



LAST EDIT MAR. 10, 2023

How Our Pollster Ratings Work

See our latest pollster ratings

the details

Longtime readers of FiveThirtyEight are probably familiar with our [pollster ratings](#): letter grades that we assign to pollsters based on their historical accuracy and transparency. Since [2008](#), we have been evaluating pollsters and using these ratings to inform both the public and [our models](#) about the quality of individual polls. Over the years, the methodology for these ratings has evolved, but the fundamental principle has remained the same: look at all the polls we have that were conducted within three weeks of an election, and try to determine how accurate each pollster has been and might be in the future.

Our pollster ratings can be found [at this dashboard](#). There, you'll see a graph with every pollster we have evaluated, organized by their most recent rating, as well as a searchable and sortable table of all the pollsters. Each pollster also has an individual rating page (for example, here's the one for [Selzer & Co.](#)) that shows details about its rating, including all the polls we've analyzed by that pollster and their accuracy. If you want even more detail, you can [download the associated datasets](#). Pollster grades are also included next to every poll we publish on our [polls page](#), to help give context to the data.

Our pollster ratings are based on a metric called Predictive Plus-Minus. This metric is based on several key factors, including:

- Simple error for polls (i.e., how far away the poll results are from the actual election margin).
- How well other pollsters performed in the same races (i.e., whether this pollster is as good as, better than or worse than others).

- Methodological quality (i.e., whether this pollster is conducting polls in accordance with professional standards).
- Herding (i.e., whether this pollster appears to just be copying others' results).

While our dataset includes several other metrics for understanding how well a pollster has historically performed, our letter grades are based entirely on Predictive Plus-Minus.

Below are all the methodological details of how we currently calculate Predictive Plus-Minus as well as a few other metrics that appear in the data. If you want to see methodology for previous versions of our pollster ratings, scroll to the bottom of this page for a series of links that contain all the methodological updates we've published over the years.

Step 1: Collect and classify polls

Almost all of the work is in this step; we've spent hundreds (thousands?) of hours over the years [collecting polls](#). The ones represented in the [pollster-ratings database](#) meet our [basic standards](#) as well as three simple criteria:

- They were conducted in 1998 or later. (We chose 1998 as the cutoff point because there are multiple sources that make polling data available from 1998 to the present, meaning that the data ought to be reasonably comprehensive. If you are aware of errors or omissions from this data, please reach out to let us know!)
- They have a median field date within 21 days of the election date.
- They were conducted for one of the following types of elections:
 - Presidential general elections
 - Presidential primaries or caucuses
 - U.S. Senate general elections
 - U.S. House general elections
 - Gubernatorial general elections

Of course, it's not so simple. A number of other considerations come up from time to time:

- Sample sizes are sometimes missing from older polls. In these cases, we've estimated a poll's sample size from its reported margin of error or from how many people a polling firm surveyed in other polls where the sample size was listed.¹ As a last resort, we use 600 as a default sample size.
- If a pollster lists results among likely voters and registered voters (or all adults), we include only the likely-voter version in the pollster-ratings database. Because the database covers the final three weeks of the campaign, and because almost all polling firms publish likely-voter polls by that time, almost all polls in the database should be likely-voter surveys.
- When a pollster publishes multiple versions of the same survey (for example, versions of the poll with and without a third-party candidate included), FiveThirtyEight's policy is to average the versions together. However, some of the older polls in our database were taken from sources that may have followed different rules, so the treatment of these cases may be inconsistent.
- Polls of special elections and runoffs are included.
- In races that use an instant runoff, polls of all rounds of the race are included. Polls are evaluated based on the results of the round(s) they polled, if those results are published, or the results of the final round if a candidate got to 50 percent of the vote before all runoff rounds were calculated.
- Polls of [all-party primaries](#) (such as in Louisiana) are included.²
- National polls for the presidential popular vote and the generic congressional ballot are included.³
- The use of tracking polls is restricted to nonoverlapping dates. For instance, if a firm's final tracking poll was conducted on the Friday through the Sunday before an election, we wouldn't also list the version that covered Thursday through Saturday.
- Polls are included in the database even if they were not used in FiveThirtyEight's forecasts.⁴
- Although virtually all polls conducted in the final three weeks of a campaign are included, there are some exceptions in the case of the presidential primaries.
 - We exclude polls of the New Hampshire primary that are conducted before the Iowa caucus.

