



Mitt Romney speaks at the Hotel Fort Des Moines on the night of the Iowa Caucuses on Jan. 3, 2012. Low turnout at events such as the Iowa caucuses makes pollsters' work difficult. CHIPSOMODEVILLA / GETTY IMAGES

AUG 25, 2014 AT 11:20 AM

## Is The Polling Industry In Stasis Or In Crisis?

By Nate Silver

Filed under Polling

There is no shortage of reasons to worry about the state of the polling industry.

[Response rates to political polls are dismal](#). Even polls that make every effort to contact a representative sample of voters now get [no more than 10 percent](#) to complete their surveys — down from about 35 percent in the 1990s.

And there are [fewer high-quality polls than there used to be](#). The cost to commission one can run well into five figures, and it has increased as response rates have declined.<sup>1</sup>

Under budgetary pressure, many news organizations have understandably preferred to trim their polling budgets rather than lay off newsroom staff.

Cheaper polling alternatives exist, but they come with plenty of problems. “Robopolls,” which use automated scripts rather than live interviewers, often get response rates in the low to mid-single digits. Most are also prohibited by law from calling cell phones, [which means huge numbers of people are excluded from their surveys](#).

How can a poll come close to the outcome when so few people respond to it? One way is through extremely heavy [demographic weighting](#). Some of these polls are more like polling-flavored statistical models than true surveys of public opinion. But when the assumptions in the model are wrong, the results can turn bad in a hurry. (To take one example, the automated polling firm Rasmussen Reports got fairly good results from 2004 through 2008, but has been [extremely inaccurate since](#).) Furthermore, demographic weighting is an [insufficient remedy](#) for the failure to include cellphone-

only voters, who differ from landline respondents in ways that go beyond easily identified demographic categories.

Another tactic is for a pollster to copy off its neighbors. As my colleague Harry Enten [described earlier this month](#), and [as other researchers have found](#), robopolls and other polls that take methodological shortcuts show better results when there are also traditional, live-interviewer polls surveying the same races. The cheap polls may “herd” off stronger polls, tweaking their results to match them. This can make them superficially more accurate, but they add little value. Where there are better polls available, the cheap poll duplicates the results already in hand. Where there aren’t, the cheap poll may stray far from an accurate and representative sample of the race.

Then there are the companies that have cheated in a much more explicit way: by fabricating data. There is strong evidence that [Strategic Vision](#) and [Research 2000](#) faked some or all of their survey results. The odds are that there are more firms out there like them.

Internet-based polling has been a comparative bright spot. In fact, the average online poll was [more accurate than the average telephone poll](#) in the 2012 presidential election. However, there is [not yet a consensus in the industry](#) about best practices for online polls. Some online methods do not use [probability sampling](#), traditionally the bedrock of polling theory and practice. This has worked [well enough in some cases](#) but [not so well in others](#).

But all of this must be weighed against a stubborn fact: We have seen no widespread decline in the accuracy of election polls, at least not yet. Despite their challenges, the polls have reflected the outcome of recent presidential, Senate and gubernatorial general elections reasonably well. If anything, the accuracy of election polls has continued to improve.

I’ve spent much of the past few weeks updating our polling database in preparation for the launch of our [2014 Senate forecasts](#). In conjunction with those forecasts, we’ll also release our [first full set of pollster ratings since 2010](#). But for now, my aim will be to consider the results for the polls as a whole: how accurate the average poll is in different types of races — presidential elections, gubernatorial elections and so forth — and how this has changed over time.

