

FIRST THOUGHTS ON POLLING PROBLEMS IN THE 2016 U.S. ELECTIONS

DOUGLAS RIVERS

We were, like most people, surprised at the outcome of the 2016 U.S. elections. We conducted a wide array of polls. These included our weekly polling for *The Economist*, 111 Battleground polls for CBS News, and a daily election model (developed by Ben Lauderdale of the London School of Economics) that produced estimates for each state as well as the national popular vote. Our results were consistent—each showed an approximate four point lead for Hillary Clinton over Donald Trump. In the final *Economist* survey, Clinton led Trump 45.1 to 41.5 percent among likely voters. The final daily model estimate (using data collected through Sunday morning) had a lead of 3.8 percent. In short, we expected Hillary Clinton to win by about four points.

Most other polls over the weekend showed a similar margin for Clinton. All of the leading media polls—ABC/*Washington Post*, CBS News, FOX News, NBC/*Wall Street Journal*—had Clinton winning by four points in their final polls. Nate Silver’s FiveThirtyEight Election Forecast predicted that Clinton would win by 3.6 percent and other poll aggregators had similar estimates. There were

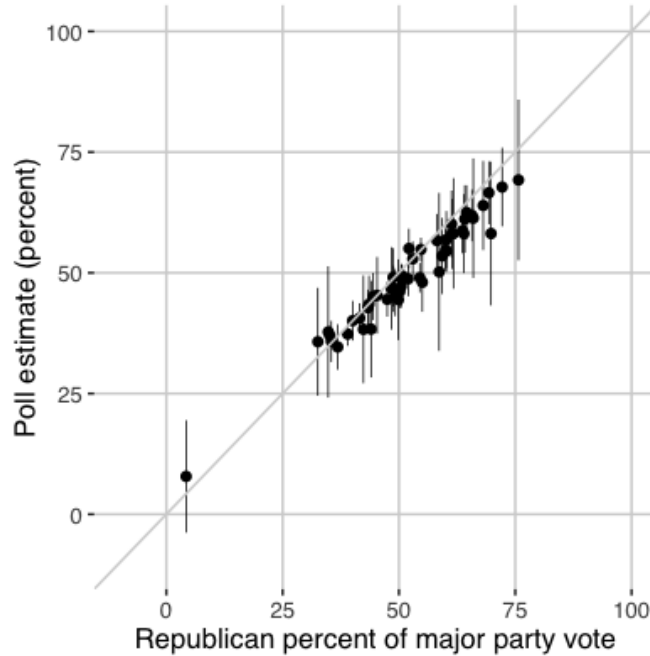
Since the election, there have been articles with titles like “Why Pollsters Were Completely and Utterly Wrong.” Most of these articles contain no data beyond the observations that (1) the pre-election polls showed a Clinton lead and (2) Trump won the election. They get a lot of things wrong. For example, Joe Flint and Luka I. Alpert in the *Wall Street Journal* write,

Only two polls consistently showed Mr. Trump in the lead—the USC Dornsife/Los Angeles Times and the IBD/TIPP tracking polls. They were often dismissed by pundits and media data analysts as outlying surveys, but in the end turned out to be the most accurate.

The authors of this piece don’t seem to be aware that Hillary Clinton, not Donald Trump, won the popular vote. Neither of these polls was the “most accurate,” not by a long shot. Hillary Clinton currently leads Donald Trump in the vote county by about 0.3 percent, so our poll with 3.6 percent Clinton margin was *more accurate* than the USC Dornsife/Los Angeles Times poll that had Trump with a 3.6 percent margin.

