POLLING IN PRACTICE: ACCURACY OF NATIONAL AND STATE POLLS IN THE 2016 ELECTION

Amanda McLean
M.A. Thesis with the Department of Communication, Stanford University
Under the direction of Prof. Jon Krosnick

Abstract

This study focuses on the impact of mode and sampling procedures on polling accuracy during the period of the 2016 presidential election. Utilizing all national and state level polls released during the final 7 days of the election, polls are classified based on the mode and sampling procedure utilized. Accuracy is then analyzed using the average absolute error of the top two candidates (Donald Trump and Hillary Clinton). The study finds that polling methodologies are utilized at very different rates, with polling in the final weeks of the 2016 election being dominated by Internet / River sampling methodologies (45% of polls). Within the sample of polls, RDD / Live Interviewer polls were the most accurate at both the state and national level. The least accurate polls were those that utilized online / river sampling methods. Significance testing was also conducted to measure the significance of the differences in average absolute errors between RDD / Live interview polls and the least accurate, Internet / River Sampling. Both a stratified Fisher-Pitman permutation test and a test using the untransformed absolute difference from the election outcome find strong support for the hypothesized difference between RDD and river surveys.

Keywords: survey methodologies, pre-election polling, mode, sampling

Acknowledgments

I would first like to thank my thesis advisor, Dr. Jon Krosnick, for his continued support, guidance, and contributions to this paper and the wider project. I would also like to thank Dr. Alex Tahk at the University of Wisconsin for his contributions and direction related to the statistical analysis and modeling for this project.

I also want to express my gratitude for the community of scholars within the Political Psychology Research Group, in particular Dr. Sebastian Lundmark, for their continued support and ideas as the project progressed through each of its stages. In addition, I would like to thank Stanford VPUE for providing funding for conference travel to present and discuss the topics found within this paper with a wider group of scholars.

Finally, I would like to thank my parents, family, and friends for their continued support and encouragement throughout my entire academic career and in the production of this thesis. Without them, none of this journey would have been possible.

Introduction

When people went to bed on November 7th, 2016, most in America were confident that Hilary Clinton would be the 45th President of the United States. However, by the end of November 8th, many were shocked both by the outcome and the perceived inaccuracy of the polls conducted. Headlines certainly bore this out during the following weeks and months, with some arguing that this was a sign that polling had lost its scientific credibility. The influence and awareness of polling within Presidential elections has continued to grow since presidential straw polls were first conducted in the early nineteenth century. The 2016 election cycle proved to be no different with high levels of spending and public awareness towards the findings of various pollsters.

New forms of quantitative journalism, such as FiveThirtyEight and the New York Times' Upshot published electoral models that predicted the outcomes of races using polling data. Polls were extensively incorporated into news broadcasts and articles detailing the predicted outcomes of the election at any given time. This narrative quickly shifted when the official results were announced and Donald J. Trump was elected the 45th President of the United States, contrary to what many polls seemed to suggest. On the day before the election, the New York Times Upshot released a model showing that the polling data suggested Hillary Clinton had an 85% chance of winning the Presidential election. When media consumers instead witnessed Trump election, a discussion began to be shaped surrounding polling's inability to accurately predict results. This was further supported by past polling failures, such as during the Brexit elections.

However, this narrative bucks the observed trend that scientific polling has become more accurate over time. Since 1958, the accuracy of pre-election polls has averaged around 2% points off the actual election result. The post-2016 media storm would seem to suggest that this number should be much higher for the 2016 election. The purpose of this study is to examine the accuracy of the 2016 election polls, analyzing if mode and sampling are predictors of a poll's accuracy. The expansion of polling in the twentieth century has led to a diffusion of polling methodologies, without a significant amount of popular analysis regarding the accuracy of these methods. While many companies claim to offer high rates of accuracy, there are also competing motivations of reducing costs and fielding time required to produce surveys. "Poll" has become a unintelligible word for many media consumers, but that does not mean it is too late to strive to differentiate between polls using the most scientifically rigorous methodologies and those that produce higher levels of error and variance. Understanding public opinion and preferences in the American electoral system is too important.