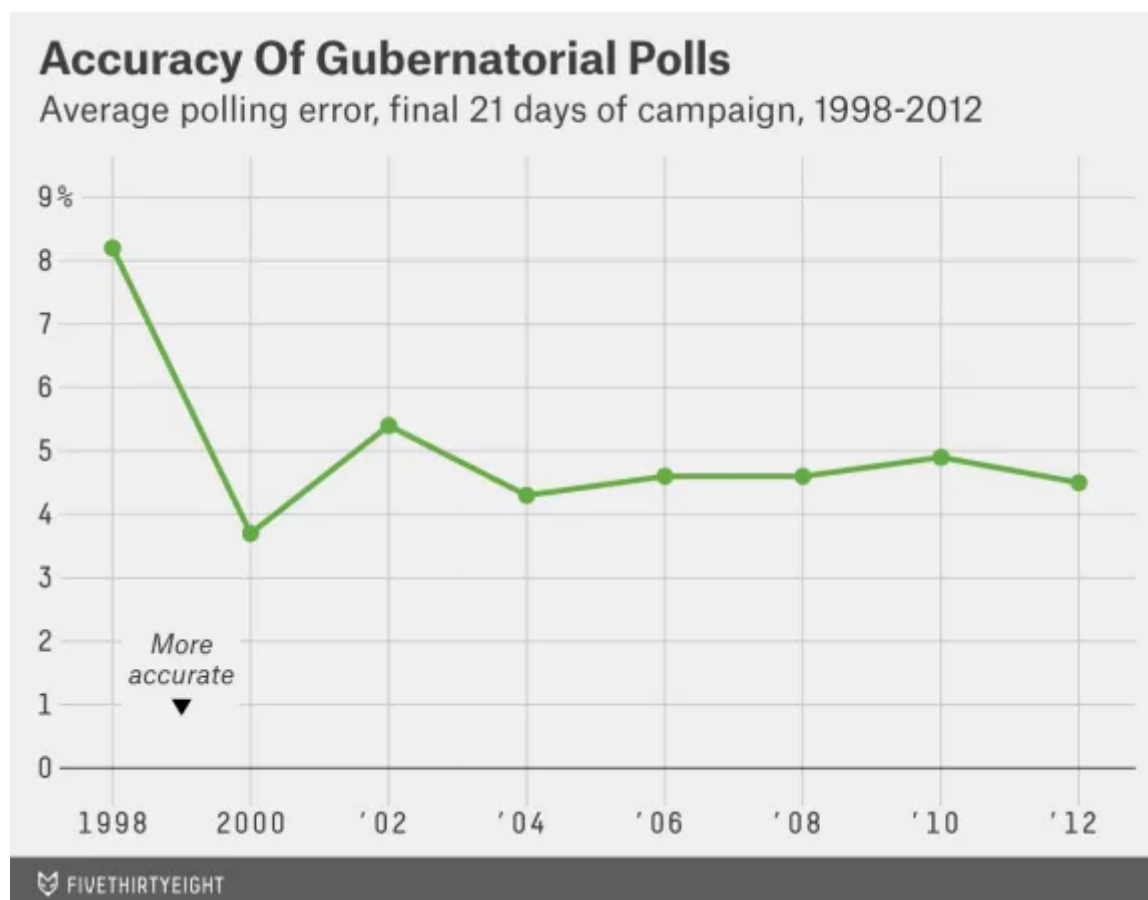
I’ll conduct this analysis in the context of the seeming contradiction we identified: The polls have managed to produce high-quality output (pretty good forecasts of election

outcomes) with worse and worse input (fewer and fewer people responding to them). It's something of a paradox.

A few unavoidable technical details: Our pollster ratings database contains about 6,600 polls conducted in the final three weeks of each campaign. In theory, every poll of presidential, Senate, gubernatorial and U.S. House general elections since 1998 should be included, along with polls of presidential primaries and caucuses.<sup>2</sup> Detail-oriented readers can find a few more notes about what's included in the polling database in the footnotes.<sup>3</sup>

I'll be evaluating poll accuracy by comparing the difference between the polled and actual margin for the top two finishers in the race. For instance, if a poll projects the Democrat to beat the Republican by 3 percentage points and she wins by 7 points, that counts as a 4-point error. In races with multiple viable candidates (as in the case of many presidential primaries), the error calculation is still based on the margin separating the top two finishers.<sup>4</sup>

Let's start by considering the polls from gubernatorial races. Like the rest of the polls, they present opportunities for different interpretations of the evidence. You can tell a tale of continuous improvement in the polls — or see a reminder about how prone to failure they can be.



