



INSIGHTS THAT COUNT™

A Market Researcher's Musings on Political Polling – Lessons Learned

On Wednesday, November 8, 2016, the emails, texts, and calls started. They came from friends, family members, co-workers, and clients. They had similar themes: *I thought the polls showed Clinton would win? What happened? What did the polls miss?* In the days that followed, the conversation shifted even more. I was getting emails from my research and insights clients about their executives wondering if market research is reliable. Political polling, and market research, were having a moment of reckoning.

At PSB, we took a step back to look at what exactly happened and how to course-correct. While there was a range of contributing factors to the 'misses' in 2016, the focus of my exploration was polling techniques to see what we as market researchers could and should learn. In the autopsy of 2016, three key themes emerged:

- 1. Manage expectations**
- 2. Sample size matters**
- 3. Be representative**

Manage expectations. Polls indicate the likelihood of an outcome, not an outcome. During the Clinton-Trump contest, many headlines presented an oversimplified narrative and an almost certain result – a Clinton presidency was inevitable. We know that was not the outcome and that is not the story that the actual data was showing.

We know that the national level polls were not "way off" but instead incorrectly positioned. Most national polls showed Clinton leading by low single digits (2 to 5 points). She ended up winning the national vote by just over two percentage points. Models, like FiveThirtyEight, showed that while Clinton had the advantage going into Election Day, Trump still had a decent shot.

Even most state polls were not as wrong as some of the headlines made it seem. In many cases, headlines and narratives overstated advantages or disadvantages for either candidate even while the data showed that the race was very close.

The question isn't "why were the polls way off in 2016" or "can we trust polls", rather we needed to be clear about what polls actually mean. We needed to better explain that even if Clinton is up a couple of points in Pennsylvania (even across multiple polls), there is a chance that Trump could still win that state.

How 2020 is different: Our industry is getting better at managing expectations and explaining what polls actually mean. Reporting and commentary around poll results are now more likely to use terms like "polling does not predict the future" or "polls are just a snapshot at that given time." Stories, and associated headlines, are doing a better job explaining context and nuance within the results. For example, the margin of error, instead of just being a footnote at the end of a story with some polling results, is fully explained or is visually shown on associated graphs or charts.

Sample size matters. In corporate market research, we are typically dealing with target audiences representing 5% or even less of the U.S. population. In some cases, we are finding needles in the haystack or unicorns. Voters, on the other hand, represent most of the adult population. While tens of millions sit out each election, most presidential elections get 55% to 60% voting.

Despite this substantial pond of potential survey respondents, too many polls in 2016 had small sample sizes. National polls with fewer than 500 respondents. State polls with 200 respondents – not sparsely populated states like Wyoming or North Dakota but states like Pennsylvania and Florida. Too often, these polls were being given equal footing with polls with significantly greater sample sizes. One had to dig into the fine print to see Poll A had a 200 respondents sample while Poll B had 1000.

Another element of size has to do with the number of polls. Some states were woefully under surveyed in 2016. There were not enough polls in the closing weeks to make an accurate assessment of one candidate's chances over another.

How 2020 is different: There are more and more polls with respondent numbers in the thousands, including at the state level. These larger sample sizes offer more confidence in the results. Instead of pulling back on polling, polling is as widely available as ever and there is an acknowledgement that we need to look at multiple polls per state to get an accurate view of that state.

Be representative. Getting the right sample composition can be very challenging – election-related polls need to predict and reflect the make-up of the electorate. Whether there are too few/many females or males, too many/few from the array of minority groups, over/under indexing on rural voters – these all impact which candidate leads as well as the margin of the lead. Determining which pollsters are doing quotas and weighting 'right' is challenging, especially since not every publically released poll share this information.

For years, we have understood the gender gap – men more likely voting Republican and women more likely voting Democrat. We knew that African-Americans heavily tilt Democrat and Hispanic voters favor Democrats. However, it is not as simple as weighting to 51% on Female and 49% on Male and getting 13% African-American. Each state is unique, and each election the electorate looks a little different.

Take Florida. The Hispanic vote is and will be a sizeable group of the electorate. Hispanics make up 17% of registered voters in Florida and simply weighting/quota-ing accordingly could skew the results. The two largest Hispanic groups in the state are of Puerto Rican and Cuban descent, and they tend to vote very differently. Too many Cubans in the sample could artificially favor the Republican. Too many Puerto Ricans could do the same for the Democrat. The sample composition needs to be nuanced even within these larger categories.

Looking at 2016, some pollsters appear to have taken a simple path to their sample – weighting at the all up level. While others were more rigorous with their sample composition making sure it was not just gender, age, race that were correct on their own, but that they had the right number of African-American males under the age of 35.

An analysis of 2016 polls showed a new, significant problem – education was a major vector for determining support for Clinton or Trump – many pollsters missed this. Trump did very well among Whites without a college degree. The states where they make up a majority or near majority of voters – Iowa, Wisconsin, Michigan, Ohio, and Pennsylvania – are states where Trump over-performed his polls.

Looking at 2018, a consistent balance of rural, suburban, and urban voters is also critical, especially in states like Pennsylvania and Ohio, where the state's population is fairly divided between the three.

One anecdote from the 2020 Election: I recently noticed a particular pollster had results that were "swinging" a bit more than others for Pennsylvania. Fortunately, they release their data in crosstab form. I found they were consistent in gender, age, race/ethnicity, and region of the state, and for education, they were controlling for race/ethnicity to get Whites without a college degree. However, I noticed that the percentage of the rural sample fluctuated from wave to wave, by +/- 7 points. In the waves with a higher percentage of rural voters, Trump was "doing better". Conversely, Trump's prospects went down when the percentage of rural voters was at its lowest. Controlling for that variable would tame the swings.

How 2020 is different: There are more opportunities for any over-zealous market researcher/data geek to dig into data. More polls are being transparent by providing the data in crosstab form. Education has definitely joined the ranks of gender, age, race/ethnicity, etc. on elements that are not just loosely controlled but have rigorous considerations. Most are being more thoughtful about their sample.

What I'm looking at in these final days

Four years later, many of us are having moments of déjà vu. But should we? Are the polls more accurate this time? Are we reporting on them with more efficacy? Did we learn and improve from 2016? Each morning I take my 15 or so minutes to catch up on the recently released polls as I try to prepare for the narrative we may have to deal with on November 4. For 2020, I am paying attention to three things.

- **Look beyond the headlines.** Many headlines continue to misrepresent the findings – skip the headlines and focus on the fuller stories and the details of the polls.
- **Ignore polls with low sample sizes.** Focusing on state polls, I like to see n=500 at a minimum, with over 1000 being preferred. You have a poll with n=300; I'm ignoring your results. I'm also looking at multiple polls per state to see trends, consistency, or outliers.
- **Pay attention to sample.** Dig into the sample composition and how the sample is handled from wave to wave. Could changes in the results be explained because the sample's composition is different or possibly because a given candidate is gaining support?

Ultimately, the only poll that really matters happens on Election Day. Go vote!

David James

PSB Insights