Background

Election polling is arguably one of the most ubiquitous examples of statistics found in everyday life, and has found increasing prominence in both popular media and public discourse. While polling can be traced back to censuses administered several millennia ago and to the straw polls conducted early in America's political history, surveys did not become more widely adopted until the twentieth century (Weisberg, 2005). The number of public polls has continued to grow since the 1990s, with polling being fully entrenched in the news stories about campaigns by the early 2000s (Frankovic, 2005). Accompanying polling's greater prominence has been an expansion of the variety of survey methodologies used by pollsters in pre-election polling, as well as proponents for different types of survey sampling and administration methods. This has placed increased importance on gaining a better understanding of sources of error in polling design and administration. While there are many potential types of error that have been covered in previous studies, this paper focuses on mode and sampling methodologies as a focus of analysis.

Over time, survey research has seen the evolution and transition of modes of data collection used in pre-election polling. While face-to-face interviewing was originally the predominant mode of data collection (Chang & Krosnick, 2009), significant increases in cost since the 1970s (De Leeuw & Collins, 2012) prompted the development and implementation of alternative data collection modes such as telephone interviewing and interactive voice recording (IVR) administration (Dillman, 2000). Internet surveys emerged in the late 1990s, with new and established polling firms adopting online data collection for practical and financial motivations. With the adoption of new modes of data collection, there continues to be debate in the survey community about new methodologies' accuracy and impact upon data quality. Within the literature, there is a long history of mode comparison studies, such as comparing face-to-face to telephone surveys (Holbrook, Green, & Krosnick, 2003) and telephone to Internet surveys (Fricker, Galesic, Tourangeau, & Yan, 2005). This study will continue this tradition, but will classify polls additionally based on the type of sampling method utilized.

While the theory of probability was established in the 18th century (Groves et. al., 2011), its principles were not fully applied until the 20th century as the idea of probability sampling in survey research was introduced. Since then, the majority of survey researchers have utilized a probability-sampling framework (Baker et al., 2013). RDDs emergence in the 1970s helped to alleviate growing concerns about the cost, difficulty, and coverage problems of earlier probability samples as RDD surveys could be done at relatively low cost and non-response (at least in this early time-period) (Glasser & Metzger, 1972). However, increasing costs associated with probability sampling methods and increasing concerns about telephone survey coverage due to the growth of cellphones have led to increased growth and interest in online sampling methods (Cooper, 2000; Baker et al., 2013). The sampling approaches within nonprobability sampling vary significantly, such as with opt-in panels that recruit participants using a variety of methods. Within the existing literature, several studies have found instances in which pre-election polling non-probability samples have yielded results similar to or better than probability-based survey samples (Terhanian et. al. 2001; Vavreck & Rivers 2008). However, leading survey research associations such as AAPOR still suggest that researchers should avoid nonprobability online panels when "one of the research objectives is to accurately estimate population values" among other situations (Baker et. al., 2010). With this issue far from settled, this study seeks to utilize pre-election polls from the 2016 Presidential election to better understand the effect of sampling and mode on poll accuracy.

Methodology

The data set utilized in this study was constructed by incorporating all the polls available on the FiveThirtyEight, Real Clear Politics, and the New York Times Upshot websites that were fielded on October 30th or later (the final week before the election). This was done to include only polls that should be the most accurate based on those polls capturing respondent's opinion nearest to the election. Each poll was then classified based on the mode and sample used. Mode refers to the mode of respondent data collection. Within the data set, polls utilized a variety of modes including:

Live telephone interviewers: a live person interviewing the respondent over telephone to collect their responses to the questions answered

Interactive Voice Response (IVR): the respondent interacting with a recorded voice and entering their responses either by voice or keypad

Internet: surveys in which the respondent answers the survey questions using the internet on the computer, a tablet, or cellphone

IVR + Internet: a mix of both the above IVR and Internet methods

IVR + Live telephone interviewers: a mix of both the above IVR and live telephone methods

In survey research, live interviewer is considered the gold standard due to its higher response rates and lower respondent error, however our study seeks to measure the accuracy of all modes during the 2016 Presidential election to better understand which proved to be the most accurate.