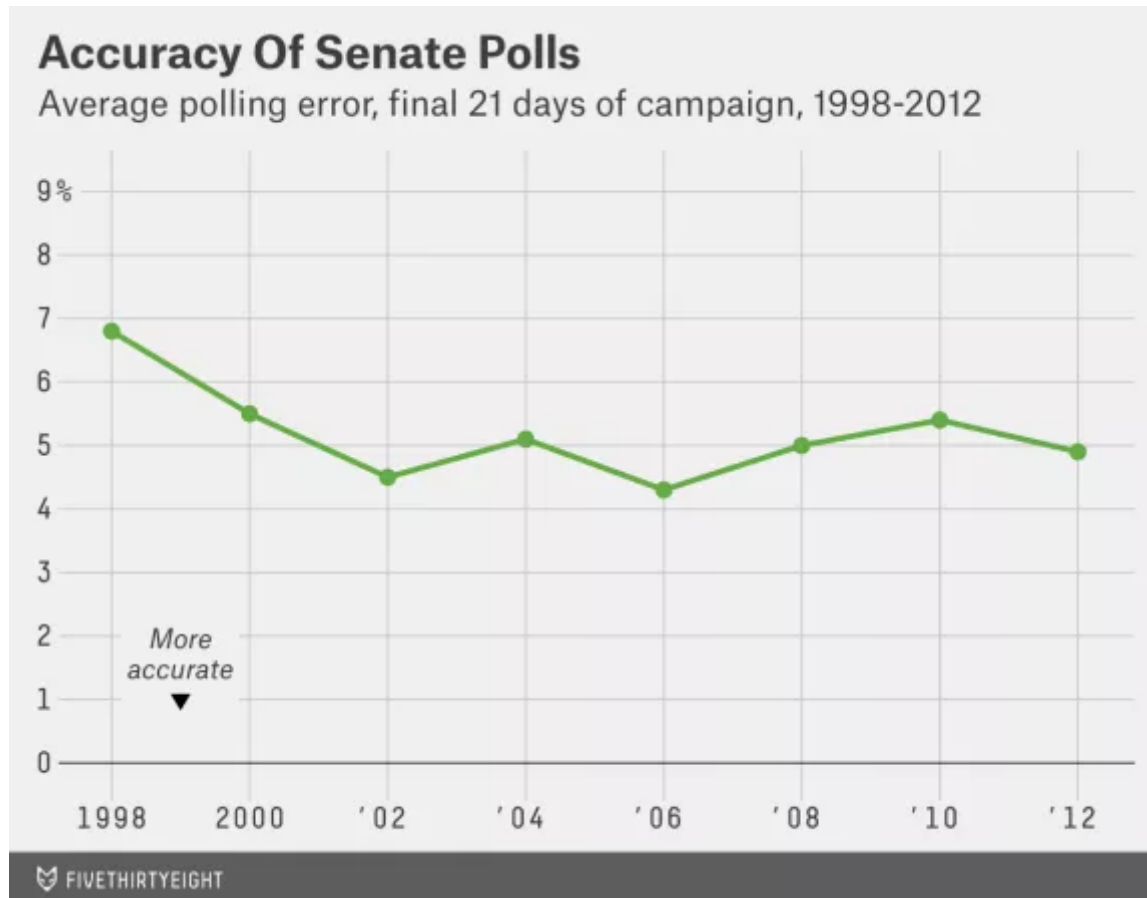
Between 2000 and 2012, gubernatorial polls did reasonably well, averaging an error of 4 to 5 percentage points. It's important to note that the accuracy of the average poll — what these figures describe — is not the same thing as the accuracy of the polling average. The polling average will cancel out some of the errors from individual polls provided that the misses come in opposite directions.<sup>5</sup>

But at other times, almost all the errors are in the same direction.<sup>6</sup> This was the case in 1998, for instance, when Democrats [did surprisingly well in the midterm elections](#). Between House, Senate and gubernatorial polls that year, about 75 percent underestimated the Democratic candidate's performance. The misses were especially bad in gubernatorial races; many polls called for Republicans to win the gubernatorial races in Iowa and Georgia, for example, when Democrats won them by a solid margin instead. Overall, the average gubernatorial poll in 1998 was 8 percentage points off the final margin.

