Spring 2024

The Effect of Age, Race, Education, Margin of Error, Undecided Voters, Poll Type, and Election Proximity on Poll Margin Accuracy

Benjamin Schneider
Shawnee State University

SHAWNEE STATE UNIVERSITY

The Effect of Age, Race, Education, Margin of Error, Undecided Voters, Poll Type, and Election Proximity on Poll Margin Accuracy

A Thesis

By

Benjamin Schneider

Department of Mathematical Sciences

Submitted in partial fulfillment of the requirements

for the degree of

Master of Science, Mathematics

July 19, 2024

Accepted by the Graduate Department

7/22/2024

Graduate Director, Date

The thesis entitled '**The Effect of Age, Race, Education, Margin of Error, Undecided Voters, Poll Type, and Election Proximity on Poll Margin Accuracy**' presented by **Benjamin Schneider**, a candidate for the degree of **Master of Science in Mathematics,** has been approved and is worthy of acceptance.

7/22/2024
Date

Graduate Director

7/19/24
Date

Student

ABSTRACT

The error between presidential polling and the results in recent elections have been abnormally large compared with relatively better polling in the decades prior. The science behind how to create a political poll remains difficult given the impossibility of knowing the sampling frame of the election before its occurrence. The public wants to know who leads and where. Media wants to provide the details and makes its money doing so. Political organizations spend hundreds of millions of dollars and countless hours of human capital because polls pointed them in a certain direction. The public and its relationship to democracy goes hand-in-hand with the idea of a fair election, and polls play a large role in it. This study investigates why polls have been missing recently by looking at key characteristics of the composition of nearly 200 polls. The researcher examines polls from presidential elections in 2012, 2016, and 2020 and compared the absolute difference of the poll and the election results to the corresponding year. Then that difference was predicted using a multiple linear regression method with seven independent variables: age, race, education, proximity to election date, poll margin of error, undecided vote share, and poll mode. Results revealed that days away from the election had statistically significant results, and that mixed-methods samples were nearly significant when compared to phone-only polls. These results imply that pollsters should continue to publish more polls as the election draws nearer. Additionally, it would be worth looking in-depth about poll modes, as mixed mode could have some relationship to the current threat of nonresponse that has recently plagued polling firms.

ACKNOWLEDGMENTS

Thank you to Dr. Darbro for always being there to talk about this thesis when things got tough. Thank you to Stefan Becket and Neal Richey for tracking down and sending me several polls whose data were useful in this study. Tom Warren also deserves a big thank you for tracking down exit data that was pretty much unavailable to me. Finally, Matt Cantelon, my editor and partner in life, kept me in check the whole way, thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

During the past 100 years, news coverage of United States presidential elections has changed dramatically. In the early 1900s, news outlets focused on the issues around candidates and the promises they made on the campaign trail. Political polls were extremely expensive to conduct and were few and far between. As time went on, polls became cheaper, albeit still expensive, and occurred more frequently. By the end of the 20th century, the 24-hour news cycle had taken over, and news coverage began to focus far more on horse-race polling as opposed to candidate issues. Who was in the lead on this particular day? Did this lead grow or shrink since 7 days ago? Since 30 days ago? What does this week's polling say about an election that is a month away? These questions moved political polling to the forefront of the election season, particularly the 21st century's hotly contested presidential elections. So how are political polls doing in presidential elections? For the most recent elections: not as good as hoped. This begs the question—with all the world's current technology, can pollsters identify why polls have accuracy issues and can they fix them?

Introduction

Poll accuracy during the past few elections has been inconsistent, ranging from large to small differences in predicted margin and favoring one party over another. Achieving better accuracy and consistency in polling means surveying the most representative sample possible. However, random sampling is impossible at the magnitude of a national election; therefore, pollsters must use the best possible survey methods. This study explores how polling firms and the media could reassess poll conduction, thereby giving the population a clearer picture of voter intentions. In turn, polling firms can report more accurately and quickly and spend fewer resources on polling. Each election offers a unique set of circumstances that might influence voters differently than a previous election, including political climate, candidate quality, and economic indicators. This study seeks to show that some factors might weigh strongly on the difference between the polling margin and the election margin in recent elections.