- We exclude polls of primaries in states beyond New Hampshire that are conducted before the New Hampshire primary.
- We exclude primary polls whose leader or runner-up dropped out before that primary was held.
- We exclude primary polls if any candidate receiving at least 15 percent in the poll dropped out before that primary was held.
- We exclude primary polls if any combination of candidates receiving at least 25 percent in the poll dropped out before that primary was held.

One challenge comes in how to identify which pollster we associate with each survey. For instance, Fabrizio, Lee & Associates and Impact Research began a partnership to conduct surveys for the Wall Street Journal in late 2021.

Theoretically, these could be classified as polls conducted by Fabrizio, Lee & Associates, Impact Research, the Wall Street Journal or some combination thereof. Our policy is to classify polls based on the pollster that conducted them, regardless of sponsorship, so these surveys are attributed to the partnership “Fabrizio, Lee & Associates/Impact Research.”

However, a few media companies have in-house polling operations. Confusingly, media companies sometimes *also* act as the sponsors of polls conducted by other firms. Our goal is to associate the poll with the company that, in our estimation, contributed the most intellectual property to the survey’s methodology. In some cases, this does include the media company that funded the poll. This is why, for example, The New York Times/Siena College is listed as a separate pollster from regular old Siena College.

When the same pollster or polling team operates multiple companies with different names but the same polling methodology, their polls are evaluated together. This is why, on some pollsters, you may see alternative names, indicating other companies operated by the same principal researchers or previous branding for that pollster’s work.

Step 2: Calculate simple average error

This part’s really simple: We compare the margin in each poll against the actual margin of the election and see how far apart they were. If the poll showed the Republican leading by 4 percentage points and they won by 9 instead, the poll’s

simple error was 5 points. We draw election results from officially certified state or federal sources.⁵

Simple error is calculated based on the margin separating the top two finishers in the election — not the top two candidates in the poll. For instance, if a certain poll of the 2008 Iowa Democratic caucus showed Hillary Clinton at 32 percent, Barack Obama at 30 percent and John Edwards at 28 percent, we'd look at its margin between Obama and Edwards since they were the top two finishers in the election (Clinton narrowly finished third).

We then calculate a simple average error for each pollster based on the average of the simple error of all its polls. This average is calculated using [root-mean-square error](#).

Step 3: Calculate Simple Plus-Minus

Some elections are more conducive than others to accurate polling. In particular, polls of presidential general elections are historically quite accurate, while presidential primaries are much more challenging to poll. Polls of general elections for Congress and for governor are somewhere in between.

This step seeks to account for that fact, along with a couple of other factors. We run a regression analysis that predicts polling error based on the type of election surveyed, a poll's margin of sampling error and the number of days⁶ between the poll and the election.⁷

We then calculate a Simple Plus-Minus score for each pollster by comparing its simple average error against the error one would expect from these factors. For instance, suppose a pollster has a simple average error of 4.6 points. By comparison, the average pollster, surveying the same types of races on the same dates and with the same sample sizes, would have an error of 5.3 points according to the regression. Our pollster therefore gets a Simple Plus-Minus score of -0.7. This is a good score: As in golf, negative scores indicate better-than-average performance. Specifically, it means this pollster's polls have been 0.7 points more accurate than other polls under similar circumstances.

