until you understand the rest of the vehicle's components and how the owner plans to use it. It is a bit baffling that so many people focus on a single indicator to assess accuracy, in this instance the horse race question that measures candidate preference, when a good poll can provide essential context for understanding an election. Truth is, no one trying to forecast a future event is wise to rely on a single indicator to make their judgments, so why should polls be treated

any differently? Would we feel differently about the polls if we allowed ourselves to broaden our perspective and think about more than just the horse race? Even in a year like 2016 where polls seemed to have had performed poorly, would we see a different picture?

## Poll Findings about Election Context in 2016 and 2020

The 2016 polling was really clear about many features of that campaign:

- Both major party candidates were historically unpopular. This unpopularity meant that a large portion of voters would need to choose between two candidates they did not like. Our polling showed that 1 in 6 (16%) voters had an unfavorable opinion of both candidates.
- The polls clearly showed that there were historically large numbers of undecided and third-party voters.
- The polls also clearly showed that there was a strong desire for change-most citizens believed the country was on the wrong track, many citizens felt economically insecure, and views about President Obama's performance were shaping people's preferences. Part of this desire for change was evident in the consistent support Trump had from white working class voters, which was a mainstay of poll reporting in 2016.
- Finally, the polls clearly showed that the race was tightening as election day approached. This tightening sped up dramatically after James Comey's press conference that re-opened the FBI investigation into Hillary Clinton's emails.

Even though the 2020 polls were mostly correct that Joe Biden would carry the state, they also offered some other important findings:

- Attitudes about the state's direction changed negatively after growing satisfaction during the first three years of the Trump administration, and only one in three (31%) voters in Pennsylvania believed the country was headed in the right direction, similar to October 2016.
- President Trump was personally unpopular, although the President's job approval rating made him competitive.
- Joe Biden was doing better in the counties Hillary Clinton won in 2016 than President Trump was doing in the counties he won that year.
- The President's supporters were enthusiastic about supporting his candidacy, while Biden's supporters were more likely to say they were voting against the President.
- More state voters expected to cast their ballot in-person (53%) than by mail (44%), but there were strong partisan differences in these intentions.

These and other indicators provided important information and a relatively clear understanding of the electoral context in both election cycles.

## Getting More from Pre-Election Polling

Polling firms have an obligation to do good work, but measuring the "accuracy" of election polls is a more complicated notion than conversations about whether "the polls got it right" or whether "the polls can be trusted" allow. The complications are many: What's the right

way of measuring accuracy?<sup>ix</sup> How close to the final margin does a poll need to be to be considered accurate; is being within the poll's margin of error close enough? What's the lifespan of a polling estimate; how close to election day should a poll be conducted for it to count as a prediction?

At its most basic, reliable and accurate survey results come from a pollster's ability to find a good list from which to draw a representative sample, their ability to encourage those sampled to participate, and their skill at asking fair questions that are understood in the same way by each participant. These technically challenging tasks are each within a pollster's control.

What is beyond a pollster's control, and what many people overlook when talking about poll accuracy, is the role of shifting voter attitudes. A question that asks how someone intends to vote is measuring a **current belief** and **not an actual behavior**. How wise is the assumption of stable voter attitudes when the major-party candidates are bombarding voters with messages? In the 2020 election, the presidential candidates spent nearly \$100 million on advertising in Pennsylvania between July and November; is it reasonable to expect that campaign spending, media coverage, and the public conversations these encourage produce unchanging, immobile attitudes about turnout and preference?

This assumption is further strained in a state like Pennsylvania that is hotly contested and evenly divided. In Pennsylvania, only a small share of voters need to change their minds about turning out to vote or about who they intend to support to make a poll that was accurate when conducted look inaccurate on election day. There is ample evidence that attitudes about voting and about vote preference change throughout the course of the campaign, even if those preferences become a bit more stable in the closing weeks of a campaign (Shirani-Mehr et al. 2018).<sup>x</sup>

Assessing the "accuracy" of polling since its widespread use in the 1930s reveals a pretty clear pattern: the accuracy of national polls has steadily improved over time and the polls are not biased toward either party. For the 21 elections between 1936 and 2016, the polls have overestimated the Republican share of the vote 10 times and overestimated the Democratic share of the vote 11 times (C. Kennedy et al. 2018). This is true for state polls, too. Analyses of state polls from 1968 to 2016 suggests that the state polling was about as accurate as it usually is.<sup>xi</sup>

Questions about the accuracy and value of polls have been with us as long as modern polling, which started in the 1936 election. As the 50th anniversary of the 1936 election approached, people were still asking the question that occupies us today--are polls accurate? (Roper 1984) The answer then, as it is now, was "Yes and no." There are limits to what polls are capable of telling us and those limits are defined by the choices that pollsters make when they design their polls, the type of topics they ask about, and by the other limitations that all polls share.<sup>xii</sup> Discussions about poll accuracy are perennial.