Background of the Problem

News organizations conduct some political polls while political campaigns sponsor others with accuracy as a top priority. The results of the polls help inform campaigns on how to invest resources, including time, money, and human capital. News outlets also have a large stake in polling because of the expense of conducting a poll. Inaccurate polls can erode voters' trust in mass media and lower advertising revenue, which could have an impact on what future endeavors that outlet can pursue. Most of all, voters might choose to abstain from voting if they see wider polling margins than what is hypothetically accurate.

For example, the past two U.S. presidential elections have been very close. In 2016, the tipping point state in the presidential election was Pennsylvania with a margin of 44,292 votes out of 6,165,478 cast statewide, or 0.72%. In 2020, it was Wisconsin with a margin of 20,608 votes out of 3,297,352 cast statewide, or 0.62% (National Broadcast Company News [NBC], 2016). Voters who could have changed the election might have stayed home because of inaccurate polls.

One of the main issues in polling margin accuracy is that of the undecided voter. A polling margin of plus-5 for a candidate hardly means anything if 20% of voters are undecided. Hypothetically, this could lead to margins from minus-15% to plus-25% for the same candidate. Knowing how to split these undecided voters could lead to better accuracy. Kimball (2020) looked at three different statistical methods of splitting these voters, and his research suggests that any of the methods studied would be sufficient to allocate these voters. However, this might only be true with a small undecided vote share, say 7%. Kimball mentioned that a reported 7% undecided voter share within 10 days of the election might be truly closer to 16% because of voter indecision.

Other studies completed after the 2020 presidential election target representative sampling as the main factor of inaccuracy. Gelman (2021) says that polling firms made strides in accuracy between the 2016 and 2020 elections pertaining to their sampling methods. However, he also states that there are other factors that pollsters are not considering, such as support bias. His idea—if a candidate has an excited voter base, then they will be more likely to respond to surveys. So, in 2020, the typical voter for Democratic presidential nominee Joe Biden would have been more likely to respond to surveys than a typical voter for Republican presidential nominee Donald Trump. Gelman continues to say that many Biden supporters were not excited about voting for him, per se, but excited to respond negatively about Trump. This, compacted with low response rates to surveys from isolated Donald Trump voters (Cox, 2020), might lead to more error than any other factor. Gelman suggests weighting this response rate similarly to how pollsters already weight for demographics; however, he does not give a concrete way to do this and notes that there are ethical considerations in doing so.

One other major factor in modern polling has been how to find the voters to sample. Through most of the 20th century, polling had been done mainly over the phone with a live pollster and a live respondent. With the growth of the internet and mobile phones, voters do not answer calls as often. Some voters might not want to, or have time to, stop and take a survey call lasting as many as 30 minutes.

Currently there are several other polling types being used, including interactive voice recognition, or IVR; text-to-Web, or SMS; and online-only survey panels. Some polling firms are still using live caller surveys. After the 2020 election, Kimball and Holloway (2022) investigated some of the differences in these methods, as one method might lead to a more representative sample. In theory, a sample more representative of voters would reduce the difference between poll margin and vote margin. Kimball and Holloway's study suggested that SMS-to-Web surveys reached the most representative audience. An SMS-to-Web survey starts as a text message sent to a potential respondent's mobile phone. The text message explains that the hyperlink within the message will take the respondent to an election survey. The user then decides whether to respond. Although this method created a more representative sample of the population, it did lean toward having too many college-educated voters, a segment that usually skews toward the Democratic candidate.

