

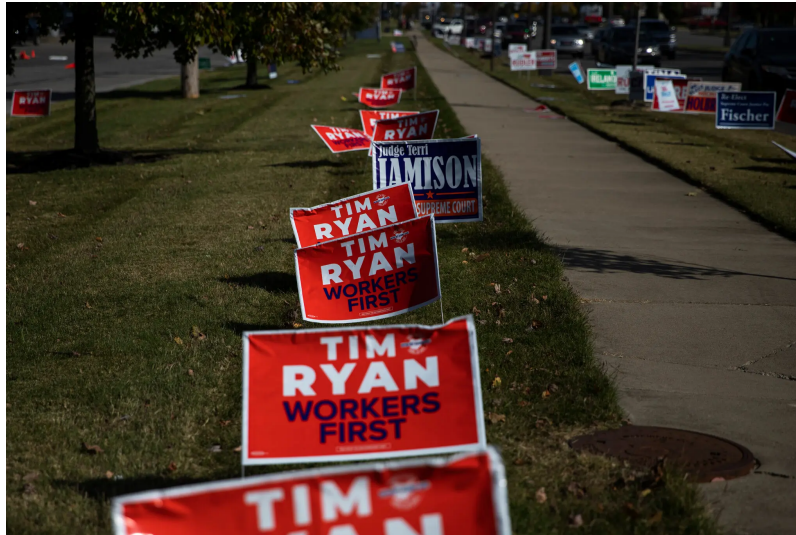
NYT/Siena Poll Is Latest to Show Republican Gains

Is four points the real margin nationally? That's a good question.



By Nate Cohn

Oct. 17, 2022, 5:00 a.m. ET



The path to keeping the House and the Senate appears to be getting tougher for Democrats, according to the most recent polling. Campaign signs for the Ohio Democratic Senate candidate Tim Ryan in Columbus on Friday. Maddie McGarvey for The New York Times

We have the result of our third New York Times/Siena College national survey of the midterm cycle to go with your coffee this morning: 49 percent of voters say they back the Republican congressional candidate in their district, compared with 45 percent backing the Democratic one.

It's a modest but notable swing from last month, when Democrats led by one percentage point among likely voters. Since then, the warning signs for Democrats have begun to add up, including Republican polling gains in key Senate races like those in Nevada, Wisconsin and Pennsylvania, and surprising Republican strength in districts in Rhode Island and Oregon where Democrats would normally be safe.

Up to this point, Democrats have maintained a narrow lead in polls asking whether voters prefer Democrats or Republicans for Congress, but there have been warning signs for the party here as well. Republicans have led in several high-quality polls, like ABC/Washington Post, CBS/YouGov and Monmouth University. Today, the Times/Siena survey adds a fourth such poll to the pile.

The evidence for a shift toward Republicans appears to be underpinned by a change in the national political environment. Gas prices went up again. The stock market is down. A variety of data suggests that the electorate's attention is shifting back to issues where Republicans are on stronger ground in public opinion, like the economy, inflation, crime and immigration, and away from the summer's focus on democracy, gun violence and abortion, where Democrats have an edge.

In other words, the conditions that helped Democrats gain over the summer no longer seem to be in place.

Is four points the real margin? (Wonkiness 4/10)

Our poll may show Republicans ahead, 49-45, and yet it may not be accurate to say they lead by four points. In fact, they actually lead by three points.

How is this possible? Rounding. By convention, pollsters round the results to the nearest whole number. In this poll, the exact unrounded figures are 48.51 (rounding to 49) to 45.47 (rounding to 45). That's a three-point lead.

This is not at all uncommon. In 2020 polling, about one-third of our reported margins, based on the difference between two rounded vote shares of the candidates, were different from what our reported margin would have been if we had rounded once on the difference between exact figures.

It's not even the only example in today's poll. Our result among registered voters is reported as a 46-46 percent tie, but Republicans lead, 46.2 to 45.6. If we reported it as a rounded margin, this might count as a one-point Republican lead.

The two rounding errors add up to an even larger disparity between the reported and actual result when it comes to the difference between likely and registered voters. The rounded result makes it seem to be a four-point gap. In reality, the difference is a 2.5-point gap.

This is a polling custom that has always left me a little cold. The case for rounding is straightforward: Reporting results to the decimal point conveys a false sense of precision. After a decade of high-profile polling misfires, "precision" is most certainly not the sense pollsters want to try to convey right now. And in this case, reporting to the one-thousandth of a point would obviously be ridiculous. We didn't even contact a thousand people; how could we offer a result to the one-thousandth?

But there's a trade-off. Characterizing this poll as a four-point Republican lead doesn't merely offer a false sense of precision — it's just false. That's not something I can gloss over.

Sometimes, the difference is enough to affect the way people interpret the poll. We've reported one party in the "lead" by one percentage point when, in fact, the figures are essentially even. These differences don't actually mean much, of course, but no one — not even those of us well versed in statistics and survey methodology — can escape perceiving a difference between R+1 and Even.