If there is anything to take away from recent elections, it is that everyone should be more circumspect about making predictions based on a single indicator of who is ahead, particularly when there is so much other data we can use to tell the story. This means not only polling data, but other indicators that might tell us what's happening. Leading up to



the 2020 election, Franklin & Marshall College made it a point to not only discuss the polling indicators we thought were important, but to also present data about unemployment rates, campaign spending, COVID deaths, early voting, and changes in voter registration as supplemental indicators worth thinking about.<sup>xiii</sup> It also means pollsters must do a better job of reporting on the uncertainty of their estimates and discussing other potential sources of error beyond sampling, including the tendency for polls to share similar biases in any given election cycle.

This also means we should rely a bit less on poll aggregators. Poll aggregators have become popular because they make it easy for those interested in the campaigns to keep score, but these sites also over-simplify our understanding of the race and of the methods used to understand it. Panagopoulos, Endres, and Weinschenk (2018) notes that poll aggregators' projections in 2016 allowed for too little uncertainty about the election outcomes by producing lopsided projections that were inconsistent with any individual poll. These sites also encourage the reporting of poll data that is consistent with the polling averages. This "herding" means that some pollsters self-censor their results so they are in line with the poll averages. An analysis of the 2019 Australian election suggests that at least one firm did not report their results because they were out of line with other polls (Pennay et al. 2020).

In the end, those who look at polls will make their own judgments about the polls' performance. Undoubtedly, many of these assessments will be motivated more by partisan and ideological than methodological or performative criteria (Kuru, Pasek, and Traugott 2017). But I'd argue the science of polling has proven itself time and again. In the long run, polling has more often advanced our understanding than misled us, including in 2016 and 2020. The task for everyone who understands those facts is to do more to make sure people understand that polling is a helpful tool for reducing uncertainty, not eliminating it.

## Poll Results for Pennsylvania Polling Organizations

### Bias Estimates for 2020 Pennsylvania Polling Organizations

Figure 3 presents the estimated bias by pollster for the 2020 election in Pennsylvania. The figure includes the estimated bias along with error bars that display the variability of the bias estimate. Biased estimates are presented in red. The error bars of a biased poll do not overlap with zero. Some pollsters produced more than one poll during the final three weeks of the election and the bias estimates and standard errors for all polls appear on a single row. As noted earlier, 14 of these 29 polls were biased.

