Surveying in Science: "Fixing" the 2016 Election Polls

February 2019

Sayem Sinha

Syracuse University

Syracuse, NY 13224,

USA

OAK RIDGE INSTITUTE
FOR SCIENCE AND EDUCATION

Background

HISTORY OF PRESIDENTIAL POLLS

While the "botched" (Siegel, 2016) opinion polling of the contentious 2016 presidential election may burn brightest in the recent memory of the American public, the incident was far from being the first occurrence of such a situation. In 1936, during the election between Republican Alfred Landon and Democrat Franklin Delano Roosevelt, a prominent national print publication at the time, *The Literary Digest* (now defunct), which had correctly predicted the outcome of the previous five presidential elections, published in their October 31st issue that Landon would win handily with a solid 370 electoral votes (Squire, 1988). FDR won that particular election 62% to Alf Landon's 38%. With 46 out of 50 states in the union, it was a veritable landslide (Squire, 1988). The *Literary Digest* poll had been based on a survey of approximately 10 million questionnaires mailed to readers and potential readers, with a response set of 2.3 million (Squire, 1988). With such a large sample size, how could the magazine have gotten it wrong? Presumably it wasn't their survey design; using the same or similar questions, they'd correctly called the previous five. For them to have a poorly designed survey and still correctly call 5 elections in a row would be a stunning coincidence. So what could have went wrong? Well, regardless of how the poll itself was structured, it was very relevant there was an issue with who actually answered that poll. When you examine their sampling method, it becomes clear it was based on a non-representative sample. The *Literary Digest* poll sampled its respondent population from telephone directories, auto registrations, and club memberships in addition to its own subscriber list (Squire, 1988). Assuming that the *Literary Digest* had a primarily wealthy readership at the time (namely people who could afford discretionary spending like a magazine subscription at the height of the Great Depression), and coupled with the fact that in the year 1936, there was a certain and very specific class of people who might own a telephone and or an automobile or a membership in a country club, it's should not be surprising that their sample skewed rich and Republican and thus missed the working class Democrat FDR wave. This poll remains to this day as one of America's most infamous examples of non-representative sampling in poor polling methodology.

A similar albeit distinct example from FDR's successor Harry S. Truman's own election was the infamous Chicago Tribune "Dewey defeats Truman" premature headline fiasco. Representative of bad survey methodology in its own but different way, the case was a situation where the at the time Republican leaning newspaper, no big fan of the Democratic presidential incumbent, prematurely overreached in attempt to get a story out first and published the erroneous headline before the votes had been called; meaning it was purely based on pre-election polling and political punditry. (Cosgrove, 2014) As Tim Jones puts it in a 2007 Chicago tribune article about the event, "Like most newspapers, the Tribune…was lulled into a false sense of security by polls that repeatedly predicted a Dewy victory….the polls and the pundits left no room for doubt: Dewey was going to defeat President Harry S. Truman. And the Tribune would be the first to report it" (Jones, 2007). The actual results of the election of 1948 went a lot different than the Tribune's headlines; Dewey lost 189 to Truman's 303 in the Electoral College.

Flash forward to the 21st century. On the eve of the 2016 presidential election, the projected election forecasts from a select few of the nation's most credible pollsters looked like this: Nate Silver's 538 projected a 71% Clinton victory, NYT said 84%, Princeton Election Consortium had her at 95-99% (The New York Times, 2017). As for the actual results, Clinton won the popular vote by two percentage points, and President Trump received 304 electoral votes, making him the current president of the country. According to Stanford Professor Jon Krosnick, who wrote in a 2016 Washington Post article "The final high quality scientific…national polls consistently overestimated Clinton's share of votes by 3 or 4 percentage points…that's a systematic error but not huge" (Clement, 2017). A later correction by the Washington Post in 2017 say "the average of national polls overestimated Clinton's share of the two-party vote by about one percentage point". (Clement, 2017)

SURVEY DEFINITIONS AND VALIDITY

Before any in depth discussion of what those results, in context, mean and how to correctly interpret them and what they say about the reliability and the validity of surveys as a measurement instrument for data collection, there are a few operational definitions regarding surveys that need to be established. Surveys are a form of applied research; they're conducted with a practical problem in mind, and researchers are working in a real-world context (Morling, 2017). Straw polls are a specific type of survey that seek to collect data regarding public opinion in an electorate, and they seek to answer the question "Who is most likely to win the upcoming election?" Surveys are a specific kind of measurement designed to collect data to evaluate frequency claims; that is, claims describing a particular rate of a single variable (Morling, 2017). In this case, a straw poll would be collecting data to support the frequency claim about the number of people who engaged in some activity, i.e. voting for a particular candidate in the upcoming presidential election. Claims are arguments journalists or researchers are trying to make, based on empirical research and evidence (Morling, 2017). Systematic observations (data) such as the responses from a straw poll, can be used to test and refine theories and claims. Based on data collected from 2016 straw polls, journalists and researchers attempted to make claims that a certain percentage of Americans would vote for a specific candidate, such as Clinton or Trump.