Hopefully, the differences between the rounded and exact results can be another reminder that polling is inexact. The results are fuzzy, and the margin of error understates the actual degree of uncertainty anyway. If we had called another 100 people, or did another round of callbacks, the results would almost certainly have been at least somewhat different, and the same if we had called another 100 more.

There isn't a great way out of this problem. We can refrain from characterizing a 49-45 lead as a four-point lead, as the main Times article on this poll does today. But even this requires us to notice when the rounded and actual margins differ, which is not easy when our crosstabs and other products use rounded numbers. When we do notice a difference worth your attention, we'll try to flag it here and elsewhere.

What's different about our polls this year?

A few weeks ago, I noted that most pollsters this cycle weren't making big methodological changes. Instead, they're doing something more like tightening the screws on an old boat after a rough storm in 2020, rather than going out and buying a new boat.

But I didn't actually mention what screws we've tightened this cycle. Here's a quick summary:

We're weighting on method of voting in 2020 — whether people voted by mail, early or absentee. It's an important predictor of vote choice, even after considering the partisanship of a registrant. Registered Republicans who voted on Election Day, for instance, were more likely to back Donald J. Trump than those who voted by mail. Weighting on this in 2020 wouldn't have made a major difference, but it would have brought some of our polls about half a point or so closer to the final result.

We now use additional information about the attitudes of respondents in determining whether they're likely to vote, including whether respondents are undecided; whether their views about the president align with their party; whether they like the candidate they intend to vote for; whether they back the party out of power in a midterm; and so on, all based on previous Times/Siena polls. At the same time, **we now give even more weight to a respondent's track record of voting** than we did in the past.

We're changing how we characterize people who attended trade or vocational school but did not receive a college degree (Wonkiness rating: 6.5/10). The effect is a slight increase in the weight given to Republican-leaning voters without any post-high-school training, and a decrease in the weight given to the somewhat fewer Republican voters who attended some college or received an associate degree.

This is a little complicated. Basically, pollsters need to decide whether people who went to technical or vocational school count as "high school graduates" or "some college" when they're adjusting their surveys to make sure they have the right number of voters by educational group. They have to choose, because the Census Bureau doesn't count a trade or vocational school as a level of educational attainment. In the view of the Census Bureau, that puts them in the category of high school graduates. The Times/Siena poll (and many other pollsters) previously counted them the same way.

But this choice isn't necessarily straightforward. Whether it's the right choice in practice depends on whether census interviewers and respondents handle this question the way the census would like. If you completed a professional technical program at, say, Renton Technical College, there's a chance you selected one of the various "some college" options on the census American Community Survey or the Current Population Survey.

I'd like to run an experiment on this at some point, but for the moment we're moving respondents like these into the "some college" category. By doing so, we modestly increase the weight we give to those categorized as high school graduates (who are pretty Republican), and decrease the weight on the other group (who still lean Republican but somewhat less so). Unfortunately, had we done this, it would have improved our result by only about a quarter of a point in 2020 — despite the number of words I just dedicated to the topic.

On a totally different topic, we now consider the source of cellphones in determining whom we'll call (Wonkiness rating: 8/10). This is the last point in this newsletter, so you can go on with the rest of your day if your eyes are glazing over, but I think it might be the most interesting to a subset of you, especially those who conduct polls.

As I've mentioned before, we get the telephone numbers for our poll off a list of registered voters called a voter file. The telephone numbers on the voter file can have two different sources: those provided by the registrant on their voter registration form (which then wind up on the file), and those matched by L2 (our voter file vendor) from an outside source.

The voter-provided cellphone numbers are the likeliest to yield a completed interview. They almost always lead to the person we're looking for (we complete interviews only with the people named on the file). And the people contacted are a little more likely to cooperate, too. But the voter-provided cellphone numbers are almost exclusively from people who registered over the last 10 years (after all, if you registered 20 years ago, you probably didn't have a cellphone number). As a result, they're relatively, young, liberal, less likely to be married, less likely to own a home, and so on.

The externally matched cellphone numbers are less likely to yield a completed interview. They're less likely to belong to the people we try to reach; these people may also be less likely to take the survey, even when we're reaching the right ones. They're somewhat more representative of the population as a whole because anyone, regardless of when they registered, could plausibly have one of these numbers.

Why does this matter? When we treated all cellphone numbers the same, we were systematically reaching fewer people who were older or married or homeowners — people more likely to have registered long ago. We called these groups in the right proportion, but we would wind up with fewer completed interviews from groups like this with more externally matched numbers.

Now, we're accounting for whether different demographic groups — like new or previously registered voters — have more self-reported phone numbers or externally matched phone numbers. As a result, we'll dial more people from the groups with relatively high numbers of externally matched numbers.

It's hard to know how our 2020 polls would have been different if we had used this year's approach. After all, we would have reached a different set of respondents. Most of the analyses I've conducted suggest that the respondents with self-reported or externally matched numbers aren't very different politically, controlling for the characteristics we're using in weighting. I'd note, though, that this would have been a real problem if we hadn't been weighting on homeownership or marital status. Most pollsters using voter file data aren't doing so; it might be worth looking at.

/