

Polls, Forecasts, and Aggregators

Mark Blumenthal, *The Huffington Post*

The *USA Today* headline was both imprecise and telling: “Election aftermath: How’d pollsters like Nate Silver do?” (M. Moore 2012).

Silver is not a pollster, but rather someone who uses poll-based modeling to “aggregate” polls. The text of the article made that distinction but, apparently, the headline writer did not.

This semantic confusion reflects long-standing tensions inherent in preelection polling: is the purpose of polling to measure the horse race or to gain a better understanding of voter choices? When does a snapshot in time become a forecast? This article reviews the recent history of poll aggregation and how its emergence dovetails with the larger debate about measurement versus forecasting in preelection polling.

ORIGINS

Poll aggregation may have had its first news media appearance in 1992. That year, the *Economist* first published a “poll of polls,” an average of a half dozen national trial heat results from surveys sponsored by other news organizations. It appeared on a weekly basis during the fall presidential campaign.

Later the same year, William Schneider, then senior political analyst for CNN, brought the “poll of polls” concept to his network. On October 28, 1992, just after reporting results from the Gallup poll that CNN had sponsored along with *USA Today*, Schneider explained that he had also taken “a poll of polls” conducted by other news organizations because all such surveys “are subject to error” and he wanted to “see how much consistency there is across all of the polls.” He summarized the results of each survey as well the overall average (Schneider 1992).

CNN’s initial poll of polls may have been a bit of a hedge, because as Schneider noted on the same broadcast, President Clinton’s lead was smaller on the CNN-USA Today-Gallup poll than on any of the others. Nevertheless, Schneider told his audience, all six produced results “within the margin of error of the averages we’re looking at right now.” Hedge or not, Schneider’s “poll of polls” became a regular feature of CNN’s political coverage during the next campaign in 1996.

Four years later, in 2000, the fledgling news website RealClearPolitics.com extended the concept, producing its own national average like Schneider, but also provided a summary of polls from 21 key battleground states, each with its own average. Unlike CNN, RealClearPolitics also included polls conducted with an automated, recorded voice methodology (known as “interactive voice response” or IVR).

Until then, compilations of state-level polling data had only been available through costly subscriptions to publications like *The Polling Report* and *The National Journal’s Hotline*. As more

media outlets started publishing their poll stories to the Internet, however, anyone with a web browser could access these results for free. Sites like RealClearPolitics added value by republishing the top line numbers in an accessible format along with hyperlinks to the source articles.

The explosive growth of the Internet brought more aggregation websites in 2004, such as Andrew Tannenbaum’s electoral-vote.com and Sam Wang’s Princeton Election Consortium. In 2006, Charles Franklin and I founded Pollster.com, a site that went beyond simple averages and presented poll trends as interactive, graphical displays.

The probabilistic forecasts of Nate Silver’s FiveThirtyEight came in 2008, capturing the imaginations of politically interested readers on a far larger scale. Silver joined the *New York Times* in 2010, and during the 2012 campaign his fame reached stratospheric heights. Meanwhile, political scientists who had long been developing their own forecasting models joined the fray, such as Simon Jackman (at Pollster, now part of *The Huffington Post*) and Drew Linzer (at votamatic.org).

The emergence of these poll aggregation sites was enabled by exponential growth in public polling. The first wave began in the 1980s, when relatively costly in-person surveys gave way to quick turnaround telephone polls. A second wave emerged when widespread adoption of the Internet provided a new business model. IVR pollsters like Rasmussen and Public Policy Polling learned they could promote their businesses by publishing results directly to company websites. Despite an absence of media sponsors, their polls often went “viral,” attracting readers via links from hundreds of blogs and news websites, as well as the popular poll aggregators.

The net result was an explosion of state-level polling. Consider that in 2000, the final RealClearPolitics averages for 21 battleground states were based on only 49 state-level presidential race trial-heat polls conducted in October and late September. During the 2012 election cycle, however, HuffPost Pollster logged 1,240 state-level polls that asked an Obama-Romney trial-heat question, including more than 500 polls conducted in October and late September. Of the total, nearly half (594) used an IVR methodology and another 10% (121) were conducted online using samples drawn from nonprobability opt-in Internet panels.

DOES AGGREGATION THREATEN POLLING?

The popularity of poll aggregation, and of Nate Silver specifically, led some in the survey research profession to speculate about impending threats. “It’s much easier, cheaper, and mostly less risky to focus on aggregating and analyzing others’ polls,” wrote Gallup poll editor-in-chief Frank Newport on his company’s website in the aftermath of the 2012 election. “Organizations that traditionally go to the expense and effort to conduct individual polls could, in theory, decide to put their

efforts into aggregation and statistical analyses of other people's polls in the next election cycle and cut out their own polling." The result, Newport worried, might lead to "fewer and fewer polls left to aggregate and put into statistical models" (Newport 2012).