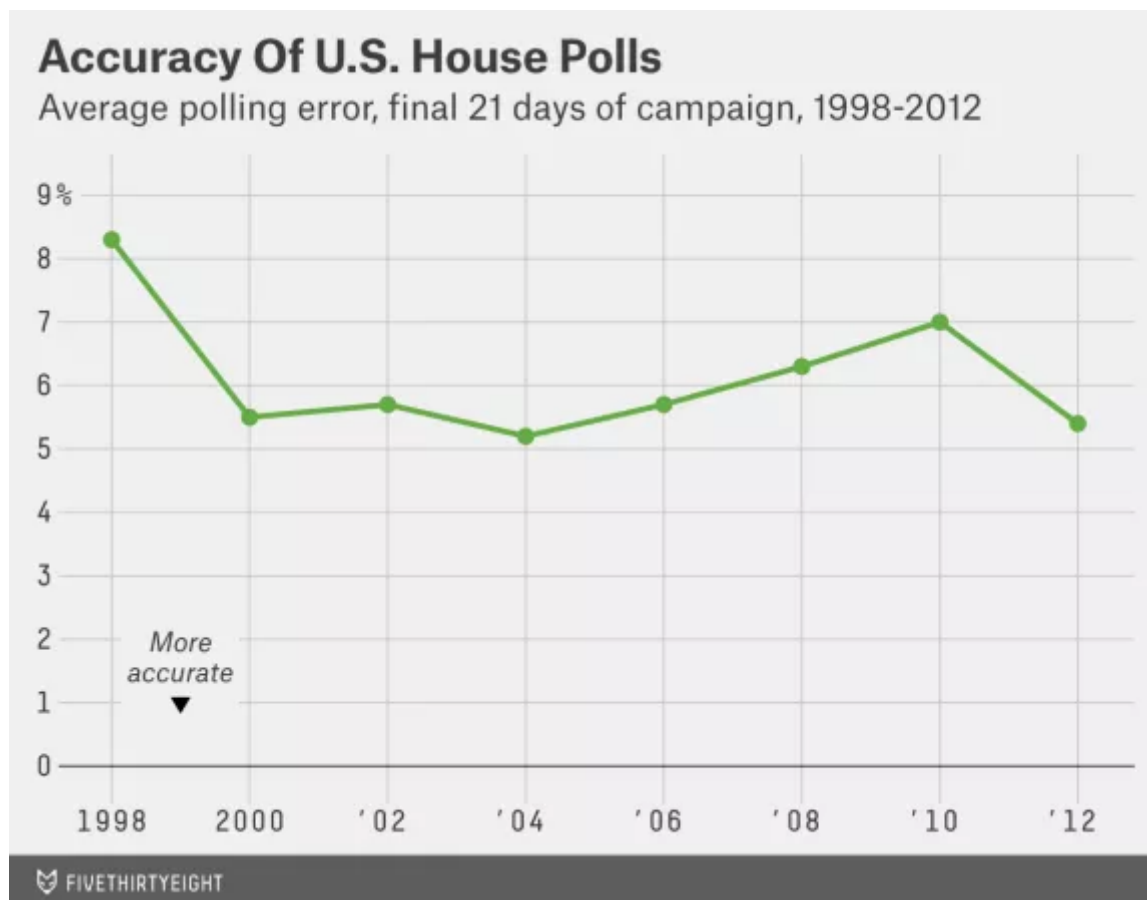
Perhaps it can be viewed as a hopeful sign that we haven't had such a disastrous year since then. But 1998 serves as testimony that when the polls have a bad year, they often make the same mistakes in a number of states, systematically underestimating the performance of Democratic or Republican candidates. That means you can have one election cycle, or several in a row, when the polls get almost every state right — followed

by another where there are misses all over the map.<sup>7</sup> Another bad polling year might be lurking around the corner.

The results for polls of Senate elections have been similar. Since 2000, the average Senate poll has missed the final margin in the race by about 5 percentage points. However, the average error was considerably larger in 1998 — 6.8 percentage points — with most of those errors underestimating the performance of the Democratic candidate.