In order to establish how useful surveys/straw polls are as a measure, we need to evaluate those claims using the lens of validities, of which there are four major types. Validity is a measure of the appropriateness of a conclusion or decisions (Morling, 2017). Valid claims are reasonable, accurate and justifiable (Morling, 2017). When we interrogate frequency claims, we examine them through the lenses of construct and external validity (Morling, 2017). Construct validity is a measure of how well conceptual variables are operationalized (Morling, 2017). When evaluating questions, we want to know how well the researchers (in this case, national pollsters) measured the variables (who a given percentage of Americans was going to vote for). Straw polls stand up quite well to the test of construct validity; each variable is reliably measured by questions like "who will you vote for?" Because such questions have high construct validity, we can be confident that straw polls accurately measure election outcomes, at least in terms of the population sampled by the survey. But how generalizable are these results? We can test

generalizability by applying a check of external validity. How did the pollsters choose straw poll participants, and how well do those participants represent the intended population of interest (in this case, American citizens of voting age)? External validity is thus a measure how well the results of the survey generalize to representative population contests besides just those included in the original study. Surveys in particular are a type of self-reporting data collection method (Morling, 2017). Participants report on their own behaviors, beliefs and attitudes by filling out a carefully constructed questionnaire.

SURVEY DESIGN

Surveys consist of different types of questions designed to collect different types of data; depending on their wording, you can end up with both qualitative and quantitative data from your responses. At the most basic level, questions are either open-ended or forced choice (Morling, 2017). Open ended allow a participant to self-report however they choose to respond, but forced-choice only allows for a select few predetermined response options. The general philosophy behind surveys is that they should start open-ended and general, get more specific by moving into forced choice, and finish off with demographic questions. Although not all surveys will be designed this way, such a survey design is likely to increase response rate. Open-ended questions mostly provide qualitative data, while forced choice questions provide quantitative; however, they can also collect categorical data too, depending on how they're worded. Useful formats for forced choice questions to collect diverse data are categorical questions (like multiple choice on a standardized test), quantitative questions (the same but with numbers instead of categorical data), semantic differentials (scales that provide data regarding attitudes, like agreement or disagreement with a construct), and Likert type scales (e.g. a scale of 1-5) (Morling, 2017). These are not the only types of questions a survey may ask, although they are ideal ones. The following are questions which are not ideal, and should be excluded from survey design.

PROBLEMATIC QUESTION TYPES

- **Leading:** Questions that are worded in such a way as to persuade the reader to one or another side of an argument or position
- **Loaded:** Questions worded in such a way that they force the respondent into an answer that doesn't reflect their actual position or opinion. Questions that make it impossible for a respondent to answer honestly and sincerely.
- **Double-Barreled:** Questions that force a respondent to answer two questions at once
- **Absolute:** Questions that force respondents into insincere absolutist positions even if they only partially agree/disagree
- **Unclear:** All other questions that are otherwise unclear, too complicated, too long, inscrutable, too technical, confusing, etc.

Interactive Example included in linked Google Form Survey: **https://forms.gle/9LNw36W9PXXJ9rDH6**

SURVEY ADMINISTRATION

The most important factor in administering surveys is to ensure that the participants are being pulled from a representative sample of the population of interest. Every member of the population of interest (in this case, Americans of voting age) has to have an equal and known chance for being selected for the sample (Morling, 2017). The best way to accomplish this is with probability sampling, and an ideal probability sampling plan for straw polls in particular is simple random sampling, the most basic form of probability sampling (Morling, 2017). In this case, pollsters can program computers to randomly select telephone numbers from a database of eligible numbers all over the country, and administer to them a well-designed survey. By making use of a simple random sample sampling plan, surveyors can preserve critical external validity and we can be confident we're working with a representative sample, meaning the results of the survey should generalize to the entire population of interest.