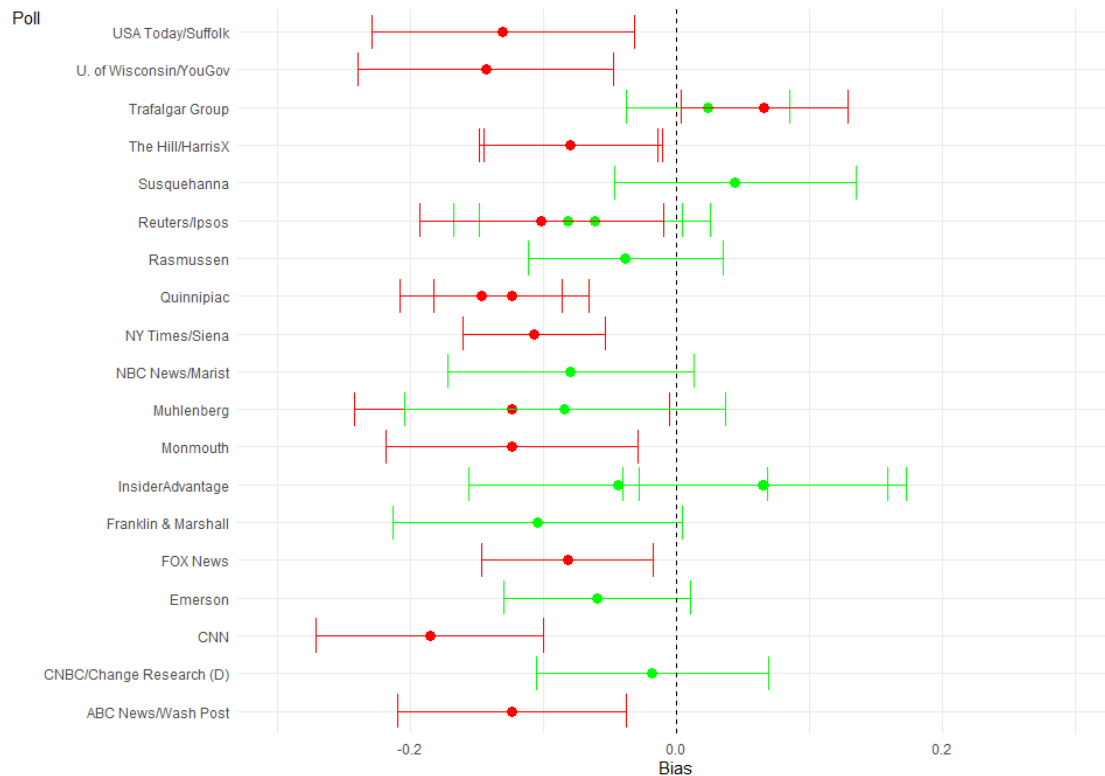


Figure 3. Poll Bias in Pennsylvania Presidential Polls, 2020

## Summarized Poll Measures for Selected Polling Organizations

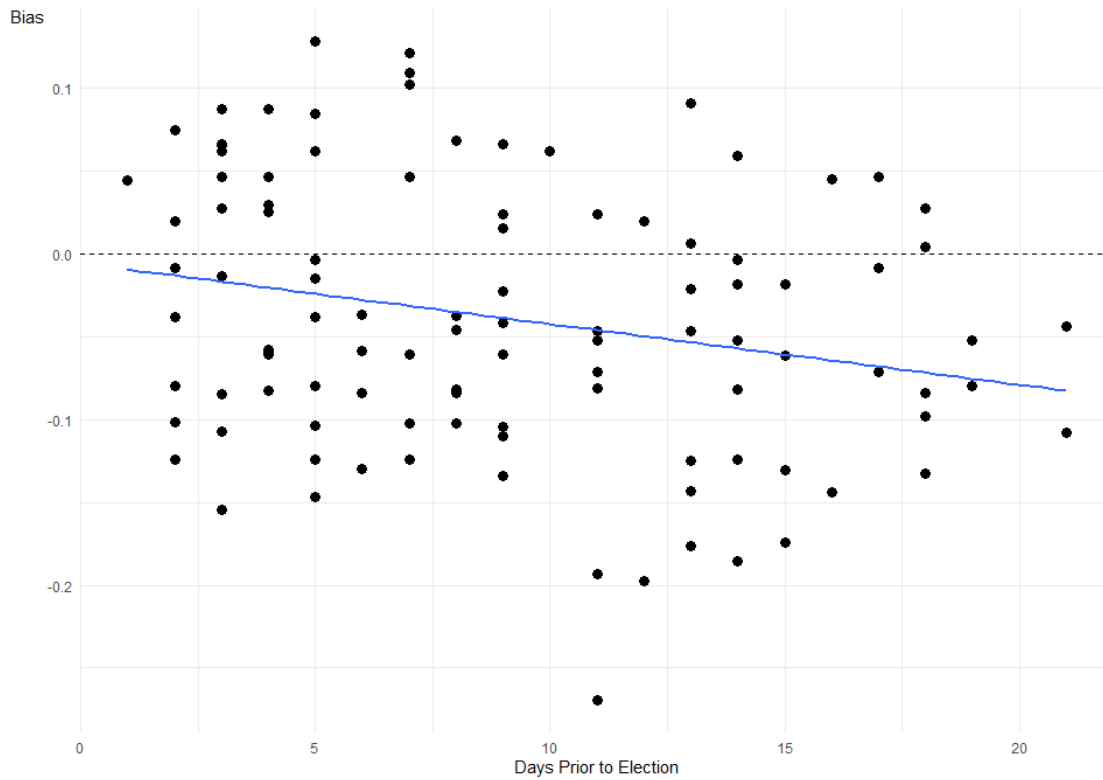
Seven organizations have released polls in at least three of the last four presidential elections in the state. The multiple measures of accuracy discussed in this paper offer an opportunity for most organizations to make claims about their "accuracy." A few notable findings: Quinnipiac has produced the largest proportion of polls that are biased and Susquehanna has called the correct winner in only one of these four elections (2016). The release schedules that these organizations follow is also notable: all of the polling organizations with the exception of Franklin & Marshall College conduct their final polls in the last week before the election.