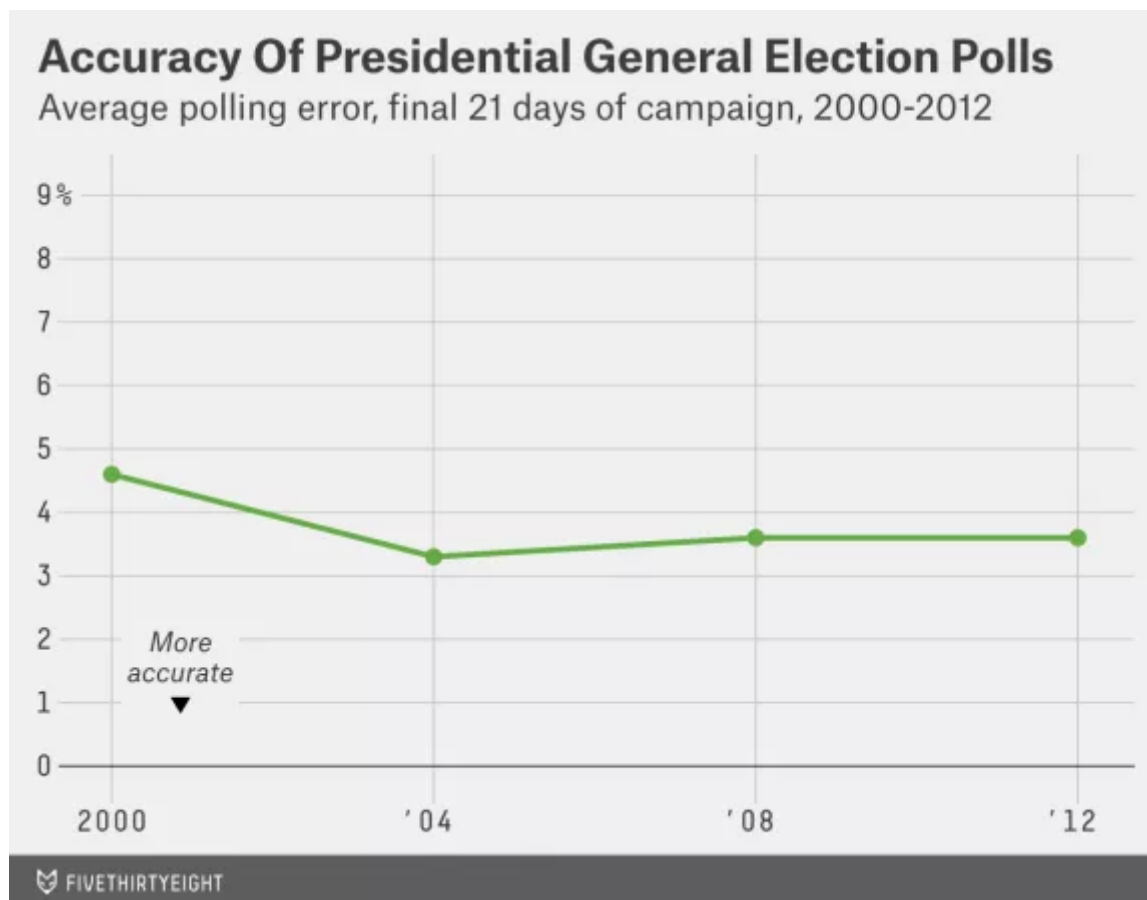


Polls for elections to the U.S. House<sup>8</sup> have been somewhat more error-prone. On average since 1998, they've missed the final margin by 6.2 percentage points (as compared to 5.2 points for gubernatorial polls and 5.1 points for Senate polls). Some of this is because a higher share of House polls are partisan surveys, which [present their own set of problems](#) and sometimes wildly exaggerate their candidate's standing.