In fact, the final Clinton margin is likely to be substantially higher (around one percent). This is an important point. millions of votes have not yet been counted, since many states allow absentee ballots to be postmarked on election day and still counted. In California alone, about 3 million mail ballots have not been counted that should add about one million votes to Clinton’s lead. The actual error in our poll is likely to be around 2.5 percent, which is quite a bit better than the 6 percent error of the “most accurate” poll. We won’t know which poll is “most accurate” until the

FIGURE 1. Accuracy of 2016 CCES State Level Estimates



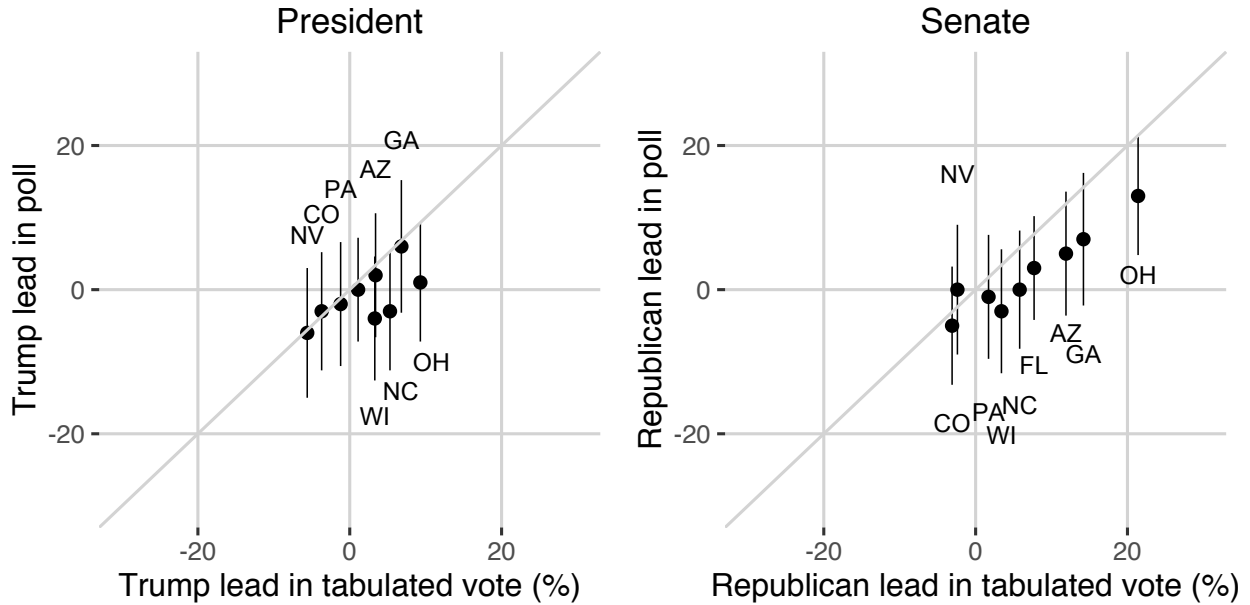
final count is completed, but there were a couple national polls (unfortunately, not ours) with one point margins for Clinton.

It's too early to do a proper postmortem. We are in the process of collecting additional data (including a reinterview of all of panelists who participated in our pre-election surveys) and will be posting detailed analyses of our polls over the coming weeks. I also serve on an AAPOR task force evaluating the 2016 polls which will be issuing a report next spring. Serious analyses cannot be done overnight. But we do appreciate that there is a lot of interest, so here are a few preliminary thoughts.

Contrary to some claims, the polling error was not “huge” or the “worst ever.” (Must everything now be exaggerated?) When we reported our *Economist* Poll on Monday with a 3.6 percent lead for Clinton, we also reported a “margin of error” of ± 1.7 percent. This is the margin of error for a sample proportion. For a lead (the difference between the proportions of respondents intending to vote for two different candidates), the margin of error is twice as large, or 3.4 percent. *The actual election outcome was within the margin of error for this poll.*

Nor was the error in the polling averages unusual. Last Friday, I told Mary Kissel on Opinion Journal (<http://www.wsj.com/video/opinion-journal-polls-and-more-polls/F464E3B3-06AD-41B8-8CEF-086B614630B2.html>) that the 2012 polling averages had been off by two percent (in the Democratic direction) so that a two-point error in the Republican direction was certainly possible. That is, in fact, about what happened. It's not fair to blame Nate Silver, but a lot of people thought he had some kind of magic ability to predict election outcomes with perfect accuracy. The real anomaly was 2012, when the particular configuration of states and relatively consistent state-level polling made it feasible for Nate (and at least five others) to predict the winner correctly in every state.

