

Challenges in adjusting a survey that overrepresents people who are interested in politics¹

Andrew Gelman² and Gustavo Novoa³

12 Sep 2023

We agree with the general point of Bailey (2023) that random sampling is a distant benchmark for real-world polls that either have very low response rates or that are constructed from panels that do not even purport to be random samples of the population. A few years ago the American Association for Public Opinion Research released a statement criticizing opt-in internet polling, saying that “these methods have little grounding in theory and the results can vary widely based on the particular method used” (Link, 2014). As discussed by DeSilver (2014) and Gelman (2014), however, there’s no “grounding in theory” that allows you to make statements about those missing 90% of respondents. Or, to put it another way, the “grounding in theory” that allows you to make claims about the nonrespondents in a traditional survey, also allows you to make claims about the people not reached in an internet survey. Whether your data come from random-digit dialing, address-based sampling, the internet, or plain old knocking on doors, you’ll have to do some adjustment to correct for known differences between sample and population. (Bailey speaks of “weighting,” but we prefer the term “adjustment,” which encompasses more general possibilities for population inference.)

The key contribution of Bailey’s article is to emphasize the relevance of differences between sample and population that have not been included in survey adjustments. When a poll oversamples Democrats or Republicans, adjustment for party identification can make a big difference (Cohn, 2016); indeed, this sort of differential nonresponse varies consistently during the campaign season and explains some variation in the polls that had been mistakenly taken as signs of large opinion swings (Gelman et al., 2016). Indeed, even the notorious *Literary Digest* poll of 1936 could have been made much more accurate simply by adjusting for respondents’ stated votes in the previous election (Lohr and Brick, 2017). Adjusting for party identification or voting history can be challenging because these variables are not tabulated in the census, but imperfectly adjusting could be better than not even trying.

Political polls oversample people who are interested in politics. This bias is well known, but pollsters are typically interested in voters more than in the general population, so we have not tended to think too much about it. Ignoring this bias might have made sense in an era when the rate of survey response was comparable to that of voter turnout, but not so much when the two rates differ by a factor of ten. It is not clear, though, what to do about this oversampling of people who are interested in politics, given that the distribution of this variable is not known in the general population.

Bailey provides an interesting clue in his Figure 3, which shows a correlation between interest in politics and support for Biden in the 2020 American National Election Study (ANES). This was not something we had expected to see, especially after all the news coverage of passionate Trump voters. If this correlation also appeared in pre-election polls, as seems likely, this would represent an adjustment opportunity that was not taken.

We explore this further by looking at other survey questions and other years. We did not find in recent years of the ANES the particular question used by Bailey regarding interest in politics, so instead we looked at a question on interest in politics that was asked in every ANES presidential election campaign pre-election poll since 1952. The plots in our Figure 1 have a similar form to Bailey’s Figure 3, with the

¹ Discussion for *Harvard Data Science Review* of “A new paradigm for polling,” by Michael A. Bailey.

² Department of Statistics and Department of Political Science, Columbia University, New York.

³ Department of Political Science, Columbia University, New York.

only differences being that (a) we showing the Republican share of two-party vote preferences (based on weighted averages of the survey responses) rather than separate bars adding to 100%, and (b) we use the sizes of circles to show the estimated proportions in each group. We see that in 2016 and 2020, Donald Trump had lower support among respondents who said they were interested or very interested in the election, with a variety of other patterns in earlier years.

In order to better understand the changes over time, in the left graph of Figure 2 we display the estimated average interest in politics in each survey for supporters of the Republican candidate, the Democratic candidate, and others. Just to check, in the right graph of Figure 2 we show the same time series but broken down by party identification. Using either measure, we see the unsurprising pattern that partisans are consistently more interested in the election. We also see a general increase in interest in elections during the past two decades, which could be attributed to increased polarization or to a change in survey respondents: as response rates have declined, perhaps those who remain are disproportionately more likely to be politically involved.

A challenge here is that many polls adjust for estimated likelihood to vote, which itself could be highly correlated with interest in politics. So, even the direction of any adjustment for interest in politics is not clear, but in any case we should recognize the potential importance of going beyond conventional adjustment variables.

References

- Bailey, M. A. (2023). A new paradigm for polling. *Harvard Data Science Review* 5 (3), <https://hdsr.mitpress.mit.edu/pub/ejk5yhgv/release/2>
- Cohn, N. (2016). We gave four good pollsters the same raw data. They had four different results. *New York Times*, 20 Sep, <https://www.nytimes.com/interactive/2016/09/20/upshot/the-error-the-polling-world-rarely-talks-about.html>
- DeSilver, D. (2014). Q&A: What the New York Times' polling decision means. *Pew Research Center*, 28 Jul. <https://www.pewresearch.org/short-reads/2014/07/28/qa-what-the-new-york-times-polling-decision-means/>
- Gelman, A. (2014). President of American Association of Buggy-Whip Manufacturers takes a strong stand against internal combustion engine, argues that the so-called “automobile” has “little grounding in theory” and that “results can vary widely based on the particular fuel that is used.” *Statistical Modeling, Causal Inference, and Social Science*, 6 Aug. <https://statmodeling.stat.columbia.edu/2014/08/06/president-american-association-buggy-whip-manufacturers-takes-strong-stand-internal-combustion-engine-argues-called-automobile-little-grounding-theory/>
- Gelman, A., Goel, S., Rivers, D., and Rothschild, D. (2016). The mythical swing voter. *Quarterly Journal of Political Science* 11, 103-130.
- Link, M. (2014). The critical role of transparency & standards in today's world of polling and opinion research. *American Association for Public Opinion Research*, 1 Aug. https://web.archive.org/web/20140806165804/http://www.aapor.org/AAPOR_Letter_on_NYT_Polling_Change.htm
- Lohr, S., and Brick, J. M. (2017). Roosevelt predicted to win: Revisiting the 1936 Literary Digest poll. *Statistics, Politics and Policy* 8, 65-84.