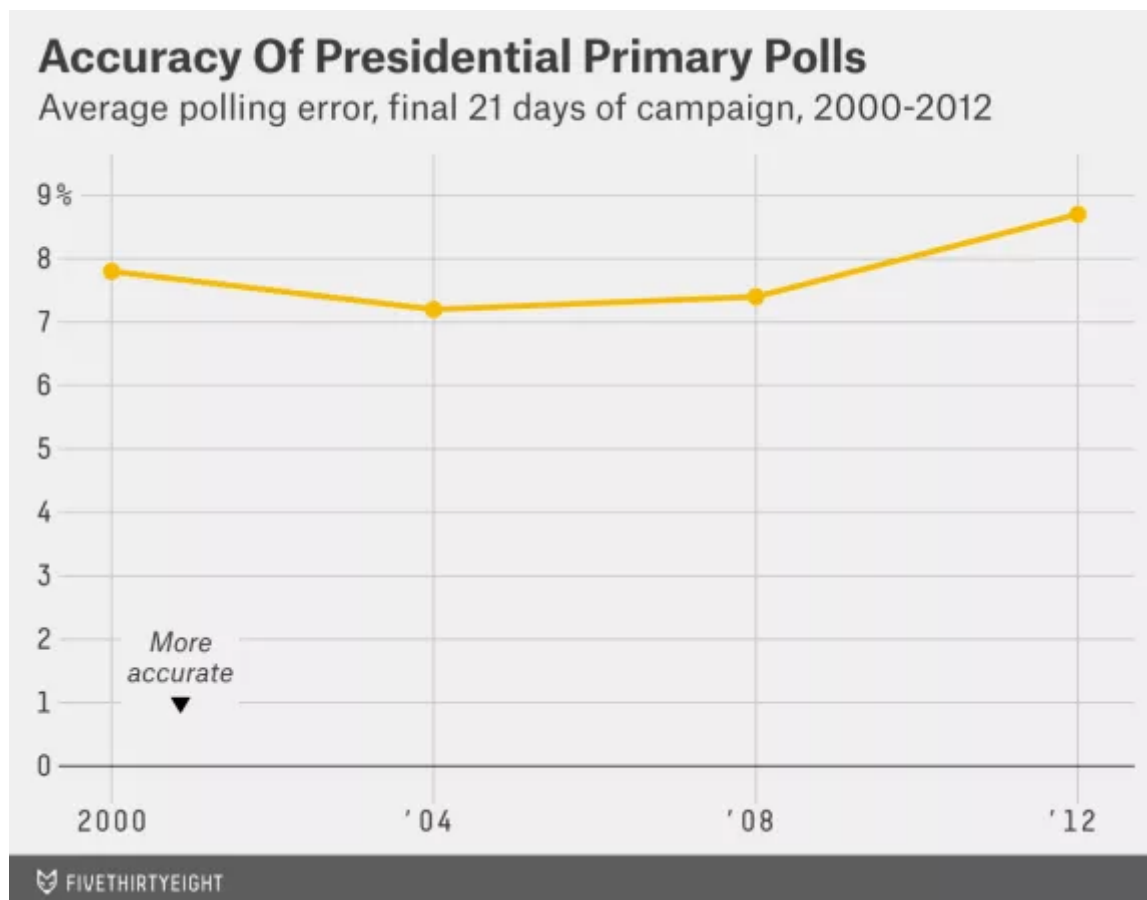


But as a rule, polling error increases the farther you go down the ballot. It's not uncommon for a House candidate to lose her race despite leading in the polls by 5 percentage points or more. Of the House polls in our database that show one candidate leading by between 5 and 10 percentage points, 23 percent picked the wrong winner. The same was true for only 8 percent of such polls of the presidential general election.

Indeed, the November presidential election has usually been the easiest one for pollsters. Since 2000, the average state poll in a presidential race has missed the final margin by 3.8 percentage points.<sup>9</sup> This is impressive given that a 900-person poll — roughly the average sample size of the presidential polls in our database — will miss the result by an average of about 2.7 percentage points on the basis of random [sampling error](#) alone.<sup>10</sup> In 2004, the average presidential poll missed the final margin in its state by just 3.3 percentage points, barely larger than than this theoretical, unavoidable minimum error; 2008 and 2012, each with an average error of 3.6 percentage points, were nearly as good. The 2000 presidential election was associated with a somewhat larger average error of 4.6 percentage points.