The second method polls are categorized by is sampling method. Polls in the 2016 Presidential election used a variety of sampling methods including:

Random Digit Dial (RDD) using landline and cellphones: the process of generating both landline and cellphone numbers through a random process of number selection

Address based sampling: a methodology that relies on a sample generated by residential addresses through the US Postal Service

Lists of people registered to vote: a sample generated by using the registered voters of a certain area and producing a list of telephone numbers from this voter data.

Automobile owners: a sample consisting of automobile owners that have purchased a particular-brand of car to create a sample during the 2016 election

River sampling: on online sampling method that involves directing visitors of specific internet pages to participate in a survey

Opt-in panels: an online sampling method in which panel members that have signed up with a specific panel company to take surveys, often for some type of reward

Unknown online samples: an online sample for which we were unable to identify the type of sample used

To measure the accuracy of each poll, poll results were broken down based on the predicted vote share of each candidate. Poll results were adjusted to eliminate the “don’t knows”, due to the expectation that don’t knows will reallocate themselves proportional to the non- don’t knows (Shuman & Presser, 1981). Actual election results were gathered from individual state’s Secretary of State, who is tasked with maintaining data regarding the results of elections. The average absolute error for two candidates, the statistic most of interest for this project, is calculated by taking a candidate’s polled result, subtracted from the actual result (taking the absolute value of this number). The same is then done for the second major party candidate, with those two values averaged together to calculate the average absolute error for two candidates. While polling and election data was collected for all candidates listed on the ballots, the focus is placed on the two major party candidates because of their presence on ballots in all states and in all polls.

Additional significance testing was also conducted using both a nonparametric and a parametric mixed-model approach. The nonparametric approach seeks to avoid making distributional assumptions about the errors while still accounting for state-level clustering and uses a stratified Fisher-Pitman permutation test to test the difference in the natural logarithm of the squared difference from the election outcome. We stratify by state to ensure that our results are not driven by the differences in the distribution of survey types between states. The second approach is a parametric mixed-model approach. This allows us to use a more complex model, taking into account variation in treatment effects between states as well as variation between survey organizations. Only RDD and river surveys within one week of the election are included in the analysis. In addition to national surveys, the data include state surveys from all 50 states and the District of Columbia. Note that the District of Columbia is treated as a “state” for the purposes of this analysis despite not actually having statehood. We include only state-level surveys in our primary analysis. This eliminates the risk that some organizations might include the same respondents in state-level and national surveys and increases comparability between polls. As a robustness check, we also ran our analysis using both state and national polls. Due to the relatively small number of national polls during the last week of the campaign, there is insufficient data to analyze national polls alone.

Findings

An important initial discovery is that different types of polling methodologies are utilized at very different rates. Polling in the final weeks of the 2016 election was dominated by a few online survey companies, Survey Monkey and Google Consumer Surveys, which fielded 47% of the polls in the lead up to election day. The methods with the highest expected accuracy, RDD and live interviewer surveys, consisted of a relatively small percentage of polls conducted (and especially at the state level). This has important implications as we consider the average accuracy of each method when compared to its rate of utilization during the 2016 election. A detailed breakdown of the frequency of each polling type can be seen in Table 1.[1]

[1] For the purpose of this analysis, it is assumed that no two polls in this analysis involve the two same responses. However, it is possible that some surveys reported national statistics based on aggregated state numbers.