Table 2: Pollster Performance in Pennsylvania, 2008 - 2020

<b>Pollster</b>	<b>% Biased</b>	<b>Candidate Error</b>	<b>Correct Winner</b>	<b>Beyond MoE</b>	<b>Signed Error</b>	<b>R Error</b>	<b>Polis</b>
Franklin & Marshall	0.25	2.21	0.75	0.25	4.42	-0.03	4
Muhlenberg	0.40	0.90	0.80	0.00	1.81	-0.01	10
Public Policy Polling	0.33	0.48	0.67	0.00	0.97	-0.01	3
Quinnipiac	0.67	1.88	0.83	0.33	3.77	-0.02	6
Rasmussen	0.25	0.04	0.75	0.00	0.08	0.00	12
Susquehanna	0.25	-0.91	0.25	0.00	-1.83	0.01	4
YouGov	0.25	1.32	0.50	0.25	2.65	-0.02	4

## Poll Accuracy and Election Proximity

Polls conducted closer to election day tend to exhibit less bias, and the Pennsylvania polls examined here show that same pattern with average bias increasing about .004 points each day prior to the election.<sup>xiv</sup>



*Figure 4. Poll Bias and Poll Interviewing Dates Pennsylvania Presidential Polls, 2008 – 2020*

## References

- Kennedy, Courtney, Mark Blumenthal, Scott Clement, Joshua Clinton, Claire Durand, Charles Franklin, Kyley McGeeney, et al. 2018. "An Evaluation of the 2016 Election Polls in the United States." *Public Opinion Quarterly* 82 (1): 1–33. doi:[10.1093/poq/nfx047](https://doi.org/10.1093/poq/nfx047).
- Kuru, Ozan, Josh Pasek, and Michael Traugott. 2017. "Motivated Reasoning in the Perceived Credibility of Public Opinion Polls." *Public Opinion Quarterly* 81 (2): 422–46. doi:[10.1093/poq/nfx018](https://doi.org/10.1093/poq/nfx018).
- Martin, Traugott, and Kennedy. 2005. "A Review and Proposal for a New Measure of Poll Accuracy." *Public Opinion Quarterly* 69 (3): 342–69.
- Panagopoulos, Costas, Kyle Endres, and Aaron C. Weinschenk. 2018. "Preelection Poll Accuracy and Bias in the 2016 U.S. General Elections." *Journal of Elections, Public Opinion and Parties* 28 (2): 157–72. doi:[10.1080/17457289.2018.1441850](https://doi.org/10.1080/17457289.2018.1441850).
- Pennay, Darren, Phil Hughes, Dina Neiger, Jill Sheppard, and John Stirton. 2020. "Report of the Inquiry into the Performance of the Opinion Polls at the 2019 Australian Federal Election." Association of Market; Social Research Organizations; the Statistical Society of Australia. [https://www.amsro.com.au/wp-content/uploads/2020/11/Inquiry\\_into\\_the\\_Performance\\_of\\_the\\_Opinion\\_Polls\\_at\\_the\\_2019\\_Australian\\_Federal-Election-Final\\_report.pdf](https://www.amsro.com.au/wp-content/uploads/2020/11/Inquiry_into_the_Performance_of_the_Opinion_Polls_at_the_2019_Australian_Federal-Election-Final_report.pdf).
- Roper, Burns. 1984. "Are Polls Accurate?" *The Annals of the American Academy of Political and Social Science* 472: 24–34. <https://www.jstor.org/stable/1043880>.
- Shirani-Mehr, Houshmand, David Rothschild, Sharad Goel, and Andrew Gelman. 2018. "Disentangling Bias and Variance in Election Polls." *Journal of the American Statistical Association* 113 (522): 607–14.
- Yost, Berwood, Jacqueline Redman, and Scottie Thompson. 2017. "The 2016 Pennsylvania Presidential and Us Senate Elections: Breaking Pennsylvania's Electoral Habits." *COMMONWEALTH: A Journal of Political Science* 19 (2): 3–26. doi:<http://dx.doi.org/10.15367/com.v19i2.152>.