Polls of presidential primaries and caucuses are another matter entirely; they haven't been much good. The average error for presidential primary polls since 2000 has been 7.7 percentage points — about twice as large as for presidential general elections. The polls were especially bad in the 2012 Republican primaries, when they missed by an average of 8.7 percentage points.



This is an important [Psephology 101](#) finding — polls of primaries are much less accurate than polls of general elections. Perhaps this isn't emphasized enough. It seems that every election cycle, people are surprised by how wild and inaccurate polls of the primaries can be, and equally shocked at how stable and reliable polls of the general election are. Keep that in mind when everyone freaks out in 2016 because Elizabeth Warren or Rand Paul wins the New Hampshire primary after trailing in the final University of New Hampshire poll by 8 points.<sup>11</sup> These outcomes are par for the course.

Nevertheless, it's worth reflecting on why there's been such a big difference in primary and general election polls. One reason is that there are a lot more swing voters in the primaries, which can make public opinion more volatile. Most Democratic voters really liked both Hillary Clinton and Barack Obama in 2008, for instance; it would not have taken much to sway them from one candidate to the other. By contrast, perhaps only 10 percent to 15 percent of voters would seriously have considered voting for both Obama and John McCain later that year. When a pollster can predict 85 percent of votes on the basis of party identification alone, it isn't so hard to get close to the eventual result.

However, the increased volatility of polling in the primaries does not account for all their differences with general election polls. If we limit our study to polls conducted only in the final three days of the campaign — right on the eve of the election, when there's little



time remaining for news events to sway voter preferences — the primary polls still had an average error of 6.4 percentage points. That compares to 3.5 percentage points for polls in the final three days before presidential general elections.

Another factor may be the lower turnout in primaries. Polls can go wrong in essentially two ways: by misestimating candidate preferences, or by misestimating who will turn out to vote. The latter mistake is easier to make when voter participation is low. In Iowa, for instance, [only 92,000 people voted in the Republican caucuses in 2012](#), as compared to the 730,000 who would vote for Mitt Romney in the general election later that year. In that sort of environment, a candidate really can win by turning out his or her supporters when other candidates don't. The importance of turnout [may be overstated](#) in general elections but not in the primaries.

But I suspect there's third factor at play — one that explains something about the polling paradox. In the general election, you can model someone's vote quite accurately by knowing a few things about him, such as his race, age, place of residence and education status. Demographic weighting can cover for a lot of problems with low and unrepresentative response rates. That doesn't work nearly as well in the primaries, where votes don't break down as cleanly along demographic lines. On the basis of demographics alone, distinguishing a Ron Paul voter from a Rick Santorum voter from a Newt Gingrich voter would have been all but impossible in 2012.

Put another way, the polling results from the primaries — which have been pretty bad and which are perhaps getting worse — may be the better reflection of polls in their naked form, before demographic weighting is applied.

Demographic weighting is a legitimate and necessary practice. The past decade or so has seen stronger and stronger partisanship, [stronger and stronger alignment](#) of voting in different states, stronger correlations between up- and down-ballot voting (there are [fewer split tickets than there used to be](#)), and [stronger predictability of voting behavior on the basis of demographics](#). All of that makes demographic weighting more powerful. It has become easier to project election outcomes on the basis of [informed priors](#) — *without* conducting polls.

If my hypothesis is right — the relatively steady accuracy of the polls is the result of the increasing demographic predictability of elections helping to offset lower response rates — we could see a disastrous year for the polls if and when political coalitions are realigned. A black or Hispanic Republican presidential candidate could scramble the demographic coalitions that prevailed between 2000 and 2012, as might a moderate

blue-collar Democratic nominee, or a certain type of third-party candidate. None of these things is especially likely to happen in the near term, but the current political coalitions won't hold forever. The [2012 presidential map](#) looks fairly similar to the one in 2008, or 2004, or 2000, for instance, but rather different from the one in 1996 or in years before that, when states now seen as locks for one party or the other were considered swing states instead.

But if we should be skeptical of the polls, we should also be rooting for them to succeed. One of the reasons news organizations bother to conduct expensive surveys is to serve as a check on the misrepresentative opinions of elites, including those of their own reporters. Even a deeply flawed poll may be a truer reflection of public opinion than the “[vibrations](#)” felt by a columnist situated in Georgetown or Manhattan.