Table 1: Poll Type Frequency (State and National)

	Live Interviews	IVR	Internet	IVR + Internet	IVR + Live Phone	Other Mixed Methods
RDD	17	2				
Address-based Sampling			1			
People Registered to Vote	16	19		5	19	1
Automobile Owners						24
River Sampling			156			
Opt-In Panel			11			
Unknown Online Sampling			5			
Mixed Sampling Techniques	4			41		

National polls make up a relatively small percentage of polls fielded during the election cycle, but have an outsized public perception due to their prevalence in national news programs. Predicting the popular vote, they have become a popular tool of publicizing the candidate's national popularity even if its Electoral College implications are limited. However, due to the presence of funding by national media outlets, RDD and live interview surveys make up a higher percentage of National polls. 23 national polls were conducted in the final week of the election, made up of 10 RDD / live interview surveys and 13 Internet surveys. The accuracy of these national polls can be seen below in Figure 2. RDD / live interview surveys had an average error of 0.96%, with a variance of 0.07. Internet polls using all sampling methods had an average error of 1.60% with a variance of 0.73.

Table 2: Average Error, National Polls Only

National Polls	Average Error	Variance	N
RDD / Live Interviewers	0.96%	0.07	10
Internet (all sampling)	1.60%	0.73	13

State polls, which make up a majority of the sample of polls, display increased diversity of polling methods. They also show a decrease in accuracy when compared to National level polls conducted over the same time-period. An overview of the average absolute error for all types can be found in Figure 2. The average absolute error for state level RDD / live interview surveys was 4.29%. However, this includes polls conducted by the University of New Hampshire, for which the UNH poll director, Andrew Smith, reported adjusted for age, gender and region but not education. The AAPOR Report on polling during the 2016 election found that had the University of New Hampshire adjusted for education in 2016, the error compared to other polls was removed. This contrasts with other state level polling firms included in our sample, who all weighted for either education or income (a close approximate of education). After receiving UNH polling data weighted for education, we have included the relevant UNH RDD/live-interview state-level poll in our dataset. The average absolute error was 1.9%, slightly more than 1% more than the national RDD / live interview surveys. Weighting proved to be a problem for other firms during the 2016 election as well. The USC Dornsife / LA Times poll utilizes Address-based sampling to generate a random sample, which had proved very accurate in previous election cycles. However, their national poll during the 2016 election had an average absolute error of 3.02%. It was later discovered that the poll did not consider the unequal probability of selection of their respondents that they intentionally incorporated in their respondent recruitment process. Once they built weights to correct for that unequal probability of selection, their pre-election survey data came into line closely with the final election outcome. These examples underscore the importance of transparency in regards to the weights used by polling firms for each of their polls.

The most common method used at the state level was Online / River Sampling. However, it also proved to be the least accurate, with an average absolute error of 4.38%. This type of methodology also showed a high level of variance between results, with a few being strikingly accurate and some with errors as high as 17.35%. Other methodologies showed mixed results, with live interview / people registered to vote producing an average absolute error of 2.62% and IVR / people registered to vote polls yielding an average absolute error of 2.96%. IVR + Live phone / people registered to vote proved to be the most accurate with 2.00% average absolute error over 19 polls. This is slightly surprising and could be a future area of research with a larger sample of polls conducted using this methodology. Overall, our study found that RDD / live interview scientific surveys were strikingly accurate, with an average absolute error of .96% at the National Level and 2.13% at the state level. However, they make up only roughly 6% of all polls conducted during the final week of the election. The most common, online / river sampling, was the least accurate. Most other methods fall somewhere in between, with state level results proving less accurate than national ones.