Newport's doomsday scenario faces three obstacles: First, much of the fuel for polling aggregation, particularly around presidential elections, comes from the inexpensive automated polls whose producers face a different set of costs and incentives than traditional pollsters. IVR surveys do face an existential threat, but it comes from legal barriers to making automated calls to mobile phones (Marketing Research Association 2013), the only telephone service now used by a third of US adults (Blumberg and Luke 2013).

Second, the segment of the public that uses and trusts polling aggregation sites remains relatively small. A recent national telephone survey finds that Americans express greater trust toward public opinion polls released by "news media organizations" (43%) than those compiled by "people or websites that average multiple polls together" (30%) (Wilner 2013).

The third and most important barrier, however, gets at the different purposes of polling and aggregation. At its best, poll aggregation can make sense of a deluge of polling data. Like Bill Schneider's first application in 1996, an average of competing polls illustrates the range of random error and puts the results of a single poll sponsored by a news organization into broader context. Moreover, these averages only work well for the small handful of survey questions that are asked with nearly identical wording and format by different organizations. As such,

A recent national telephone survey finds that Americans express greater trust toward public opinion polls released by "news media organizations" (43%) than those compiled by "people or websites that average multiple polls together" (30%) (Wilner 2013).

poll aggregation is of little use for the questions that make up the vast majority of news media polls. If the funding of these polls is grounded in more than just the measurement of voter preferences near elections, the threat from aggregation should be slight.

FOR WHAT PURPOSE?

A tension often exists between the survey researchers who produce preelection polls and consumers who watch or read poll stories in the news media during campaigns. Pollsters typically stress that their results should be considered only a "snapshot in time," not a forecast. They frequently downplay measurements of vote preference—the so-called horse race—as least important to their efforts to help explain who voted and why.

Yet those who follow political news during campaigns appear to gravitate to polls for their apparent forecasting value. That interest is implicit in the pattern of Google searches for the word "poll" during the past 10 years. As seen in figure 1, such searches show massive "spikes" of 10 to 20 times baseline search volume just before the general elections of 2004, 2008, and 2012. Smaller

but equally prominent spikes have occurred on the eve of the 2006 and 2010 off-year elections.

From these data we can surmise that as tens of millions of Americans start paying attention to election campaigns, the most partisan among them become increasingly interested in the prospects of their chosen candidate. Like baseball fans checking the standings, these political enthusiasts never tire of checking the latest polls to see how their "team" is doing and whether it is on track to win the big prize.

IMPLICATIONS FOR SCORING ACCURACY

The tension between polls as election forecasts and polls as a measurement of everything *but* the outcome is especially pronounced when it comes to measuring poll accuracy. While some pollsters dismiss the notion of treating preelection polls as forecasts, many in the field are happy to treat the apparent accuracy of polling near elections as a sign of its overall health.

For example, the Pew Research Center conducted an exhaustive study on response rates in 2012. Despite "dramatically" declining response rates, they found evidence that telephone polls with adequate sample coverage continue to provide accurate data on most measures, a finding that "comports with the consistent record of accuracy achieved by major polls when it comes to estimating election outcomes" (Pew Research Center 2012a).

The tension between forecasting and measuring preelection snapshots of voter preferences is reflected in the decades-long debates about measuring polling accuracy. From the report on the polling failures of 1948 edited by the famed statistician Frederick Mosteller (Mosteller et al. 1949) to Nate Silver's more recent model-driven

scoring (Silver 2010), pollsters have proposed a series of new ways to measure the accuracy of trial-heat election preference questions and debated their merit (Blumenthal 2010; Crespi 1988; Martin, Traugott, and Kennedy 2005)

These many approaches have struggled to resolve three key issues: respondents who say they are undecided, the gap between the poll's field dates and Election Day, and whether to focus on just the winning candidate, the margin between the top two, or the other candidates who finish further back in the pack.

Implicit in all of these challenges is, again, the fundamental tension between polls as forecasts and snapshots in time. Yet also implicit in the scoring of poll accuracy is Crespi's resolution a quarter century ago that ruling out the forecasting value of polls conducted "immediately before an election" is to "impugn the meaningfulness of all polls. If polls cannot achieve such accurate predictability, why should we accept any poll results as having meaning relevant to real life?" (Crespi 1988, 5).

However, the extension of Crespi's logic—that the accuracy of late polls provides "an empirical basis" for judging poll accuracy in the campaign—is questionable. At issue is the frequently

observed phenomenon that polling error declines close to Election Day as polls “converge” or “herd” around the averages (Blumenthal 2008; Lavrakas et al. 2008; Linzer 2012; D. Moore 2008). This topic too is a subject for considerable debate. Do results converge because voters grow increasingly certain about their choices and less volatile or because pollsters have adjusted their methods with an eye toward avoiding a late outlier?