Problem and Assessment

THE FINAL PRE ELECTION CBS-TIMES POLL

A collaboration between CBS News and The New York Times, done on their behalf by polling agency SRS of Media, PA. Conducted over the phone on November 2nd, in both English and Spanish using live interviewers. Among a random sample of 1,753 adults nationwide, 1426 of whom were registered voters (Sarah Dutton, 2016). Phone numbers were taken from both land-line and mobile phones, employing a random digit dialing method. For landlines, participants were randomly selected from among all the adults in a household. For mobile, whoever answered the call became the interviewee. The data was weighted to reflect Census figures concerning demographic variables. Sampling error was estimated to possibly be plus or minus three percentage points (Sarah Dutton, 2016). The margin of error was also plus or minus three (Sarah Dutton, 2016). Margin of error slightly enlarged the sampling error by accounting for the weighted sample.

Interactive Example included in linked Google Form Survey: **https://forms.gle/mzu8NcsJacPRjiPNA**

AUTOPSY:

To evaluate any potential issues with survey design, a panel of experts from the put together an autopsy from the American Association of Public Opinion Research (AAPOR), which they later published in a 2018 edition of *Public Opinion Quarterly*. These are the highlights from their evaluation:

> "While the general public reaction was that the polls failed, we found the reality to be more complex. Some polls, indeed, had large problematic errors, but many polls did not. In particular, the national polls were generally correct (with respect to the popular vote) and accurate by historical standards." (Keeter, 2018).

> "…there is only mixed evidence that misspecified likely voter models were a major cause of the systematic polling error." (Keeter, 2018).

> "…the 2016 election featured a number of unusual circumstances that are perhaps unlikely to repeat (e.g., both major party candidates being historically unpopular, split

outcomes in the popular vote and Electoral College, nearly 14 million votes across three states breaking for a candidate by about one-half of one percentage point..." (Keeter, 2018).

"Finally, a late swing in favor of one candidate (as appears to have occurred in 2016) is not something that pollsters can necessarily guard against…" (Keeter, 2018).

What seems to be the case is that contrary to public perception, the polls were not systematically flawed in any significant way. What flaws they do seem to have are either negligible or difficult, even impossible to control for. Furthermore, 2016 was an anomalous election year fraught with confounding factors. Predictions are made based on patterns and consistencies, and therefore, even the best predictions will fail to account for breaks in the patterns and inconsistencies.

Alternative Solutions and Their Problems

MACHINE LEARNING

It might seem tempting to eliminate the main source of error in surveying (the human factor) by replacing human surveyors in the equation altogether. One of the potential solutions to this issue currently being discussed in the literature is the application of AI technology and machine learning models to replace traditional pollsters and purportedly predict the outcome of elections more accurately. In an article published in the Journal of Industrial Engineering International, Zolghadr et al. (2018) proposed that by making use of artificial neural networks (ANN) and support vector regression (SVR) models (two different types of machine learning algorithms), they claim they can increase the accuracy of prediction models for forecasting the U.S presidential election by a whopping 50%. The way they claim they can do this is unique to each method. For SVR, it's based on what they call "the statistical learning theory", which essentially "characterizes properties of learning machines to make them able of generalization of unseen data". (Zolghadr, 2018). As for ANN, their recommendation for its application in election forecasting is based on four factors; being nonlinear, data driven, able to generalize, and not needing to make assumptions on the part of the input data. While this would be an exciting development for survey methodology if proved effective, the research is still mainly theoretical side, and it faces critical obstacles. First, the results of the researchers AI predictions are based purely on nationally reported statistics; meaning they neglect including and fail to account for political and economic variables of each individual state. Which, by their own admission, "are also significantly influential on people's decisions in elections". (Zolghadr, 2018). Just think of how not accounting for the state specific political realities of battleground states like Wisconsin, Ohio and Michigan could affect a model's predictive ability when it came to the 2016 presidential elections. Based on past data, these states were blue; but many of the same people who voted for Obama twice also ended up voting for Trump in the next election cycle. This is well within the realm of possibility in the perspective of human pollsters, who have access to the wealth of qualitative research about why that happened, but would seem anomalous and critically incongruent to current AI, which are simply not equipped with the appropriate parameters to reliably handle that sort of information.