Table 3: Average Absolute Errors (State Only)

	Live Interviews	IVR	Internet	IVR + Internet	IVR + Live Phone	Other Mixed Methods
RDD	1.9%**					
Address-based Sampling						
People Registered to Vote	2.6%	3.0%			2.0%	
Automobile Owners						2.4%
River Sampling			4.4%			
Opt-In Panel						
Unknown Online Sampling						
Mixed Sampling Techniques				2.6%		

**Inc. UNH poll with weighting by ed.
***Dropped all categories with 5 or less polls

Significance testing was also conducted to measure the significance of the differences in average absolute errors between RDD / Live interview polls and the least accurate, Internet / River Sampling. Our first permutation test, a stratified Fisher-Pitman permutation test look primarily at the difference the log of the squared difference from the election outcome to reduce the possibility that a small number of very large errors would dominate our results. We also rerun our test using the untransformed absolute difference from the election outcome. The p-value for a test using logged squared differences from the election outcome is 0.0038 with an observed difference in means of 2.73. Repeating this test using the absolute difference from the election outcome yields $p = 0.0017$. In both cases, these provide strong support for the hypothesized difference between RDD and river surveys.

Table 4: Descriptive Statistics

	N	Mean	Std. Dev	Min	Max
Absolute Error	175	3.99	2.94	0.13	17.69
Log Squared Error	175	2.28	1.49	-4.11	5.75
RDD	175	0.11	0.31	0.00	1.00
River	175	0.89	0.31	0.00	1.00
National	175	0.09	0.28	0.00	1.00

Our second mixed-model approach uses linear mixed models to account for multiple sources of variation in our error. In the appendix, we show a variety of linear mixed models and generalized linear mixed models that provide further evidence that mean-squared error is higher for river surveys than for RDD surveys even when we account for variation across states. But this does not distinguish between two possible sources of greater variation for river surveys: bias and variance. Aggregating across multiple surveys will reduce variance, but do nothing to address bias.

A linear mixed model will allow us to explore this difference. Our mixed models assume that the estimate of Clinton's share of the two-party vote in a survey is equal to the sum of Clinton's actual share on election day, an error term, and, possibly, a bias term. The error terms are assumed to have mean zero but different variances for different error types. Survey bias, when present, is assumed to vary across state and survey type. We estimate four models: one in which surveys are unbiased (Model 1), one in which only river surveys are biased (Model 2), one in which only RDD surveys are biased (Model 3), and one in which both are biased (Model 4). The estimates of these models appear in Table 5.

Table 5: Linear mixed-models predicting difference between Clinton's two-party vote shares in surveys and election outcomes, excluding national surveys

	Model 1	**Model 2**	**Model 3**	**Model 4**
Mean of RDD survey bias	—	0.85	—	0.89
		(1.18)		(1.15)
Mean of River survey bias	—	—	4.57***	4.57***
			(0.93)	(0.93)
Variance of RDD survey bias	—	10.75	—	12.55
Variance of River survey bias	—	—	24.07	24.07
Correlation between RDD and River bias	—	—	—	11.13
N	160	160	160	160
Number of States	51	51	51	51
Log Likelihood	–589.36	–566.37	–587.10	–563.73
Test against Model 1 (*p*-value)	—	0.103	< 0.00001	< 0.00001
Test against Model 4 (*p*-value)	< 0.00001	< 0.00001	0.152	—

***p < 0.001, **p < 0.01, *p < 0.05

Looking at possible bias in RDD surveys, we cannot reject the null hypothesis of a model that does not include bias terms for RDD surveys. This is true regardless of whether we compare models that include river-survey bias (Model 4 against null hypothesis of Model 3, $p = 0.152$) or those that do not (Model 4 against null hypothesis of Model 2, $p = 0.103$). Thus, we find no evidence that RDD surveys provide a biased picture of vote intention, even allowing for the possibility that that bias varies between states.

