Polls, Damned Polls, and Statistics

Jeffrey S. Rosenthal, University of Toronto

(November 24, 2020)

In the recent U.S. presidential election, public opinion polls indicated that Joe Biden would defeat Donald Trump handily. His actual victory was much tighter than expected – a popular vote margin about half of the predicted 8–10%, and narrow victories in states he was supposed to carry easily. These errors, amplified by the delayed count of certain pro-Biden mail-in ballots, and intensified by many people's hatred of Trump, led to howls of protest that the polls had betrayed us and could never be trusted again.

Some of the complaints came from statisticians themselves. One colleague wondered what was possibly left to say about polls, now that their inaccuracies had been so exposed. Another leaned in conspiratorially and whispered, "I would like to talk to a pollster after a few drinks, to find out what *really* happened." They felt a sense of shame, and saw the unreliable polls as a harsh and public repudiation of the very concept of random sampling upon which so much of statistics is based.

Due to the success (especially in Canada) of my general-interest book *Struck by Lightning: The Curious World of Probabilities*, I am often asked about polls by news media and various organizations, so I have had to confront these issues head on. And I have come to think that we should regard high-profile polling errors not as a failure, but as an opportunity.

Consider a typical statistics exam question: *An urn contains N balls of different colors. A sample of n balls is taken, of which exactly $n/2$ are red. Compute a 95% confidence interval for the fraction of red balls in the urn.* Every statistician knows that the answer to this question. The sample proportion is $\hat{p} = 0.5$, so if $1 \ll n \ll N$, then the interval has endpoints $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n} = 0.5 \pm 0.98/\sqrt{n}$. Indeed, that is how most polling companies compute their margin of error. Easy, right?

However, this answer requires the assumption, either implicit or explicit, that *the sample was drawn uniformly at random.* But suppose it wasn't. Suppose the question instead said: *the sample was drawn using an unknown, arbitrary scheme.* Then it is no longer easy. In fact, it is now completely impossible! Any statistics instructor assigning such a question would face an angry student revolt.

And yet, this second version is essentially what confronts pollsters. Sure, they phone people randomly, but most people do not answer – Pew Research reports that their response rates have declined to just 6%. If the non-respondents were missing at *random*, then they

would be of little consequence (aside from requiring more phone call attempts), and the usual confidence intervals would still apply. But what if they're not?

In fact, response rates do appear to be increasingly correlated with voting preferences, for reasons which remain unclear. Perhaps Trump supporters were less inclined to reveal their preferences to "elite" pollsters, or were harder to reach due to work responsibilities, or were less likely to follow COVID-19 safety protocols which would make them be home and available? All we know is that somehow, Trump supporters were significantly underrepresented in pre-election polls, in both 2016 and 2020. And all of the efforts to reweight the poll samples to match general population covariates such as race, age, gender, and education level, still failed to overcome these biases.

Despite these challenges, poll results haven't actually been that far off. They accurately predicted the 2018 U.S. midterm elections, and the 2008 and 2012 presidential elections. In 2016, they just slightly overestimated Hillary Clinton's popular vote margin as 4% instead 2%, and failed only because they predicted narrow wins in several states which ended up as narrow losses. Even their 2020 forecasts correctly predicted the winner (Biden) and most of the states that he eventually won, albeit with excessive spreads. These outcomes, achieved under impossible circumstances, are worthy of statisticians' praise, not scorn.

Bias-correction efforts for polls raise many interesting statistical questions. Which population covariates are relevant, and how should samples be reweighted to match them? How should past election results be incorporated into forecasting models to increase accuracy? Can other kinds of sampling, from online panels to social media scraping to intercepted web browsers, replace or supplement traditional random phone calls to produce better data? These questions should intrigue statisticians, not depress them.

One statistics instructor recently enthused that the Trump/Biden polling errors were an actual, real-life example of sampling bias in action. I didn't particularly share the amazement, since sampling bias is all around us and easy to find. But I do agree with the sentiment. High-profile missed forecasts provide compelling ways to teach our students the importance of statistical assumptions, as well as new opportunities to investigate innovative ways to overcome their limitations. Inaccurate polls should fill statisticians not with shame, but with excitement.