Beyond this fundamental design flaw, there are a few technical difficulties that need to be worked out as well, and hitherto lay unresolved. "In applying SVR and ANN…lack of a specific method to specify some parameters…in SVR and the number of hidden layers in ANN." (Zolghadr, 2018). The technology shows significant promise but ultimately, in its current state, it's still in its infancy and not a realistic solution for the problem at hand. Beyond this, as mentioned above, one of the fundamental issues with a surveys' "failure" to report accurate outcomes comes from misinterpretation of the results, and lack of understanding amongst laypersons on how surveys are designed or analyzed. This is a process already made confusing just by utilizing simple descriptive statistical analysis; such a problem, already pervasive as it is, would only be exacerbated by upgrading that complexity to something perhaps only a handful of people with PhD's in the world have any clue how to make heads or tails of. Not only are there exceeding advanced mathematical calculations involved with these as of yet incomplete learning algorithms, the software itself is poorly understood by even the people who have designed it. Whatever promise this may show in terms of increased accuracy were the scientists behind it able to figure out all the nuts and bolts, it would be counteracted tenfold by the necessity of one needing several PhD's just to interpret the results, let alone run of the mill journalists and average Joe consumers of science.

ABANDONING SURVEYING

Some have suggested surveys have proved themselves to be so unreliable a measurement tool that we as a public should, or should at least seriously consider, abandoning them and their use completely. In a 2016 news article, Peter Barnes, senior elections analyst at the BBC wrote "The presidential election result came as a shock to people….Just like…the EU referendum, the outcome was different from what most polls had suggested. So should we give up on them altogether?" (Barnes, 2016). It's a troubling question. Although its understandable laypeople and journalists felt betrayed by the results of the surveys and their "failure" to predict accurate outcomes, it's an overreaction to just give up on attempting to gauge public opinion altogether, and completely erroneous to think this to be the best course of action available to us. We can concede that the year 2016 was an exceptionally shocking, and therefore perhaps traumatic, year for polling. When so many well established and highly visible pollsters ended up being so wrong in such conspicuous, nationally attention grabbing ways, it can really discourage the average person from placing any real faith in polling ever again. It's easy to see how this could cause mistrust, suspicion, even resentment; left unaddressed, such sentiments can lead to a rift between the public and the pollsters, and erode the latter's credibility.

But at what point does sympathetic disillusionment with polling and pollsters, under unusual circumstances, give way to abject paranoia and poor judgement? Journalistic organizations place a premium on credibility, and understandably so; their entire business model relies upon the degree of trust the public places in their institution's ability to report the truth. Additionally, journalists rely on their partnerships with professional pollsters, publishing their predictions for their seasonal election coverage. When the polls are wrong, and in such a public manner to boot, there is an unintended consequence for the people who publish and report on those polls as well. It is for presumably this reason that even pollsters themselves have become

hesitant to enter into the arena of forecasting presidential elections. In 2016, neither Gallup nor Pew Research, two of the most prominent names in mainstream polling, decided to publish "horse race" polls for the 2016 elections (the latter abstained from publishing such polls in October and November specifically). (Jackson, 2017). Presumably this was done out of fear of repercussions to their reputation in the event they called the race wrong. Another mainstream pollster shut its doors entirely, and ceased all polling work. The 70 year old California Field Poll closed its doors. (Cadalgo, 2016). There was even a brief scare where it seemed like even the NYT CBS poll would close; certainly, with the departure of key personnel, its future was uncertain. (Pompeo, 2016). French newspaper *The Parisien* has taken that drastic step of forgoing opinion polls entirely. In a radio interview, Editor Stéphane Albouy said "Instead of just talking about the errors and faults in polls, we have decided to return to the core of our profession. That is to say: the field. We want proximity. It is that way of working that we want to keep going forward." (Birnbaum, 2017). But refusing to conduct opinion polls anymore and shutting down entire polling institutions isn't a satisfactory solution by any means, and it fails to address the fundamental problem. If we simply abandon an instrument of data collection that occupies an important research niche because we perceive it to be defective or at the very least not worth the trouble, without finding an adequate or equivalent replacement, we actively harm science as a field. There is no suitable alternative to surveying when it comes to gauging public opinion at large. And must we carry on as if anomalies don't exist? 2016 was an unusual year for polling; it was an unusual year for the political process in general. Even the most well designed survey cannot account for every joker, wildcard or ace up the sleeve, and surveying as a research practice shouldn't be abandoned because of it. And regardless, a Trump victory was already within the margin of error of the majority of the mainstream polls; for anyone familiar with survey methodology, they would not have been blindsided by the ultimate outcome. Barnes himself points out the ludicrousness of the sentiment, when he writes "…if the polls suggest the correct outcome, small or moderate inaccuracies are barely noticed. But when they get the outcome wrong it's considered a disaster." (Barnes, 2016). This isn't an acceptable way to resolve the problem. Science is a cyclical process. (Morling, 2017). Just because your initial prediction was proved wrong doesn't mean you shouldn't make the necessary changes to your model and try again, except this time, with a better understanding of what went wrong. If a survey fails to forecast accurate election outcome, better to do an autopsy and learn from you mistakes then give up entirely and swear off the whole practice of polling.