---

## Footnotes

1. The math here is simple. If you want a survey with 1,000 respondents, it's more expensive when you need to call 10,000 people to get them (with a 10 percent response rate) instead of 3,000 (with a 35 percent response rate).
2. We've searched far and wide for these polls, though we're undoubtedly missing some from the earlier years of the coverage period.
3. What's included in the database is a simple matter in principle — every survey during the last three weeks of a presidential, U.S. Senate, U.S. House, gubernatorial or presidential primary campaign since 1998, including special elections and runoffs. But there are some definitional issues that come up in practice:
  - Sometimes, a polling firm publishes results both among [likely voters](#) and among registered voters, or all adults. In these cases, we use the likely voter version of the poll. Perhaps 95 percent of the polls in the database are surveys of likely voters.
  - The use of [tracking polls](#) is restricted such that there is no overlap in the sample. For example, if a pollster publishes results from a three-day tracking poll that spans from Monday to Wednesday, and rolls the sample forward one day at a time, we'd wait until it publishes results spanning Thursday through Saturday before we include it as a separate entry in the database. (More precisely, we start with the version of the tracking poll conducted nearest to the election date and then go backward in time until we find non-overlapping dates. A firm's final tracking poll before an election is always included.)
  - As I mentioned, the database covers polls released in the final three weeks (the last 21 days) before each election. I feel strongly that this approach is superior to evaluating only a firm's final poll — particularly given the tendency of polls to converge or herd toward one another late in a race. However, there are some exceptions in the case of presidential primaries. No polls of the New Hampshire primary are included until the Iowa caucus is complete, and no polls for states beyond New Hampshire are included until New Hampshire is done. We also exclude primary polls when a leading candidate withdrew from the race after the poll was conducted.

- The database includes polling from firms that are on FiveThirtyEight’s “blacklist” because we know or suspect they faked their data. The problem with excluding polling from firms like Research 2000 or Strategic Vision is that this assumes knowledge we wouldn’t have had at the time of the election. There are probably other fake polling firms whose numbers are being included in polling averages today — we just don’t know their identities yet.
- 4. In the New Hampshire Republican primary in 2012, for example, we’d look at the difference between Mitt Romney and Ron Paul — Paul finished second, about 14 points behind Romney — and compare it to the difference between Romney and Paul as projected by each poll.
- 5. If Poll A shows the Republican winning by 10 points and Poll B shows the Republican winning by 4 points, and the Republican eventually wins by 7 points instead, the average poll will have missed the outcome by 3 percentage points. But the polling average will have gotten it exactly right.
- 6. A provocation in the Nerd Wars: Any election model worth its salt needs to account for the property that errors are somewhat correlated from poll to poll. If you select a random poll from our database and compare it to the average of other polls from the same race, you’ll find that they miss in the same direction about 70 percent of the time.
- 7. As a consequence of this, models that are calibrated on past data but that don’t include at least one bad polling year like 1998 in their training sample are likely to produce overconfident results. Although our polling data is less comprehensive prior to 1998, the FiveThirtyEight Senate model is calibrated based on polling from elections since 1990, and our presidential model is calibrated based on elections since 1968.
- 8. Our polling database also includes [generic congressional ballot polls](#), which project the popular vote across all House races nationwide. I haven’t included generic ballot polls in the averages described in this article, so as to produce better comparability to Senate and gubernatorial polls. For the same reason, I don’t include national polls of the presidential race in these calculations.
- 9. The average national poll has been even better, missing the final margin by 3.0 percentage points.
- 10. This calculation is not quite the same thing as the margin of error as reported in polls. The margin of error also reflects a poll’s sampling error, but it does so by describing the [95 percent confidence interval](#) associated with a poll rather than the average error. Also, the margin of error is typically reported to reflect the sampling error associated with one candidate’s vote share, whereas our focus here is on the margin separating the top two candidates.
- 11. Be equally wary when, that November, some pundit describes New Hampshire as “too close to call” when one candidate is ahead by 8 points.