Republican share of two-party vote preference among people with low, moderate, or high interest in the election

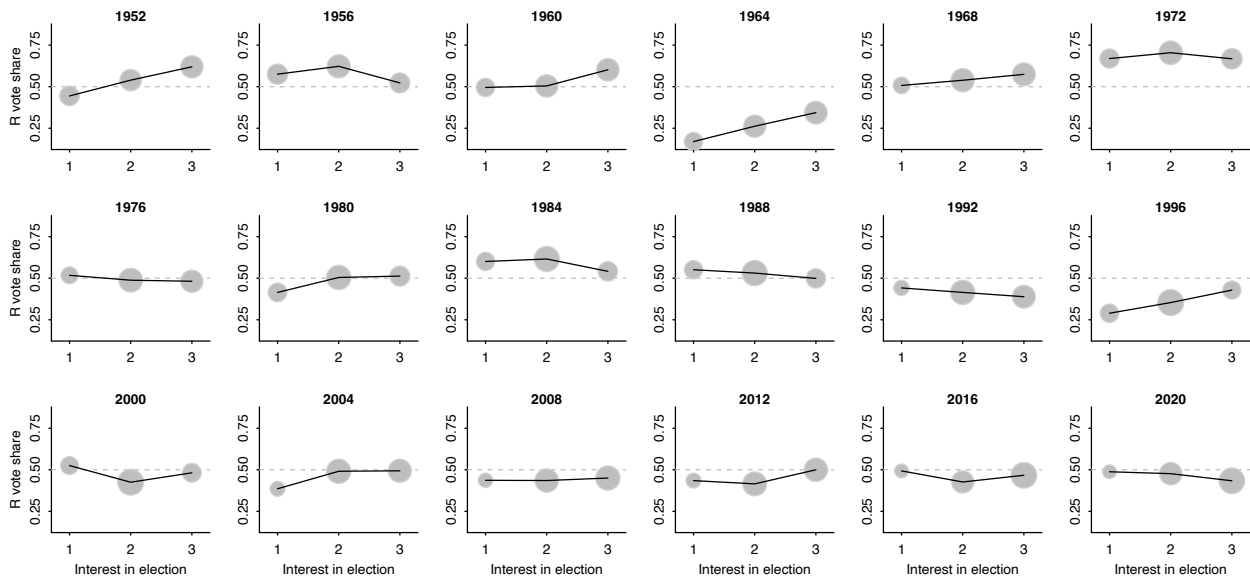


Figure 1. For each election since 1952, estimated share of two-party support for the Republican candidate for president, among pre-election survey respondents who say they are “not much interested,” “somewhat interested,” or “very interested” in the upcoming election. In 2016 and 2020, respondents who expressed more interest in the election were less likely to support the Republican candidate; patterns were different in previous elections. Within each graph, the areas of the circles are proportional to the estimated proportion of each of the three levels of interest. Data from American National Election Study.

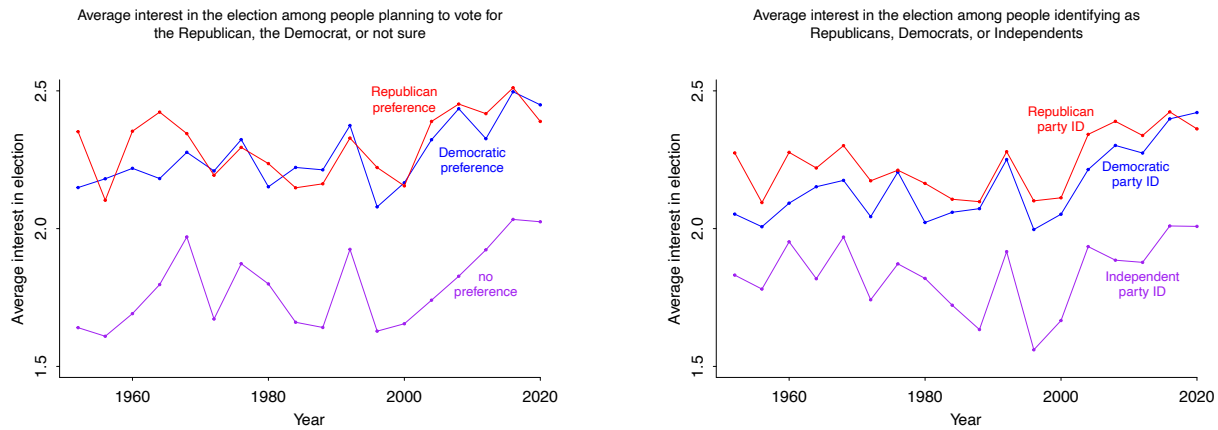


Figure 2. For each election since 1952, estimated average interest in the upcoming election in pre-election polls (on a 1-3 scale), among Republicans, Democrats, and others. The left graph shows the breakdown by vote intention; the right graph shows party identification. Unsurprisingly, partisans (under either coding) are consistently more interested in politics than nonpartisans. There has also been an increase in interest in politics in the past twenty years, which could be attributed to increased partisan polarization or to increased selection bias accompanying declining survey response rates. Data from American National Election Study.