## Notes

---

<sup>i</sup> <https://www.inquirer.com/opinion/commentary/election-polling-failure-wrong-trump-biden-20201104.html>.

<sup>ii</sup> The measure of predictive accuracy  $A$  developed by (Martin, Traugott, and Kennedy 2005) that compares the ratio of preferences for the major party candidates to the ratio of the final vote tally for each is used to measure accuracy and bias. The natural log of this odds ratio ( $\ln$  odds) is used because of its favorable statistical properties and the ease of calculating confidence intervals for each estimate. The confidence interval for a poll that reasonably predicts the final outcome of the election will overlap zero. Donald Trump's poll preference and vote total are the numerators in all the calculated ratios, which means a negative value for the  $\ln$  odds is an overestimate in favor of Joe Biden and a positive value is an overestimate in favor of President Trump. The 2020 polling results used in this analysis were taken from the RealClearPolitics website and included polls released in the three weeks prior to election day. For the 2020 calculations, an effective sample size calculation was used to estimate the poll variance since the reported margin of error for many polls reflected design effects that reflect more variability than the reported sample size. The polls for the 2008, 2012, and 2016 elections were downloaded from 538. For the calculations prior to 2020 and consistent with the  $A$  methodology, the Republican candidate is used in the numerators. In all years, the sample sizes used to estimate variance are probably too large since they tend to include all respondents and not just those who had a preference for the major party candidates.

<sup>iii</sup> Candidate error is a method used by the National Council on Public Polls (NCPP) to determine how closely published polls match the final election returns. The candidate error is one-half the error of the difference between the top two candidates. For example, the Rasmussen survey released two days prior to the election showed that President Trump trailed Joe Biden 50% to 47%, a three-point gap. Subtracting the estimate (3) from the actual result (1.2) shows the poll was off by 1.8 points. Dividing this result by two produced the final candidate error estimate of 0.9 for the survey.

<sup>iv</sup> The actual vote margins in the state were 10.4 points in 2008, 5.4 points in 2012, 0.7 points in 2016, and 1.2 points in 2020.

<sup>v</sup> CNN, Franklin & Marshall College, Marist, Muhlenburg, NY Times/Siena, and Reuters/Ipsos appear to have incorporated design effects into their margin of error estimates.

<sup>vi</sup> Source: RealClearPolitics, accessed (1/5/2021)

<sup>vii</sup> according to C. Kennedy et al. (2018), that is more than double the second largest margin in the time series, with was a seven point advantage among switchers for George W. Bush in 2000.

<sup>viii</sup> The Center for Opinion Research re-contacted all the respondents it interviewed prior to the election during the month of December, 2020. Re-interviews were completed with 1,093 respondents. More details about this study will be made available in the coming months.

<sup>ix</sup> For example, three recent papers assessing poll accuracy by Shirani-Mehr et al. (2018), Panagopoulos, Endres, and Weinschenk (2018), and C. Kennedy et al. (2018) each use different measures of accuracy based on polls conducted within three weeks, one week, and two weeks, respectively, of election day.

<sup>x</sup> the github repo for Shirani-Mehr et al. (2018)'s analysis can be found at <https://github.com/stanford-policylab/polling-errors>.

<sup>xi</sup> see <https://fivethirtyeight.com/features/the-real-story-of-2016/>.

<sup>xii</sup> Roper (1984) specifically mentioned distinctions between factual questions, opinion questions and future intentions by saying, "*Surveys that measure facts, (such as habits or ownership,) tend to be more accurate than surveys that measure opinion, and surveys that measure opinion are, in turn, more accurate than surveys that attempt to measure future intent.*"

<sup>xiii</sup> see F&M's pre election dashboard <https://www.fandm.edu/fandmpoll/fm-collegepoll-indicatordashboard>

<sup>xiv</sup> The regression statistics predicting bias ( $A$ ) from interviewing proximity to election day are  $F(1)=6.020$ ,  $p=.0158$ . A similar regression predicting candidate error from proximity to election day show the same significant result,  $F(1)=1949$ ,  $p. < .001$ , and reduces candidate error by .04 each day.