

We Gave Four Good Pollsters the Same Raw Data. They Had Four Different Results.

By NATE COHN SEPT. 20, 2016

How four pollsters, and The Upshot, interpreted 867 poll responses:



You’ve heard of the “margin of error” in polling. Just about every article on a new poll dutifully notes that the margin of error due to sampling is plus or minus three or four percentage points.

But in truth, the “margin of sampling error” – basically, the chance that polling different people would have produced a different result – doesn't even come close to capturing the potential for error in surveys.

Polling results rely as much on the judgments of pollsters as on the science of survey methodology. Two good pollsters, both looking at the same underlying data, could come up with two very different results.

How so? Because pollsters make a series of decisions when designing their survey, from determining likely voters to adjusting their respondents to match the demographics of the electorate. These decisions are hard. They usually take place behind the scenes, and they can make a huge difference.

To illustrate this, we decided to conduct a little experiment. On Monday, in partnership with Siena College, the Upshot [published a poll](#) of 867 likely Florida voters. Our poll showed Hillary Clinton leading Donald J. Trump by one percentage point.

We decided to share our raw data with four well-respected pollsters and asked them to estimate the result of the poll themselves.

Here’s who joined our experiment:

- **Charles Franklin**, of the [Marquette Law School Poll](#), a highly regarded public poll in Wisconsin.
- **Patrick Ruffini**, of [Echelon Insights](#), a Republican data and polling firm.
- **Margie Omero, Robert Green and Adam Rosenblatt**, of [Penn Schoen Berland Research](#), a Democratic polling and research firm that conducted surveys for Mrs. Clinton in 2008.
- **Sam Corbett-Davies, Andrew Gelman and David Rothschild**, of Stanford University, Columbia University and Microsoft Research. They’re at the forefront of using statistical modeling in survey research.

Here’s what they found:

Pollster	Clinton	Trump	Margin
Charles Franklin Marquette Law	42%	39%	Clinton +3%
Patrick Ruffini Echelon Insights	39%	38%	Clinton +1%
Omero, Green, Rosenblatt Penn Schoen Berland Research	42%	38%	Clinton +4%
Corbett-Davies, Gelman, Rothschild Stanford University/Columbia University/Microsoft Research	40%	41%	Trump +1%
NYT Upshot/Siena College	41%	40%	Clinton +1%

Well, well, well. Look at that. A net five-point difference between the five measures, including our own, even though all are based on identical data. Remember: There are no sampling differences in this exercise. Everyone is coming up with a number based on the same interviews.

Their answers shouldn't be interpreted as an indication of what they would have found if they had conducted their own survey. They all would have designed the survey at least a little differently – some almost entirely differently.

But their answers illustrate just a few of the different ways that pollsters can handle the same data – and how those choices can affect the result.

So what’s going on? The pollsters made different decisions in adjusting the sample and identifying likely voters. The result was four different electorates, and four different results.

Pollster	Result	White	Hisp.	Black	Sample
Charles Franklin Marquette Law	Clinton +3	68%	15%	10%	+5 Dem.
Patrick Ruffini Echelon Insights	Clinton +1	67%	14%	12%	+1 Dem.

Pollster	Result	White	Hisp.	Black	Sample
Omero, Green, Rosenblatt Penn Schoen Berland Research	Clinton +4	65%	15%	12%	+4 Dem.
Corbett-Davies, Gelman, Rothschild Stanford University/Columbia University/Microsoft Research	Trump +1	70%	13%	14%	+1 Rep.
NYT Upshot/Siena College	Clinton +1	69%	14%	12%	+1 Rep.

There are two basic kinds of choices that our participants are making: one about adjusting the sample and one about identifying likely voters.

How to make the sample representative?

Pollsters usually make statistical adjustments to make sure that their sample represents the population – in this case, voters in Florida. They usually do so by giving more weight to respondents from underrepresented groups. But this is not so simple.

What source? Most public pollsters try to reach every type of adult at random and adjust their survey samples to match the demographic composition of adults in the census. Most campaign pollsters take surveys from lists of registered voters and adjust their sample to match information from the voter file.

Which variables? What types of characteristics should the pollster weight by? Race, sex and age are very standard. But what about region, party registration, education or past turnout?