A few words about the other factors Simple Plus-Minus considers. In the [past](#), we've described the error in polls as resulting from three major components: sampling error, temporal error and pollster error (or "pollster-induced error"). These are related by a sum of squares formula:

$$\text{Total Error} = \sqrt{\text{Sampling Error} + \text{Temporal Error} + \text{Pollster Error}}$$

Sampling error reflects the fact that a poll surveys only some portion of the electorate rather than everybody. This matters less than you might expect; theoretically, a poll of 1,000 voters will miss the final margin in the race by an average of only about 2.5 points because of sampling error alone — even in a state with 10 million voters.⁸ Unfortunately, sampling error isn't the only problem pollsters have to worry about.

Another concern is that polls are (almost) never conducted on Election Day itself. We refer to this property as **temporal (or time-dependent) error**. There have been elections when important news events occurred in the 48 to 72 hours that separated the final polls from the election, such as the New Hampshire Democratic presidential primary debate in 2008.

If late-breaking news can sometimes affect the outcome of elections, why go back three weeks in evaluating pollster accuracy? Well, there are a number of considerations we need to balance against the possibility of last-minute shifts in the polls:

- The overwhelming majority of elections do not feature important late-breaking developments. There will often be head-fakes and media-hyped “game changers,” but the evidence suggests [they rarely make much difference](#).
- Herding (see below) becomes more prominent in the final few days before an election. It's fairly common for a pollster to publish some wild-seeming results earlier in the cycle — which can affect media coverage of the campaign — only to “fall in line” with its final poll.
- Some of the apparent movement in the polls in the late days of the election is [probably artificial](#), reflecting response bias (i.e., voters for a certain candidate might be more likely to respond to polls after the candidate has a strong news cycle) and badly designed turnout models rather than genuine changes in public opinion.
- “Election Day” is something of a misnomer. Most states allow people to vote by mail or early in person; in the 2022 Senate election in Arizona, for example, [over 80 percent of votes](#) were cast by early or mail-in ballot rather than at a polling place on Nov. 8.

- Accounting for all polls in the final three weeks of the campaign increases the sample size of polls we can analyze, making us much more confident in our evaluations.

Three weeks is an arbitrary cutoff point; we have found no significant difference between ratings based on polls conducted three, four or five weeks out from an election. But we feel strongly that evaluating a polling firm's accuracy based only on its very last poll before an election is a mistake.

Nonetheless, the pollster ratings account for the fact that polling on the eve of the election is slightly easier than doing so a couple of weeks out. So a firm shouldn't be at any advantage or disadvantage because of when it surveys a race.

The final component is **pollster error** (what we've referred to in the past as "pollster-induced error"); it's the residual error component that can't be explained by sampling error or temporal error. Certain things (like projecting turnout or ensuring a representative sample of the population) are inherently pretty hard. Our research suggests that even if all polls were conducted on Election Day itself (i.e., no temporal error) and took an infinite sample size (i.e., no sampling error), the average poll would still miss the final margin in the race by about 2 points.

However, some polling firms are associated with more of this type of error. That's what our Simple Plus-Minus scores seek to evaluate.

Step 4: Calculate Advanced Plus-Minus

In 2014, House Majority Leader Eric Cantor [lost the Republican primary in Virginia's 7th Congressional District](#) to David Brat, a college professor. It was a stunning upset, at least according to the polls. For instance, a [Vox Populi Polling/Daily Caller poll](#) had put Cantor ahead by 12 points. Instead, [Brat won by 11 points](#). The poll missed by 23 points.

According to Simple Plus-Minus, that poll would score very poorly. We don't have a comprehensive database of House primary polls and don't include them in the pollster ratings, but we'd guess that such polls are off by something like 10 points on average. Because the aforementioned poll missed by 23 points, it would get a Simple Plus-Minus score somewhere around +13.

That seems pretty terrible — until you compare it with the only other poll of the race, an [internal poll released by McLaughlin & Associates on behalf of Cantor's](#)

[campaign](#). That poll had Cantor up by 34 points — a 45-point error! If we calculated something called Relative Plus-Minus (how the poll stacks up against others of the same race), the Vox Populi/Daily Caller poll would get a score of -22, since it was 22 points more accurate than the McLaughlin & Associates survey.