On the other hand, we find strong evidence that river surveys are biased. We can reject the null hypothesis of a model that assumes river surveys are unbiased in each state regardless of whether the model assumes RDD surveys are biased (Model 4 against null hypothesis of Model

2 or Model 3 against null hypothesis of Model 1, $p < 0.00001$ in both cases). Further, while the models allow the bias to vary from state to state, the mean bias across states is also significantly different from zero ($p < 0.00001$) with an estimated bias in favor of Clinton of 4.6 percentage points. We can also reject the null hypothesis in Model 4 of both survey types providing an equal bias across states ($p = 0.028$). Thus, we find strong evidence that river surveys exhibited not only higher variance but systematic bias that tended to overstate Clinton's share of the vote in most states.

General Discussion

The utilization and growth of diverse polling methodologies is not in itself a negative occurrence, however an examination of the polls conducted during the 2016 election has shown that the least accurate methodologies are often being deployed at the highest rate. The most accurate polling methodology in our study, RDD / Live interviewer was utilized by only 21 state and national polls. This contrasts with Internet / river or opt-in polling methods that were applied in 182 polls and were found to have the highest average absolute error. This is particularly concerning when the methodologies used in "battle-ground states" that saw small margins of victory. In Michigan, Wisconsin, Pennsylvania, Minnesota, and Florida 60 polls were conducted in the final week of the election. Only one poll was fielded using optimal RDD / live interviewer methods. That poll had an average error of 1.13%. This is strong evidence that the most scientifically rigorous methods produce accurate results, but the ocean of survey research is currently polluted by many unscientific and inaccurate polls. Polling aggregators such as FiveThirtyEight and the NYT Upshot, rely on or are strongly improved by quality state level polling which drives their election outcome modeling. The low-accuracy methods utilized in states where Trump had a small margin of victory led to inaccuracy both in these models and in the general perception of Hillary as the overwhelming favorite to win the election. In addition, the electoral college necessitates an accurate understanding of state-level voter preferences, but optimal survey methods that could reveal accurate results are not being fielded.

The polls conducted during the 2016 election show that mode and sampling are powerful determinants of a poll's accuracy. Following the election, a narrative was shaped around polling's failure to accurately predict the outcome that damaged polling's public reputation. However, this narrative fails to distinguish between the types of methodologies certain polls use, whether scientific or non-scientific. While there should be continued debate over whose responsibility it is to argue this distinction, both pollsters and the news media need to improve the way in which they communicate poll's methodologies and results to the public. There also needs to be greater transparency on the part of companies conducting polls to ensure that their methods and weighting is available for public study. The consequences of not acknowledging these realities extends further than simply a decrease in public trust in polling. The fundamental benefits of polling are also diminished, creating a voting population less educated about politics, decreasing the effectiveness of political campaigns, and generating a class of politicians who understand their constituents less.

Conclusion and Future Work

Even with the public criticism towards pre-election polling after the 2016 election it is unlikely that polls importance will diminish in the national political conversation, with polling aggregator sites continuing to receive high levels of web traffic and attention. It therefore is crucial that more research and funding is devoted to ensuring that high-quality polling is produced at the state level and that poll results are reported accurately. It is also important to note that there exists a tension between survey researchers wanting to minimize error in their work but being limited by practical restraints, one of the most common being cost (Groves, 1989). The most accurate method overall within our study (RDD/Live Interview) is also one of the most expensive to administer. Technology and the internet have allowed for a proliferation of technologies that claim to simplify and lower the costs of survey administration. However, more research is required to analyze the accuracy these techniques. Finally, this project highlights the importance of transparency in poll methodology design by survey firms, without which it would be impossible to conduct research such as this. While initiatives such as AAPOR's Transparency Initiative have taken strong steps in this area, more must be done to ensure that poll results and methods are communicated accurately to researchers and an interested public.

References

Baker, R., Brick, J. M., Bates, N., Battaglia, M., Couper, M., Dever, J., . . . Tourangeau, R. (2013, June 01). Non-Probability Sampling. Retrieved March 10, 2018, from https://www.aapor.org/Education-Resources/Reports/Non-Probability-Sampling.aspx

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., . . . Zahs, D. (2010). Research Synthesis: AAPOR Report on Online Panels. Public Opinion Quarterly, 74(4), 711-781. doi:10.1093/poq/nfq048

Brehm, J. O. (2009). Phantom respondents: Opinion surveys and political representation. Ann Arbor, MI: University of Michigan Press.