My Recommendation

The failure of pollsters to accurately forecast the winner of the 2016 presidential election is a situation in which an instrument of data collection (surveys) presumably interfered with their own measurement (of public opinion). We know that surveys, as a method of data collection, are particularly prone to error. This is not because surveys are some invalid, unreliable or otherwise inferior research method; it is because surveys are a delicate instrument that require a skilled designer and skilled administrator. They are not inaccurate because of some immutable flaw of their fundamental design, but rather, because they are incompetently designed or applied (in most cases). In a perfect world, a perfect survey is perfectly accurate 100% of the time. In our

world, a well-designed survey is reasonably accurate most of the time. After examining one of the most prominent surveys used by a mainstream pollster, CBS-Times (the partnership between CBS News and the New York Times to survey electorates for critical elections), for any glaring defaults, it becomes apparent that that survey in particular is quite well designed. It's questions hold high ecological validity, it does not sacrifice construct validity for external validity, it does not use poorly worded question types. Its random sampling plan, too, seems more than sufficient to ensure that its administration should have garnered a representative sample. While it is not devoid of *all* room for improvement, for the most part, there is little if any reason to believe that such an excellent survey should fail to measure what it seeks to measure. Acting in good faith, and reasonably understanding that at the professional level many of the nation's mainstream pollsters and polling aggregates, including but not limited to the Princeton Election Consortium, Monmouth, 538, etc. operate at this same degree of survey design and administration excellence, how then, to account for the near unanimous discrepancy between the prediction and the result in 2016? The error in this event does indeed lie with the human factor; but not with the creators and administrators of the nation's election surveys. Rather, the source of the polling error in this case is actually in its interpretation, by laypersons in journalism, and the average Joe consumer of science.

The *correct* interpretation of survey results is not merely a salient factor in any worthwhile discussion of the validity of surveys as a measurement tool, it is an absolutely crucial one. Where we to make any corrections to existing survey design, we must first take into account the fact that not every consumer of research whom those results reach is someone who is adequately equipped to handle proper analysis and reporting of results. The fact of the matter is, for many of these polling intuitions, their polls accounted for a Trump victory within the margin of error. It was a close race across the board, and for those familiar with proper survey design and interpretation, the outcome was not unexpected. The degree of shock and disbelief displayed by journalistic institutions and the general public at the outcome of the election demonstrate however, that this expertise in survey interpretation is not widespread. Yes, there are some problems with surveys as instruments of data collection; often, it is because they are incompetently designed or administered. In this specific situation, however, their issues lies not with the ability to measure their construct, or generalizability to the real world, but rather, our understandings and attitudes, as a public, towards the results they produce. *We* do not understand survey results; margin of error remains a mystery to the vast majority of the public.

The polls said Clinton would win, then she did not, ergo ditch the polls. This is the simple, intuitive conclusion. But it's a judgement made in ignorance. The polls did not, in fact, say Clinton would win. They said she might win, and there was a significant chance that Donald Trump might win instead. And then he did, and all of this was accounted for by design. But not in the minds of the public. Because the public does not understand survey design. Or survey application. But most importantly of all, the public does not understand survey interpretation. Rather than seek to fix an instrument of measurement that is not broken, we should seek to remediate the lacking of proficiency in its technicians. The remedy for a situation like the 2016 is a change in our attitudes towards and our relationship with surveys, and our interpretation and understandings of survey results.

Firstly, we must all temper our expectations of what we can get from survey results. Surveys are not crystals balls that tell us with certainty what the future will be; they are merely gauges of public opinion, and their capacity to forecast outcomes, no matter how well designed, are fundamentally limited. And that's perfectly fine. We ask too much of surveys. A scale tells you your weight in pounds, not your exact body fat composition. Understanding that some of that weight may be water weight, or the principle that weight fluctuates throughout the day, or the weight is not an exact measure of body fat composition, and that measurements in scale readings have acceptable margins of error, are all necessary parts of making appropriate use of weight scales as tools of watching our health and fitness. Similarly, understanding and accepting the intrinsic limitations of survey's abilities to predict the future is an important part of being a responsible technician of that measurement tool. Surveys, properly designed and applied, are the best instrument available for gauging public opinion. Whether this means we now know the results of an election before it occurs is an entirely different question, and one we should not expect surveys to be able to accurately address; it's simply not their function.