Advanced Plus-Minus, the next step in the calculation, seeks to balance these considerations. Advanced Plus-Minus is a combination of Relative Plus-Minus and Simple Plus-Minus, weighted by the number of other polling firms that surveyed the same race (let's call this number n). Relative Plus-Minus gets the weight of n , and Simple Plus-Minus gets a weight of three. For example, if six other polling firms surveyed a certain race, Relative Plus-Minus would get two-thirds of the weight and Simple Plus-Minus would get one-third.

In other words, when there are a lot of polls in the field, Advanced Plus-Minus is mostly based on how well a poll did in comparison to the work of other pollsters that surveyed the same election. But when there is scant polling, it's mostly based on Simple Plus-Minus.

Meticulous readers might wonder about another problem. If we're comparing a poll against its competitors, shouldn't we account for the strength of the competition? If a pollster misses every election by 40 points, it's easy to look good by comparison if you happen to poll the same races it does. The problem is similar to the one you'll encounter if you try to design college football or basketball rankings: Ideally, you'll want to account for the strength of a team's schedule in addition to its wins and losses and margins of victory. Advanced Plus-Minus addresses this by means of iteration (see a good explanation [here](#)), a technique commonly applied in sports power ratings.

Advanced Plus-Minus also addresses another problem. Polls tend to be more accurate when there are more of them in the field. This may reflect herding, selection bias (pollsters may be more inclined to survey easier races; consider how many of them [avoided the Kansas gubernatorial race](#) in 2022) or some combination thereof. So Advanced-Plus Minus also adjusts scores based on how many other polling firms surveyed the same election. This has the effect of rewarding polling firms that survey races few other pollsters do and penalizing those that swoop in only after there are already a dozen polls in the field.

Two final wrinkles. Advanced Plus-Minus puts slightly more weight on more recent polls.⁹ It also contains a subtle adjustment to account for the higher

volatility of certain election types, especially presidential primaries.¹⁰

Step 5: Calculate Predictive Plus-Minus

If you're interested in a purely retrospective analysis of poll accuracy, Simple Plus-Minus and Advanced Plus-Minus can be useful. You'll also find a number of other measures of historical accuracy in [our pollster-ratings database](#). The version we'd personally recommend is called "Mean-Reverted Advanced Plus-Minus," which is retrospective but discounts the results for pollsters with a small number of polls in the database.

However, that may not be your purpose. At FiveThirtyEight, we're more interested in predicting which polling firms will be most accurate going forward. This is useful to know if [you're using polls to forecast election results](#), for example. For that purpose, we use a measure called Predictive Plus-Minus.

The difference with Predictive Plus-Minus is that it also accounts for a polling firm's methodological standards — albeit in a slightly roundabout way. A pollster gets a boost in Predictive Plus-Minus if it is a member of the [American Association for Public Opinion Research's Transparency Initiative](#) or contributes polls to the [Roper Center for Public Opinion Research's archive](#). Participation in these organizations is a [proxy variable](#) for methodological quality. That is, it's a correlate of methodological quality rather than a direct measure of it.¹¹

We've previously [discussed at length](#) the [value](#) of including this sort of methodological component in our pollster ratings. In every cycle we have evaluated, pollsters that participate in professional organizations such as these¹² have performed significantly better than pollsters that do not.

But let's say you have one polling firm that passes our methodological tests but hasn't been so accurate, and another that doesn't meet the methodological standards but has a reasonably good track record. Which one should you expect to be more accurate going forward?

That's the question Predictive Plus-Minus is intended to address. But the answer isn't straightforward; it depends on how large a sample of polls you have from each firm. Our finding is that past performance reflects more noise than signal until you have about 30 polls to evaluate, so you should probably go with the firm with the higher methodological standards up to that point. If you have more than

30 polls from each pollster, however, you should tend to value past performance over methodology.

