Response to Gelman and Azari (2017)

Corrie V. Hunt
Hart Research, Washington, DC

As Gelman and Azari make clear, there is no single smoking gun to point to as the primary explanation for the 2016 election that took so many of us by surprise. As a pollster at a progressive public opinion research firm, I will admit the election floored me in the most depressing and sickening of ways. It was not because I did not think it was possible. In fact, in the final weeks leading up to the election, I and many of my colleagues grew increasingly fearful that the tightening we saw in internal polls meant that a Clinton victory was far from certain. But I let myself be reassured by the confidence of the analytics projections. One of the most important lessons practitioners and consumers of public opinion research can learn from this experience is to take a much closer examination of election prediction models (lesson #3) and how nonresponse bias (lesson #5) affects polls in general and the polls that feed into forecast models. And finally, we cannot let ourselves get so fixated on the horserace numbers that we forget to listen to what voters are actually telling us in the rest of the poll and in qualitative research.

Overconfident Pundits Get Attention (Lesson #3)

Not only do overconfident pundits get attention but their confidence is contagious. People are impatient for answers and uncomfortable with uncertainty. So pundits and analysts that provide the allusion and allure of “more certainty” and make irrefutable projections are rewarded. Gelman and Azari suggest that we would be better served by reporting the results of individual polls plus the margin of error. I agree that individual horserace numbers and especially margin of error need to be prominently featured in discussions of predictions. But I am not convinced yet that election forecast models should be discarded—at least not yet. Instead, we should approach these predictions with a higher degree of skepticism (when warranted) and take a much closer look at the “black boxes” behind these election forecast models to see where they can be improved.

Without question, there are significant improvements that can be made in the way models treat polls that are used as input into the models. In 2016, the three most prominent election forecast models available to the public and media all made “adjustments” to the polls used in the models. FiveThirtyEight (Silver 2016) is the most transparent about what adjustments it makes in their methodology description; the model adjusts for likely versus registered voters, omitted third party candidates, and how polls conducted by the same firm are trending relative to their previous polls (e.g., whether a specific firm shows the race tightening, holding steady, or widening). Sam Wang (2016) is probably the least transparent (it is unclear how and to what degree pollster “house effects”—the degree to which a specific firm’s polls typically lean Democrat or Republican—are incorporated into the uncertainty assessment). None of these models talk about adjustments for sample composition, what I would argue is the most influential.

As we know, not all polls are created equal. And while it is impossible to perfectly predict the composition of the electorate ahead of time, there is plenty of information available from the voter file¹ and the Census to give us a solid idea of the basic demographics. And through examinations of poll after poll, we learn the relationships between demographic groups and vote choice. Pollsters should be able to recognize when the demographics of a sample are off and how that might affect the results. Election forecast models should account for sample quality too. A sample that underestimated white voters without a college degree or rural white voters would have underestimated Trump support. It just so happens that because of nonresponse bias (see below), many polls systematically under-represented these two voting blocs so critical to Trump’s base. Why would election forecast models not incorporate the same kinds of sample composition adjustments into predictions? It certainly stands to reason that forecast models could be improved by more careful attention to the demographics of the samples that feed into the models. Polls with higher quality sample—with demographic profiles closer to expectations to the likely electorate and those with smaller design effects—should be given greater weight in models.

Another critical component of these forecast models is how to factor in undecideds. The proportion of undecideds creates more uncertainty in an election outcome. Take the final NBC/Wall Street Journal Poll (Nov 2016) which had Clinton

---

¹Pollsters can draw samples from the voter file by purchasing a sample from a voter database vendor. Voter registration (including names, address) and vote history (whether a person voted, not who they voted for) are public information. Voter database vendors compile voter information from state and county election authorities into one database and make it available to pollsters and campaigns for purchase.

²A design effect is an index of the amount of weighting adjustments applied to a sample to account for designs other than simple random sampling. Larger design effects decrease the precision of measurement and increase the margin of error.

© Corrie V. Hunt
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named author(s) have been asserted.

Published with license by American Statistical Association
Nonresponse Bias is a Thing (Lesson #5)

Nonresponse bias is absolutely real and I do not know a single pollster who did not realize this before the election. But where nonresponse bias is coming from and how pollsters adjust for it in the future is a critical lesson and one that deserves serious consideration. It is not just about partisan fluctuations in who decides to answer the phone and complete a political survey (as Gelman and Azari rightly point out). Consistent lower responses among less educated and rural voters across polls created a systematic bias that favored Clinton. This is a problem that can be addressed for future polls and frankly, it is one that could have been addressed pre-election. For the record, some pollsters did. The only way to address it is to start from a sample that allows you to track response rates. The only way currently that we have to track response rates across demographics and partisanship in real-time as a survey is being fielded is to draw the sample from the voter file. At present, the use of voter file versus RDD sampling is inconsistent across public pollsters. I admit there is a real argument to be made for the value of an RDD sample; it is true that given the significant number people on the voter file who lack valid phone numbers, RDD is the only method that truly gives everyone an equal chance of being contacted. And RDD methodology does allow pollsters to make corrections for sample composition based on Census demographics variables. But you cannot systematically track response rates with RDD samples in the way that is necessary to correct for nonresponse bias.