One need not resolve the issue—nor make a judgment of the state of pollster ethics—to see evidence of late tinkering in the methods used for final polls. In 2004, for example, Gallup changed its assumption about turnout for their last poll before the election (Newport and Moore 2004). For their final poll in campaign 2012, the Pew Research Center opted to add a weight for past vote choice not applied to earlier surveys (Blumenthal and Edwards-Levy 2013; Pew Research Center 2012b). More importantly, the vast majority of public polls withhold details

to those that are offered earlier. The surveys most likely to influence press coverage or candidate fundraising occur early in the campaign, not late. Yet the empirical basis for their accuracy is weak.

Can the recent advances in poll aggregation and forecasting help provide a solution?

One possibility is to use the trend estimates produced by advanced poll-averaging models, like the one created by Simon Jackman (Jackman 2012). These models can correct their trend lines after the election to better match the outcome, so their estimates of pollster “house effects” are, in essence, a measure of error throughout the campaign.

Such an approach has obvious drawbacks: only the end result can be truly validated, and a greater emphasis on house effects throughout the campaign risks creating even more incentive for pollsters to conform to poll averages.

Do results converge because voters grow increasingly certain about their choices and less volatile or because pollsters have adjusted their methods with an eye toward avoiding a late outlier?

of their likely voter models, so any last minute changes are hidden.

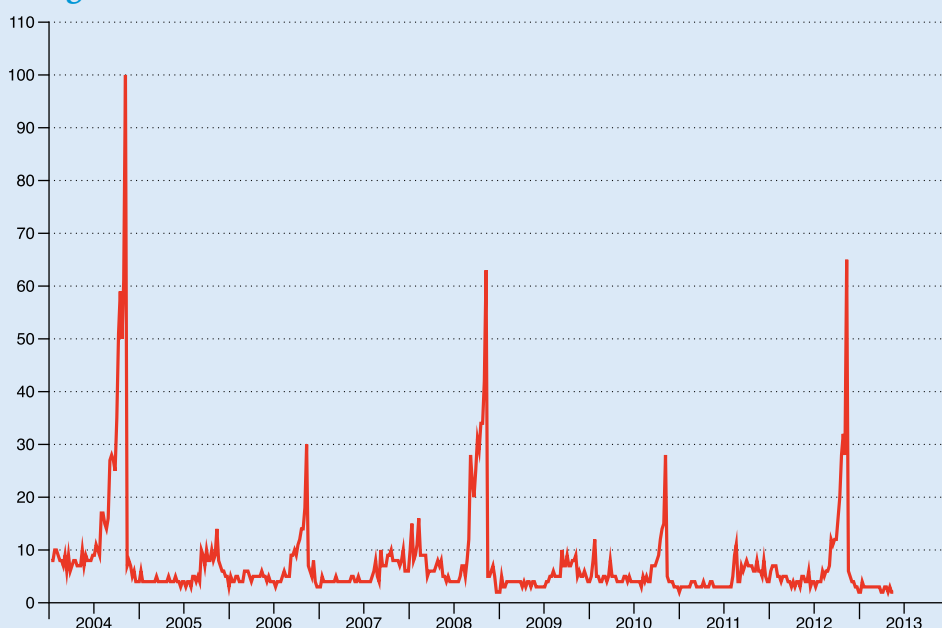
At the very least, the intense pressure on pollsters to produce an accurate result on their final poll creates incentives for the sort of tinkering that makes late polls less comparable

A second possibility might be to focus more on the composition of the likely electorate than on vote preference. Official lists of registered voters include demographic data as well as the actual history of turnout of individual voters. The much celebrated data analytics advances of Obama’s 2012 campaign hint at the possi-

bility of better demographic measures of past electorates and perhaps even the ability to model demographics of likely voters during the campaign (Issenberg 2012).

Yes, both turnout intention and vote preference can vary during the course of the campaign, and pre-election polls aim to measure both. Yet while we have considerable evidence on the variability of candidate preference, we know less about the volatility of electorate composition. If past voter history allows for relatively accurate predictions of the demographics of an electorate relatively early in the campaign—as a new generation of campaign data analysts claims it can—then, in theory, those predictions might also facilitate more accurate polls.

Figure 1
Google Searches on “Poll” in the U.S.



Note: Data used to create this chart was located at <http://www.google.com/trends/explore#q=poll&cmpt=geo&geo=US>.
Source: Google.com/trends.