Chang, L., & Krosnick, J. A. (2009). National Surveys Via Rdd Telephone Interviewing Versus the Internet. Public Opinion Quarterly, 73(4), 641-678. doi:10.1093/poq/nfp075

Crespi, I. (1988). Pre-election polling: Sources of accuracy and error. New York: Russell Sage Foundation.

Leeuw, E. D., & Collins, M. (2012). Data Collection Methods and Survey Quality: An Overview. Survey Measurement and Process Quality, 197-220. doi:10.1002/9781118490013.ch8

Dillman, D. A. (2000). Mail and Internet Surveys: The tailored design method. New York: John Wiley & Sons.

Erikson, R. S., & Wlezien, C. (1999). Presidential Polls as a Time Series: The Case of 1996. Public Opinion Quarterly, 63(2), 163. doi:10.1086/297709

Frankovic, K. A. (2005). Reporting "The Polls" in 2004. Public Opinion Quarterly, 69(5), 682-697. doi:10.1093/poq/nfi066

Fricker, S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An Experimental Comparison of Web and Telephone Surveys. Public Opinion Quarterly, 69(3), 370-392. doi:10.1093/poq/nfi027

Glasser, G. J., & Metzger, G. D. (1972). Random-Digit Dialing as a Method of Telephone Sampling. Journal of Marketing Research, 9(1), 59. doi:10.2307/3149607

Groves, R. M. (1989). Survey Errors and Survey Costs. Wiley Series in Probability and Statistics. doi:10.1002/0471725277

Groves, R. M., Fowler, F., Couper, M., Lepkowski, J., Singer, E., & Tourangeau, R. (2011). Survey Methodology. Hoboken, NJ: John Wiley & Sons.

Heimlich, R. (2004, November 23). Pre-Election polls largely accurate. Retrieved February 12, 2018, from http://www.people-press.org/2004/11/23/pre-election-polls-largely-accurate/

Hillygus, D. S. (2011). The Evolution of Election Polling in the United States. Public Opinion Quarterly, 75(5), 962-981. doi:10.1093/poq/nfr054

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires. Public Opinion Quarterly, 67(1), 79-125. doi:10.1086/346010

Kennedy, C., Blumenthal, M., Clement, S., Clinton, J., Durand, C., Franklin, C., . . . Wlezien, C. (2017, May 4). An Evaluation of 2016 election eolls in the U.S. Retrieved January 21, 2018, from https://www.aapor.org/Education-Resources/Reports/An-Evaluation-of-2016-Election-Polls-in-the-U-S.aspx

Rhee, J. (1996). How polls drive campaign coverage: The Gallup/CNN/USA todaytracking poll andUSA today'scoverage of the 1992 presidential campaign. Political Communication, 13(2), 213-229. doi:10.1080/10584609.1996.9963108

Schuman, H., & Presser, S. (1981). Questions and answers in attitude surveys: Experiments on question form, wording, and context. Thousand Oaks, Calif.: Sage Publ.

Terhanian, G., Siegal, J. W., Overmeyer, C., Bremer, J., & Taylor, H. (2001). The Record of Internet-Based Opinion Polls in Predicting the Results of 72 Races in the November 2000 US Elections. International Journal of Market Research, 43(2), 1-10. doi:10.1177/147078530104300203

Vavreck, L., & Rivers, D. (2008). The 2006 Cooperative Congressional Election Study. Journal of Elections, Public Opinion and Parties, 18(4), 355-366. doi:10.1080/17457280802305177

Weisberg, H. F. (2009). The total survey error approach a guide to the new science of survey research. Chicago, IL: University of Chicago Press.