FIGURE 2. Accuracy of YouGov Battleground Polls



Polls predict the popular vote and only indirectly the winner and electoral vote. In a race that's 50/50, you should predict the wrong winner about half the time. Three states were decided by less than a point and another three were within 1.5 points. We do not evaluate polls based on whether they predict the correct winner, but on how close the poll estimate is to the popular vote. Furthermore, one accurate or inaccurate estimate does not tell us much about the quality of a poll—good polling needs to get it right most of the time, but we shouldn't expect polls to be right all of the time. The purpose of a margin of error is to give an indication of a poll's expected accuracy. We aim for 19 of 20 polls to be within the margin of error of the actual result. About two thirds of the time, the error should be less than half this amount.

How well did we do compared to our published margins of error? The Economist/YouGov poll was within its margin of error, but toward the outer edges, so this doesn't prove much one way or the other. A better approach is to compare many independent estimates. The Cooperative Congressional Election Study (CCES) had a very large sample (over 117,000 respondents), so it's big enough to produce reasonable estimates for each state. In Figure 1, we have plotted the poll estimate against the actual outcome. The vertical bars are 95 percent confidence intervals (plus or minus one standard error). If the confidence interval crosses the 45 degree line (in grey), then the election outcome was within the margin of error. In 46 of 51 states, Trump's actual margin was within the margin of error. That's slightly over 90 percent of the time. We would like for this to be 95 percent instead of 90, so this isn't great, but it's not terrible either.

We can do the same calculation for our Battleground polls. We polled ten different states after October 1. For comparison, we have taken all other polls published on RealClear Politics in the same period. (When the same pollster had multiple polls, we eliminated all but their last poll.) Figure 2 shows the range of estimates in each of these states along with the reported margins of

TABLE 1. Accuracy of Battleground Polls

| Pollster | Mean Absolute Error |
|-------------------------|---------------------|
| YouGov | 3.2 |
| Gravis Marketing | 3.7 |
| Emerson College | 3.8 |
| Ipsos | 3.9 |
| SurveyMonkey | 5.1 |
| Google Consumer Surveys | 6.5 |

error. YouGov polls are indicated using a red dot. Trump’s lead was computed using his share of the major party vote (by dropping third party and undecided voters from each poll).

In some states, such as Arizona, Colorado, and Florida, the polling was reasonably good, with some estimates too high and others too low, but only a few outliers (off the lead by more than 7 or 8 points). In other states, such as Wisconsin, the polling was uniformly awful, with every single poll underestimating the Trump lead, usually by a substantial amount. Our polls were quite good in six of these nine states. We were outside the margin of error in three (North Carolina, Ohio, and Wisconsin) and in all three cases the direction of the error was the same, which needs further investigation.

Table 1 shows the mean absolute error of the six organizations that polled in all nine of these states. Our mean absolute error was 3.2 percent. The samples included from 800 to 1000 registered voters, so the mean absolute error is approximately what would be expected from a simple random sample of this size.¹ Our accuracy was somewhat better than the others, whose mean absolute error in these states ranged from 3.7 percent to 6.5 percent.