Secondly, we must all exercise healthy skepticism when applying survey results to our decision making. Even the best survey results can sometimes suffer from extemporaneous factors. Consider the fact that 2016 may be one of the most unusual election years in American history. Anomalies happen, and any instruments ability to account for, and accurately report on anomalies has limitations. In the case of the 2016 election, they were many different anomalies in the electorate. People crossed parties, people stayed home, those who had not voted in decades voted for the first time, and there were myriad ways in which things were "out of the ordinary". Being skeptical of any and all results in the eventuality that such a phenomenon occurs is a good insurance against meeting the same fate as the unfortunate Chicago Tribune during the election 1948. Having faith in surveys is good, but don't count your chickens before they hatch remains a prudential adage. A responsible consumer of science makes sure they're using research results appropriately and judiciously, not jumping the gun and rashly following an ill-advised publication strategy to get one over on competition without regard for the possibility that the polls may have gotten it wrong (for whatever reason). And shifting the blame after the fact after making a fool of oneself is exactly the kind of behavior that only an irresponsible consumer of science would do.

Such a situation is akin to blaming the scale you have at home for reading a number lower than the one you have at the weigh in for an athletic competition; it's not the scales' fault you've failed to account for any minor discrepancies and are now out of your weight class by 1-2 pounds. You shouldn't have made the assumption in the first place and committed to your current weight before it was too late. Journalistic organizations and laypersons have the same relationship with surveys and their results. Researchers will continue to do their best to design good, valid surveys that accurately gauge public opinion give or take margin of error. As responsible, good faith consumers of science, it is our responsibility to educate ourselves in the appropriate way to interpret those results and whether or not to allow them to shape our decision making. We do not sue the weatherman for making us wear a raincoat on what later turns out to be a sunny day; because we have common sense understandings that the morning weather report is not an absolute or guarantee. Similarly, we should not hold grudges or sulk over a handful of

survey results that turn out not to match the exact outcomes of unpredictable, wildcard elections, especially when those results mostly turn out to be within the margin of error regardless. That would be asking too much of surveys, failing to be skeptical about their results, and misunderstanding and misinterpreting survey results. Taking responsibility for correctly interpreting survey results is the solution to the problem of why an instrument, in this case surveys, seems to have interfered with its own data collection in the specific situation and context of the 2016 election. By being more informed and responsible consumers of science, we do not have to experience the mass hysteria and baseless finger pointing of the aftermath of the 2016 election, all without laying on finger on what seems to be largely excellent survey design and administration.

POLLSTER'S REFLECTIONS

The biggest consequence of the 2016 election for the practice of data collection and public opinion polling has been the substantial hit to credibility and faith in polling organizations and polls as an idea. This is evidenced by statements that have come out since from sources like professional pollsters and analysts at some of the nation's most elite private polling organizations. Ashley Koning (in an interview with Scott Keeter), director of the Rutgers University Eagleton Center for Public Polling, in a 2018 article written in an AAPOR affiliate e-journal *Survey Practice,* had this to say about the blowback:

> The number of times we've had to reassure a client that the process still works and the number of skeptical questions I get when giving talks has certainly increased since 2016. I think this new era of politics has brought about a significant amount of distrust…this distrust further hampers our job to provide an accurate depiction of public sentiment. (Keeter, 2018).

This frustration and feeling of having been the scapegoat or fall guy for a crime they had not committed, was a sentiment shared by other prominent pollsters and polling experts. Mostly it's a sentiment expressed through Meta analyses of autopsies or commentaries on and responses to the many, many criticisms and take down of the elite pollsters by journalists and pundits. In a 2018 article on the private polling website 538, owner and statistician Nate Silver reflected on what lessons he and his team had learned in the aftermath and dissection of the 2016 elections:

> Polls were never as good as the media assumed they were before 2016 — and they aren't nearly as bad as the media seems to assume they are now. In reality, not that much has changed….Media organizations need to do a better job of informing their readers about the uncertainties associated with polling and of distinguishing cases in which polls were within a few percentage points of the correct result, like in the Brexit vote, from true polling failures, like the 2016 Michigan Democratic primary. But they also need to recognize that polls pick the winner correctly most of the time — about 80 percent of the time — and that media's attempts to outguess the polls have a really bad track record. (Silver, 2018).