One further complication is “[herding](#),” or the tendency for polls to produce very similar results to other polls, especially toward the end of a campaign. A methodologically inferior pollster may be posting superficially good results by manipulating its polls to match those of the stronger polling firms. If left to its own devices — without stronger polls to guide it — it might not do so well. When we [looked at Senate polls from 2006 to 2013](#), we found that methodologically poor pollsters improve their accuracy by roughly 2 points when there are also strong polls in the field. As a result, Predictive Plus-Minus includes a “herding penalty” for pollsters that show too little variation from the average of previous polls of the race.

The full formula for how to calculate Predictive Plus-Minus has evolved over the years. The formula we currently use is as follows:

$$PPM = \frac{\max(-2, APM + \textit{herding_penalty}) \times (\textit{disc_pollcount}) + \textit{prior} \times 18}{18 + \textit{disc_pollcount}}$$

The variables *PPM* and *APM* are Predictive Plus-Minus and Advanced Plus-Minus, of course. The variable *disc_pollcount* is the discounted poll count, in which older polls receive a lower weight than more recent polls.^{[13](#)} The exact values used to calculate *prior* change every time we update the pollster ratings,^{[14](#)} but it is currently calculated as $0.66 - \textit{quality} \cdot 0.57 + \min(18, \textit{disc_pollcount}) \cdot -0.03$. The variable *quality* is the methodology boost discussed above; it has a value of 1 if the pollster meets the AAPOR/Roper transparency standard and 0 if it doesn’t.

Finally, to calculate *herding_penalty*, we start by calculating how much the pollster’s average poll differs from the average of previous polls of that race — specifically, polls whose median field date was at least three days earlier.^{[15](#)} This is a pollster’s Average Distance from Polling Average (ADPA), and it’s also a column in [our pollster-ratings database](#). The herding penalty for each pollster is one-half of the difference between a pollster’s actual ADPA and its theoretical minimum ADPA based on sampling error (both of the pollster’s polls and the polling average they’re being compared with).

Basically, Predictive Plus-Minus is a version of Advanced Plus-Minus in which scores are reverted toward a mean, where the mean depends on both the methodological quality of the pollster and the recency of its polls. The fewer recent polls a firm has, the more its score is reverted toward this mean. So Predictive Plus-Minus is mostly about a poll's methodological standards for firms with only a few recent surveys in the database, and mostly about its past results for those with many recent surveys.

Step 6: Convert Predictive Plus-Minus into a letter grade

As a final step, we've translated each firm's Predictive Plus-Minus rating into a letter grade, from A+ to D-. One purpose of this is to make clear that the vast majority of polling firms cluster somewhere in the middle of the spectrum; about 83 percent of polling firms receive grades in the B or C range.¹⁶ Another, of course, is to make the ratings intuitive; not everyone reads entire methodology statements and understands what a "+0.76 Predictive Plus-Minus" means in practical terms.

Finally, there are a couple circumstances in which we'll assign a pollster a rating that isn't a simple letter grade between A+ and D-. First, pollsters that are [banned by FiveThirtyEight](#) automatically receive a grade of F. There is no Predictive Plus-Minus bad enough that it merits an F grade; if a pollster is rated "F," that means it did something much worse than simply being bad at polling. (The most common reason we ban pollsters is that we know or suspect they [fabricated data](#), but we also ban pollsters for [engaging in betting markets](#) that may be directly impacted by their survey work.)

And second, pollsters with a relatively small sample of polling get a provisional rating rather than a precise letter grade. An "A/B" provisional rating means that the pollster has shown strong initial results, a "B/C" rating means it has average initial results and a "C/D" rating means below-average initial results. It takes roughly 20 recent polls (or a larger number of older polls) for a pollster to get a precise pollster rating.