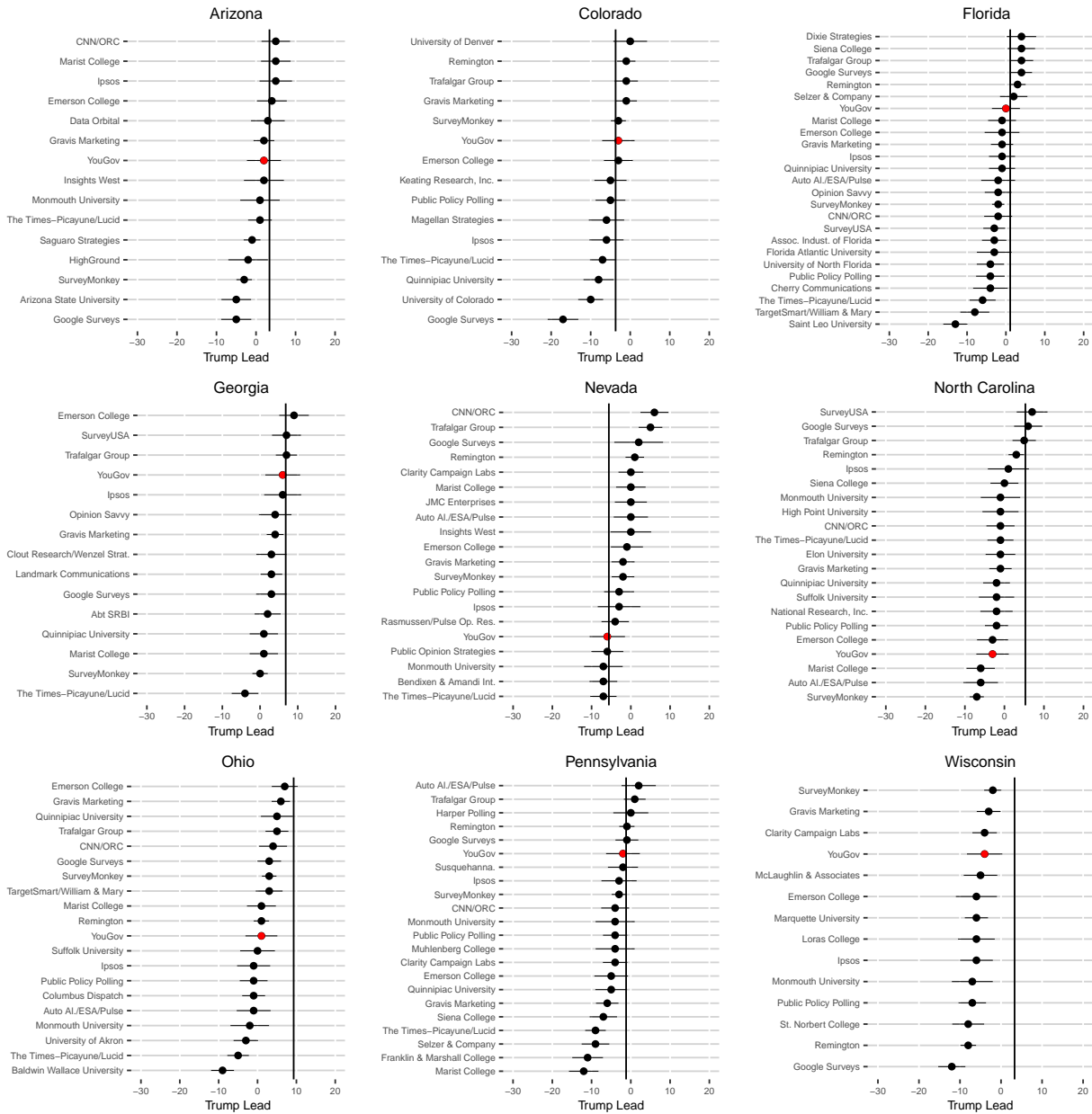
The most striking feature of polling in 2016 was not that a few polls had large errors. That always happens. Rather, it was the error in the polling averages. If all the polls overestimate support for one candidate (as most of the 2016 polls did for Clinton), then there will be no negative errors to cancel out the positive errors and averaging fails. After 2012, some had mistakenly assumed that polling averages didn’t have any error. This was a misreading of 2012 (35 of the last 36 polls in 2012 underestimated Obama’s lead), but hardly anyone noticed, since we got the outcome of the election right.

Another strange feature of this year’s polling that three weeks before the election there was wide disagreement among the polls. A few polls had Clinton with double-digit leads, while others showed it to be a close election. A few, like ours, showed very little movement during this period. It is unclear why the polls converged. It could be due to bad behavior on the part of some pollsters (“herding”) or, more likely, the use of similar methods that had some defect pushing the polls in Clinton’s direction. In particular, most polls underestimated Trump support in the midwest, where he did better than expected with white working class voters. This too deserves further investigation.

At this point, we don’t know what happened in the places where our polls were off. We have lots of hypotheses, but instead of speculating, we’re collecting more data. Every person that we

¹Our samples are not simple random samples, but with no selection bias and proportional allocation within strata, the same standard error calculation applies.

FIGURE 3. Comparison of Battleground Polls



interviewed before the election is being reinterviewed after the election, to see if there was much switching or last minute deciders. As state voter files become available, we will check on who voted and when and how they voted. Stay tuned.