For the most part, what the 538 team had to say about the ultimate, underlying cause of the discrepancies between prediction and results in 2016 echoed this proposal's conclusions. In

fact, there seems to be a growing consensus about the real issue being one of interpretation and expectations by the consumers of science, rather than of survey construction and administration by the producers. One of the very first sources to come to this conclusion, all the way back in November of 2016, just a few short days after the election when the zeitgeist amongst most journalists and even some pollsters themselves was that the polls had been botched in some way (this was before all the extensive post-election autopsies like the one conducted by the AAPOR), was Sean Trende with the journalistic organization/publication RealClear Politics:

> What occurred wasn't a failure of the polls. As with Brexit, it was a failure of punditry. Pundits saw Clinton with a 1.9 percent lead in Pennsylvania and assumed she would win. The correct interpretation was that, if Clinton's actual vote share were just one point lower and Trump's just one point higher, Trump would be tied or even a bit ahead….So don't blame the polls. They did what they were supposed to do, and in fact, did their job as well as they did in 2012. Instead, blame the analysts and pundits, and their stubborn resistance to considering the possibility of a Trump presidency. (Trende, 2016).

There's been a growing alarm about this disconnect between the sensations and perceptions of the pollsters and polling experts, and the laypeople who were using and critiquing these same polls according to their own agendas. There was not just misattribution and unfair, sometimes even hypocritical hypercriticism of the polling institutions, but even more concerning, blatant misunderstanding, To put in the words of HBO's Last Week Tonight's John Oliver (in a 2016 episode talking about the average consumer of sciences chronic misinterpretation of scientific studies and the pitfalls of that parallel situation):

> "…except that's not what the study said! It's like a game of telephone, the substance gets distorted at every step….Some of this is on us, the viewing audience. We like fun, poppy science that we can share like gossip, and TV news producers know it…this is dangerous, this is really dangerous! If you start thinking science is a la carte and if you don't like one study, don't worry, another will be along soon, that is what leads people to think manmade climate change isn't real, or that vaccines cause autism… Science is by its nature imperfect, but it is hugely important. And it deserves better than to be twisted out of proportion and turned into morning show gossip." (Oliver, 2016).

Conclusion

Surveys are an instrument of data collection that play a key role across many scientific disciplines. They are a delicate instrument; in the wrong hands, they are prone to error. Therefore they require both skilled designers and skilled administrators. A situation in which they seemed to have interfered with their own measurement was the 2016 presidential election. A near unanimous failure amongst even the nation's most elite pollsters to accurately predict the outcome of the election occurred. One of the proposed solutions to the problem was to incorporate machine learning into the design and administration of future surveys. While promising, the technology in its current state is inadequate; it remains mostly unfinished and what does exist is plagued with parameter problems, among other concerns. Another proposed solution is to abandon the data collection method of surveying altogether. However, no superior

or equivalent instrument for gauging public opinion currently exists which is a vital part of data collection across many scientific fields. Therefore this solution completely fails to address the problem. Furthermore, it's evident from the rhetoric surrounding the "solution" that it's merely a hysterical overreaction, not grounded in sincere and rational decision-making. Such a proposal cannot be taken seriously or in good faith. So how can we fix surveys? Autopsies of the polls after the fact by experts suggest the results of the polls were within the margin of error, and substantive issues with either their design or administration were negligible. I propose that in this situation, the problem was the incorrect interpretation of the survey results, not the initial data collection. Meaning the fault, if any exists, lies with consumers of science, not with the producers. Therefore, the solution is for laypeople, such as the journalists who published the polls, and the audience who closely followed those polls, to change their general attitude towards, and relationship with surveys. By improving how well they understand sound survey design, valid survey administration and correct interpretation, average consumers of science increase the usefulness of surveys. Surveys whose results laypeople do not understand, published by journalists who also do not understand what they are publishing, help no one; least of all scientists interested in collecting useful data and making accurate predictions, who need support and understanding from the public in these kinds of situations when it appears like a valuable measurement instrument may have interfered with its own data collection. One of the surveying concepts the public at large should become more familiar with is margin of error, which was crucial to this situation. An attitude the public can adopt is healthy skepticism, and lowering their expectations of the predictive powers of surveys to forecast complex events, and their need for them to do so consistently to shape their personal decision making. The alternative solutions are inherently flawed; and beyond that, fail to address the real problem discovered by close examination of an extensive autopsy on what potentially went wrong. Therefore, the proposal that the solution to this issue is more responsible consumers of science who are capable of making correct interpretations of survey data is the most likely to be successful. To put it succinctly, after close examination of the available evidence, we're assigning blame to the wrong things. It is not the case that the polls failed us; rather, I propose it is we who have failed the polls, and the remedy is to improve ourselves before looking to improve polls that do not have (nor do they warrant) much room for improvement. This is evidenced by elite, high-profile national pollsters changing almost nothing about their polling methods, citing this same reasoning, having come to the similar conclusions. And it is truly important that we do this, because one of the most devastating effects that this misinterpretation of survey results has had in the wake of the 2016 election has been widespread faithlessness in the ability of polls to accurately predict anything.