Firms like TargetSmart and Catalist that provide pollsters with samples from the voter file not only have information about the age, gender, and location (hence, ability to track rural response rates) but also model scores predicting the likely partisanship of individuals on the file. In addition to tracking basic demographics, pollsters using the voter file as a sample source can easily track response rates for modeled Republicans and modeled Democrats and either adjust by weighting or calling more members of the under-represented party. This is something most campaign pollsters do regularly and is something more media pollsters should consider.

The trickier problem to deal with right now is education. As most pollsters should know by now simply by comparing their sample profile to Census figures, it is significantly harder to get less educated people on the phone to complete a survey (see Keeter et al. 2017). The voter file right now does not have accurate information about voters’ highest level of education completed and the education models that currently exist are not sufficiently accurate to rely upon. This means that there is not a way to systematically track response rates by education level (other than comparing the profile of the sample to known characteristics) or to accurately target people by education level. If those who build predictive models are looking for something to do before the next election, creating a more accurate model for education is a good place to start.

Overvaluing of Analytics Projections and Under-Valuing of Qualitative Research (New Lesson #20)

Election forecast models care about one question in polls: the horse race. Polls of course provide much more information that tells us where the electorate is heading, which gets ignored when the sole focus is on that one single question. But qualitative research in which researchers listen to voters explain their views and feelings about the candidates and the country provides a much deeper and richer understanding than a single number. This kind of valuable qualitative research is based on rigorous methodological procedures to recruit participants and carefully crafted focus group guides. There is no way any public opinion researcher conducting qualitative research could have walked out of a discussion with genuine swing or undecided voters feeling certain about the outcome. Unfortunately, focus groups cost money and media organizations typically prefer to spend their money sponsoring polls rather than focus groups. That leaves campaigns that pay for focus groups the primary beneficiaries of this deeper understanding.

But occasionally, media outlets or organizations sponsor this kind of research and it can be incredibly valuable for adding nuance to the conversation. The Annenberg Public Policy Center the last several election cycles has partnered with Peter Hart to conduct a series of focus groups as part of their ”Voices of the Voters” project (Annenberg Public Policy 2016). I had the privilege of working with Mr. Hart during 2015 and 2016. For each of these discussions, we invited journalists to view and report and then Annenberg made video of the full groups available on their website (“Peter Hart’s Voices of the Voters”). It is an incredible resource and one the public is not typically privy to. These discussions revealed that undecided voters were deeply dissatisfied with their options but still looking for change, not only in improving the economy but of changing the culture in Washington. That desire for something different and the ambivalence that many voters who were leaning toward Clinton felt about supporting her painted a much less stable picture of the electorate than analytics projections foretold. I sincerely hope that media organizations will learn from 2016 and invest
in qualitative research to help inform discussions about the election and provide a deeper, more nuanced understanding of the electorate. These conversations provide an important context, as Amy Walter (2016) observed after a focus group with blue collar voters in Pittsburgh:

It’s important not to read or project too much onto the opinions of 11 people trapped in a windowless room and talking politics for two hours in June. But, they are important windows into how “regular” people are dealing with and processing the political environment. Or, as Hart puts it, it’s important for us in the media/political world, to not “jump ahead of the story.” And, while the “story” here in DC is that Trump is flailing and doomed, he’s still holding strong with his base. That base alone won’t win him the election. He’s got to expand beyond it. But, as long as he still has these voters, he has not gone into free fall.

While many political commentators were talking as if the race was over after a rough few weeks for Trump in June, these voters were telling us not to write him off. We should have listened.

References

Annenberg Public Policy (2016), Peter Hart’s Voices of the Voters. Available at https://www.annenbergpublicpolicycenter.org/events/peter-harts-voices-of-the-voters/[2]

Blake, A. (2016, November 17), "How America Decided, at the Last Moment, to Elect Donald Trump," The Washington Post. Available at https://www.washingtonpost.com/news/the-fax/wp/2016/11/17/how-america-decided-at-the-very-last-moment-to-elect-donald-trump/?utm_term=.5a1244ae5ecf[2]

Keeter, Sa, Hatley, N., Kennedy, C., and Lau, A. (2017, May 15), “What Low Response Rates Mean for Telephone Surveys,” Pew Research Center. Available at http://www.pewresearch.org/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/[2]

NBC/Wall Street Journal (2016), “Survey of 1,282 Likely Voters,” Question 8 of 22, November 5, 2016. Available at https://www.scribd.com/document/330170543/9amET-NBCWSJ-release[1]

Sanger, T. (2017, November 4), “Free Press Poll: Donald Trump Gains on Hillary Clinton as Race in Michigan Tightens,” Detroit Free Press. Available at http://www.freep.com/story/news/politics/2016/11/04/free-press-poll-donald-trump-gains-hillary-clinton-race-michigan-tightens/93287658/[2]

Silver, N. (2016, June 29), “A User’s Guide to FiveThirtyEight’s 2016 General Election Forecast,” Fivethirtyeight Blog. Available at https://fivethirtyeight.com/features/a-users-guide-to-fivethirtyeights-2016-general-election-forecast/[1]

Walter, A. (2016), “When Conventional Wisdom Gets Ahead of the Voters,” The Cook Political Report Blog. Available at http://cookpolitical.com/[3]

Wang, S. (2016, June 30), “The Presidential Meta-Analysis for 2016,” Princeton Election Consortium Blog. Available at http://election.princeton.edu/2016/06/30/the-2016-presidential-meta-analysis/[1]