Editor's note: This article is adapted from a [previous article](#) about how our pollster ratings are calculated.

ratings creators

Development by [Aaron Bycoffe](#). Research by [Mary Radcliffe](#), [Cooper Burton](#) and [Dhruvil Mehta](#). Copy editing by [Andrew Mangan](#). Statistical model by [Nate Silver](#).

version history

1.0

First formal methodology.

related articles

2008 pollster ratings

2010 pollster ratings

2014 pollster ratings

2016 pollster ratings

2018 pollster ratings

2019 pollster ratings

2020 pollster ratings

2021 pollster ratings

2023 pollster ratings

Footnotes

1. For example, if a certain polling firm usually conducts 800-person polls, we'd use that as the sample size in a case where it was unreported.
2. We include only polls for elections where a winner can be determined by the outcome. So, for example, we include polls for Louisiana's all-party primary but not for California's all-party primary, where the top two candidates must compete in a runoff election regardless of the primary outcome.
3. The accuracy of generic-congressional-ballot polls is evaluated by comparing them to the total number of votes cast for each party's candidates, calculated as a sum of the election results in each district.
4. This may occur if a pollster has been banned by FiveThirtyEight, or if a poll was released after we froze our models the night before the election.
5. In the rare case that we are unable to locate officially certified results from a particular state, we use [Dave Leip's Atlas of U.S. Presidential Elections](#) as an alternative source for election results.

6. In the regression, this is specified as the square root of the number of days between the poll's median field date and the election; the relationship between the time a poll is conducted and its accuracy is slightly nonlinear. For presidential primary polls, a separate coefficient is used for timing, since it is more important in those races.
7. As a control, the regression also includes a dummy variable for the type of election. The purpose of this is to check whether differences in polling accuracy between different types of elections are the result of the mix of polling firms that happen to survey those races rather than factors intrinsic to the races themselves. This doesn't make much difference — but the modest exception is polls for the U.S. House, which have historically been less accurate than polls of Senate and gubernatorial races. The regression analysis suggests most of this difference is the result of worse pollsters tending to survey House races — in particular, the proportion of partisan polls is much higher in House races.
8. The reason to decompose this factor from other sources of polling error is that pollsters sometimes vary their sample sizes. If a certain polling firm had been getting inaccurate results because it conducted only 300-voter samples but then shifted to using 1,500-voter samples instead, you'd expect its results to get better. FiveThirtyEight's models account for a poll's sample size, so it's helpful to know how much this contributed to a poll's accuracy (or lack thereof) compared with other factors.
9. The weights are a decaying exponential function of the year in which the poll was conducted and are calculated as $0.93^{(cycle-year)}$, where *cycle* indicates the most recent election cycle that we are evaluating and *year* indicates the year in which the poll was conducted.
10. Polls of presidential primaries are associated not only with a higher average error, but also with a higher standard deviation in the error term. Advanced Plus-Minus normalizes the error term so that it's equal across different types of races.
11. AAPOR's Transparency Initiative vets applicants, and both AAPOR and Roper require pollsters to meet certain disclosure and professional standards. This vetting process, along with self-selection in which firms choose to participate in these groups, tends to screen out firms with poorer methodological standards.
12. Previously, we also gave pollsters the same extra credit for belonging to the National Council on Public Polls, but the NCPP is no longer an active organization, and our pollster ratings stopped considering it in 2023.
13. Each poll's weight in the count is calculated as $0.93^{(cycle-year)}$, where *cycle* is the most recent election cycle that we are evaluating and *year* is the year of the poll.
14. Specifically, the coefficients are optimized.
15. The average includes only the most recent poll from each pollster in each race and excludes partisan polls and pollsters banned by FiveThirtyEight. It is weighted based on the square root of the number of other polls in the field for each race.
16. This also includes pollsters that receive a provisional grade of B/C but not A/B or C/D. The percentage is calculated excluding banned pollsters, who receive a grade of F in our pollster ratings.