How? There are subtly different ways to weight a survey. One of our participants doesn’t actually weight the survey in a traditional sense, but builds a statistical model to make inferences about all registered voters (the same technique that yields our pretty dot maps).

Who is a likely voter?

There are two basic ways that our participants selected likely voters:

Self-reported vote intention Public pollsters often use the self-reported vote intention of respondents to choose who is likely to vote and who is not.

Vote history Partisan pollsters often use voter file data on the past vote history of registered voters to decide who is likely to cast a ballot, since past turnout is a strong predictor of future turnout.

Our participants’ choices

The participants split across all these choices.

Pollster	Who is Likely Voter? Type of weight		Tries to match...
Charles Franklin Marquette Law	Self-report	Traditional	Census
Patrick Ruffini Echelon Insights	Vote history	Traditional	Voter File
Omero, Green, Rosenblatt Penn Schoen Berland Research	Self-report	Traditional	Voter File
Corbett-Davies, Gelman, Rothschild Stanford University/Columbia University/Microsoft Research	Vote history	Model	Voter File
NYT Upshot/Siena College	Report + history	Traditional	Voter File

Their varying decisions on these questions add up to big differences in the result. In general, the participants who used vote history in the likely-voter model showed a better result for Mr. Trump.

At the end of this article, we’ve posted detailed methodological choices of each of our pollsters. Before that, a few of my own observations from this exercise:

- These are all good pollsters, who made sensible and defensible decisions. I have seen polls that make truly outlandish decisions with the potential to produce even greater variance than this.
- Clearly, the reported margin of error due to sampling, even when including a design effect (which purports to capture the added uncertainty of weighting), doesn’t even come close to capturing total survey error. That’s why we didn’t report a margin of error in our original article.
- You can see why “herding,” the phenomenon in which pollsters make decisions that bring them close to expectations, can be such a problem. There really is a lot of flexibility for pollsters to make choices that generate a fundamentally different result. And I get it: If our result had come back as “Clinton +10,” I would have dreaded having to publish it.
- You can see why we say it’s best to average polls, and to stop fretting so much about single polls.

Finally, a word of thanks to the four pollsters for joining us in this exercise. Election season is as busy for pollsters as it is for political journalists. We’re grateful for their time.

Below, the methodological choices of the other pollsters.

Charles Franklin [Clinton +3](#)
Marquette Law

Mr. Franklin approximated the approach of a traditional pollster and did not use any of the information on the voter registration file. He weighed the sample to an estimate of the demographic composition of Florida’s registered voters in 2016, based on census data, by age, sex, education, gender and race. Mr. Franklin’s likely voters were those who said they were “almost certain” to vote.

Patrick Ruffini **Clinton +1**

Echelon Insights

Mr. Ruffini weighted the sample by voter file data on age, race, gender and party registration. He next added turnout scores: an estimate for how likely each voter is to turn out, based exclusively on their voting history. He then weighted the sample to the likely turnout profile of both registered and likely voters – basically making sure that there were the right number of likely and unlikely voters in the voter file. This is probably the approach most similar to the Upshot/Siena methodology, so it is not surprising that it also is the closest result.

Sam Corbett-Davies, Andrew Gelman and David Rothschild **Trump +1**

Stanford University/Columbia University/Microsoft Research

Long story short: They built a model that tries to figure out what characteristics predict support for Mrs. Clinton and Mr. Trump based on many of the same variables used for weighting. They then predicted how every person in the state would vote, based on that model. It's the same approach we used to make the pretty dot maps of Florida. The likely electorate was determined exclusively by vote history, not self-reported voice choice. They included 2012 voters – which is why their electorate has more black voters than the others – and then included newly registered voters according to a model of voting history based on registration.

Margie Omero, Robert Green, Adam Rosenblatt **Clinton +4**

Penn Schoen Berland Research

The sample was weighted to state voter file data for party registration, gender, race and ethnicity. They then excluded the people who said they were unlikely to vote. These self-reported unlikely voters were 7 percent of the sample, so this is the most permissive likely voter screen of the groups. In part as a result, it’s also Mrs. Clinton’s best performance. In an email, Ms. Omero noted that every scenario they examined showed an advantage for Clinton.