Kimball and Holloway (2022) suggested that IVR reached older, slightly more conservative voters, and more rural voters than SMS-to-Web did. Additionally, the study suggested that the online panels reached mostly urban voters. While this study did not detail whether a method produced more accurate polling results, it did propose ways to change sampling methods for less representative samples of the population. For example, a pollster could use mixed-methods sampling, gathering more urban and college-educated voters with SMS-to-Web, then call voters using IVR to gain a more rural portion of the sample inclusive of isolated and older voters.

Problem Statement

The inaccuracy of political polling in recent elections, namely in the past three presidential elections shall be addressed during this study. There have been suggestions that certain polling factors have larger than suspected importance in polls, leading to inaccurate results when compared to the actual election results. These include college-educated voters being over-represented (Kennedy, 2020) and the uncertainty of undecided voters as the election approaches (Jackson et al., 2020).

While political scientists and journalists have posited many theories about why presidential polling has been subpar recently, the specific set of variables being used in this study has not yet been investigated.

Purpose of the Study

There is one main objective in the study: the researchers will investigate whether certain characteristics of polls and demographics affect the absolute difference between polling margin and election margin.

Research Question and Hypothesis

Are margin of error, percent of undecided voters, age, race, education, days from election, and poll type significant predictors of the absolute difference between a poll's predicted margin and the election margin?

H_0: There is no relationship between the absolute difference of a poll's predicted margin and the election margin regarding the predictors of poll margin of error, percent of undecided voters, difference in age from poll to voter demographic, difference in education from poll to voter demographic, difference in race from poll to voter demographic, days from election, and type of poll.

H_a: At least one of these variables has a relationship with the absolute difference of a poll's predicted margin and the election margin.

Research Design

The data collection occurred during the past three U.S. presidential election cycles. A majority of polling conducted from the 2012, 2016, and 2020 presidential elections is made publicly available at the Web sites fivethirtyeight.com and realclearpolitics.com. The data will be entered into Microsoft Excel documents. Additionally, some polls did not have all the data necessary to include the poll in the study. After contacting the pollster, the poll was either dropped from the study or kept in the study if the data from crosstabs was obtained. The independent variables consist of margin of error, percent of undecided voters, age, race, education, days from election, and poll type. Publicly available data represents the following dependent variable: the absolute difference between polling margin and vote margin; and the following independent variables: margin of error, percent of undecided voters and days from election. Data from polls concerning age, race, education, and poll type must be obtained from crosstabs or demographics sheets.

Analyzed data will consist of any poll that ended its survey 37 days to 7 days before the presidential election in the years of 2012, 2016, and 2020. The data encompass states whose actual election margin would have been categorized as "likely" or "lean" for one candidate or as a tossup during polling. This allows for analysis of only states that might be in play for either candidate and not for more-or-less predetermined outcomes. For this study, the researchers will use a cutoff of 10 percent as the likely margin.

Because this is a multiple linear regression study, the F-Test will be used. For additional analysis, other techniques might also be used, such as a χ^2 test for analysis of variance. The statistical software R will be used to perform these tests. (R Core Team, 2021). Power analysis will be computed using G*power (Faul et al., 2009).

Ethical Considerations

The ethical considerations for the researchers are few in this case. Other firms collected the data that the researchers are using, and the data is available to the public. No identities are linked to the surveys. Thus, there are no live subjects in the study and the study was exempt from the institutional review board.

The most important ethical consideration revolves around the accuracy of data entry, ensuring no bias to either political party. As there are only two nominees being investigated in each election, any improperly entered or fabricated data would be unethical. When possible, the researchers will cross-check the polls between the two main data-collection Web sites being used.

Theoretical Framework

There are three main voter theory models, sociological, psychosocial, and rational choice, of which two apply to the study: the sociological and the psychosocial.