At the very least, the convergence of polling and more advanced analytics, which has been most visible in polling “aggregation,” may help point the way. ■

REFERENCES

- Blumberg, Steven J., and Julian V Luke. 2013. *Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July-December 2012*. Centers for Disease Control and Prevention. <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201306.pdf> (September 17, 2013).
- Blumenthal, Mark. 2008. “More on the ‘Convergence Mystery.’” *Pollster.com*. http://www.pollster.com/blogs/more_on_the_convergence_myster.php?nr=1 (September 18, 2013).
- . 2010. “Rating Pollster Accuracy: How Useful?” *Pollster.com*. http://www.pollster.com/blogs/rating_pollster_accuracy_predi.php?nr=1 (September 17, 2013).
- Blumenthal, Mark, and Ariel Edwards-Levy. 2013. “HUFFPOLLSTER: When Is It OK to Weight Polls by Past Vote?” *Huffington Post*. http://www.huffingtonpost.com/2013/09/13/weight-polls-past-vote_n_3923178.html?utm_hp_ref=pollster (September 18, 2013).
- Crespi, Irving. 1988. *Pre-election Polling: Sources of Accuracy and Error*. New York: Russell Sage Foundation.
- Issenberg, Sasha. 2012. “The Definitive Story of How President Obama Mined Voter Data to Win a Second Term.” *MIT Technology Review*. <http://www.technologyreview.com/featuredstory/509026/how-obamas-team-used-big-data-to-rally-voters/> (September 18, 2013).
- Jackman, Simon. 2012. “Model-Based Poll Averaging: How Do We Do It?” *Huffington Post*. http://www.huffingtonpost.com/simon-jackman/modelbased-poll-averaging_b_1883525.html (September 18, 2013).
- Lavrakas, Paul, Michael Traugott, Michelene Blum, Cliff Zukin, and Don Dresser. 2008. “The Experts Reply on the Poll Convergence Mystery.” *Survey Practice*. <http://surveypractice.org/2008/12/19/the-experts-reply-convergence-mystery/> (October 6, 2009).
- Linzer, Drew. 2012. “Pollsters May Be Herding.” *VOTAMATIC*. <http://votamatic.org/pollsters-may-be-herding/> (September 18, 2013).
- Marketing Research Association. 2013. “TCPA Restrictions on Using Autodialers to Call Cell Phones | Marketing Research Association.” <http://www.marketingresearch.org/tcpa-restrictions-on-using-autodialers-to-call-cell-phones> (September 17, 2013).
- Martin, Elizabeth A., Michael W. Traugott, and Courtney Kennedy. 2005. “A Review and Proposal for a New Measure of Poll Accuracy.” *Public Opinion Quarterly* 69 (3): 342–69.
- Moore, David. 2008. “Evaluating the 2008 Pre-Election Polls – The Convergence Mystery.” *Survey Practice*. <http://surveypractice.org/2008/11/25/a-question-for-the-experts-evaluating-the-2008-pre-election-polls-%E2%80%93-the-convergence-mystery/> (October 6, 2009).
- Moore, Martha. 2012. “Election Aftermath: How’d Pollsters Like Nate Silver Do?” *USA Today*. <http://www.usatoday.com/story/news/politics/2012/11/07/election-poll-results/1690553/> (September 17, 2013).
- Mosteller, Frederick, Herbert Hyman, Philip J. McCarthy, Eli S. Marks, and David B. Truman. 1949. *The Pre-Election Polls of 1948: Report to the Committee on Analysis of Pre-Election Polls and Forecasts*. Social Science Research Council.
- Newport, Frank. 2012. “Polling, Likely Voters, and the Law of the Commons.” *Gallup.com*. <http://pollingmatters.gallup.com/2012/11/polling-likely-voters-and-law-of-commons.html> (September 17, 2013).
- Newport, Frank, and David Moore. 2004. “Final Poll Shows Presidential Race to Be Dead Heat.” *Gallup News Service*. <http://www.gallup.com/poll/13873/final-poll-shows-presidential-race-dead-heat.aspx> (September 18, 2013).
- Pew Research Center. 2012a. “Assessing the Representativeness of Public Opinion Surveys.” *Pew Research Center for the People and the Press*. <http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/> (September 17, 2013).
- . 2012b. “Obama Gains Edge in Campaign’s Final Days.” *Pew Research Center for the People and the Press*. <http://www.people-press.org/2012/11/04/obama-gains-edge-in-campaigns-final-days/> (September 18, 2013).
- Schneider, William. 1992. “Poll of Polls’ Reflects Polling Consistency.” *CNN News*.
- Silver, Nate. 2010. “Pollster Ratings V4.0: Methodology.” *FiveThirtyEight: Politics Done Right*. <http://www.fivethirtyeight.com/2010/06/pollster-ratings-v40-methodology.html> (September 17, 2013).
- Wilner, Elizabeth. 2013. “Kantar’s ‘Path to Public Opinion.’” <http://us.kantar.com/public-affairs/politics/kantars-path-to-public-opinion/> (September 17, 2013).