References

Barnes, P. (2016, November 11). Reality Check: Should we give up on election polling? BBC News, pp. 1-5.

Birnbaum, M. (2017, January 3). Opinion polls missed Trump and Brexit. This French newspaper says it has the solution. The Washington Post.

Cadalgo, C. (2016, December 9). The Field Poll is shutting down after 70 years of charting political highs and lows. The Sacramento Bee.

Clement, S. (2017, February 6). The 2016 national polls are looking less wrong after final election tallies. The Washington Post, pp. 1-4.

Cosgrove, B. (2014, May 4). Behind the Picture: 'Dewey Defeats Truman' and the Politics of Memory. TIME, p. 1.

Courtney Kennedy, C., Blumenthal, M., Clement, S., Clinton, J.D., Durand, C., Franklin, C., McGeeney, K., Miringoff, L., Olson, K., Rivers, D., Saad, L., Witt, G.E. & Wlezien, C., An Evaluation of the 2016 Election Polls in the United States, Public Opinion Quarterly, Volume 82, Issue 1, Spring 2018, Pages 1–33, https://doi.org/10.1093/poq/nfx047

Jackson, N. (2017, January 6). Don't Swear Off Polls, Help Fix the Problems. Huff Post, pp. 1-4.

Jones, T. (2007, December 19). Dewey defeats Truman. Chicago Tribune, pp. 1-2.

Pompeo, J. (2016, August 25). Future of Times-CBS News poll is up in the air. Politico, p. 1.

Keeter, S. (2018). Ask the Expert: Polling the 2018 Elections. Survey Practice, Vol.11 Issue 2.

Morling, B. (2017). Research Methods in Psychology: Evaluating a World of Information. New York: W.W Norton & Company.

Oliver, J. [LastWeekTonight]. (2016, May 6). Scientific Studies: Last Week Tonight with John Oliver (HBO) [Video file]. Retrieved from https://www.youtube.com/watch?v=0Rnq1NpHdmw

Pompeo, J. (2016, August 25). Future of Times-CBS News poll is up in the air. Politico, p. 1.

Presidential Election Results: Donald J. Trump Wins. (2017, August 9). Retrieved from The New York Times: https://www.nytimes.com/elections/2016/results/president

Sarah Dutton, J. D. (2016, November 7). CBS News poll: State of the race the day before Election Day. Retrieved from CBS News: https://www.cbsnews.com/news/cbs-news-poll-state-of-the-race-the-day-before-election-day/

Siegel, E. (2016, November 9). The Science of Error: How Polling Botched the 2016 Election. Forbes, pp. 1-9.

Silver, N. (2018, May 30). The Polls Are All Right. Retrieved from FiveThirtyEight: https://fivethirtyeight.com/features/the-polls-are-all-right/

Squire, P. (1988). WHY THE 1936 LITERARY DIGEST POLL FAILED. Public Opinion Quarterly Volume 52, 125-133.

Trende, S. (2016, November 12). It Wasn't the Polls That Missed, It Was the Pundits. Retrieved from RealClear Politics: https://www.realclearpolitics.com/articles/2016/11/12/it_wasnt_the_polls_that_missed_it _was_the_pundits_132333.html

Zhou, L. (2018, October 24). How pollsters are responding to problems that came up in 2016. Vox, pp. 1-7.

Zolghadr, M., Niaki, S.A.A. & Niaki, S.T.A. J Ind Eng Int (2018) 14: 491. https://doi.org/10.1007/s40092-017-0238-2

Image Citations

Blazek, L. (Photographer). (2018, November 02). Person Writing on Notebook [digital image]. Retrieved from https://www.pexels.com/photo/person-writing-on-notebook-669615/

[Digital image]. (2018, November 06). Retrieved April 1, 2019, from https://www.businesswire.com/news/home/20181106005838/en/ORISE-Accepting-Applications-2019-Summer-Research-Team