The sociological model revolves around how social groups, and the media—including news and advertising—influence a voter's choice. The group conducting this early research, Lazarsfeld et al. (1944), found evidence that the media does not influence voters as much as the voter's social group does. Fewer than 10% of voters changed their mind about candidate selection during the 7-month study. Mostly the voter's socio-economic status, religion and area of residence impacted vote choice

(Antunes, 2010). Further studies updated in the third edition of Lazarsfeld et al.'s work (1968) suggest that people watching election news closely had already made their choice long ahead of the election date. However, if an unsure voter belonged to a social group featuring a respected member with strong views about the candidates, the unsure voter might be more likely to side with the person with strong views. For example, a business owner in the city might represent a respected member of a social group.

The sociological theory can explain long-term voting habits, but not why a voter might change party for just one election. Other voter models might be able to better explain this (Antunes, 2010). The psychosocial model of voting behavior fills in many of the areas that the sociological model cannot explain.

Several studies by the University of Michigan in the mid-20th century suggested that partisanship shares an extremely strong link with voting habits. Much of this partisanship does come through socialization, like the sociological voter theory. The authors liken this to choosing a religion. Partisanship will not simply choose a candidate, though it does give a high likelihood of the probable choice (Antunes, 2010). In current polling, much of the data in crosstabs show that this is true, also. Democrats and self-identified liberals tend to vote for Democrats, while Republicans and self-identified conservatives tend to vote for Republicans. However, the middle third of people garner the most interest from pollsters.

Partisanship can give a lens for the voter to look through; however, it will not mold how a voter feels about candidates. This will happen through many lenses, including a social lens, economic lens, election issues, specific candidates, and other factors. These lenses will ultimately help the voter make a choice (Antunes, 2010).

This could help explain large gap changes in certain states between the 2012 to the 2016 presidential elections and might also help explain vote preference changes among some demographics during this same time frame.

Assumptions, Limitations, and Scope

The population for the study will be future state-level polls regarding U.S. national elections, namely presidential elections. However, the results might also generally be applied to polls regarding national Senate elections, though that generalization remains unclear at this time and will not be answered by this study. The sample will be a couple of hundred previous state-level presidential polls from 2012, 2016, and 2020. Because the data comes from only three elections, generalization might not be possible to future elections, as more data would need to be collected to confirm any findings.

There are several issues that can limit the reach of the study. Although the data summarizes a nationwide presidential election, these types of elections are dynamic from year to year based on political environment, candidates running for office, incumbency, and possibly other factors. So, while there might be some broad knowledge gained from this study, it might not apply to every future presidential election, or even one future presidential election.

The only other type of election whose data is collected in the same way as this study is the statewide race for U.S. Senator. Some Senate elections might coincide with the results of this study, but a comparative study with several cycles of Senate elections would have to be conducted using the same variables and statistical methods. Then researchers would use an analysis of variance to compare to this study to verify.

Definitions of Terms

pollsters: a person or agent who surveys voters to find out for whom they are voting, or to obtain other information or opinions

horse-race polling: a type of election coverage more akin to a horse race because of the focus on polling data and public perception instead of a focus on candidate policy

interactive voice recognition, or IVR: a phone technology created such that the respondent can access information and respond to a prerecorded message without speaking to a live person

text-to-Web, or SMS, polling: a text message sent to a potential respondent's mobile phone that explains an embedded hyperlink within the message; the message will take the respondent to an election survey

representative sampling: a sample that mimics the characteristics of the population that is being studied

poll's predicted margin: for this study, the difference in percentage points between candidate A and candidate B in a singular poll

vote margin: for this study, the difference in percentage points between candidate A and candidate B in a state's election.

poll type: the method of data collection used by the pollster, also known as poll mode; for this study, there are three types: phone—which includes IVR, online, and mixed methods

margin of error: the random sampling error in the results of a survey; this percentage is based on the number of people in the sample

absolute difference: the absolute value of the difference of two numbers *crosstabs*: a tool that is used to compare the result of one variable to the result of another variable

likely margin: an election whose outcome based on polling models is thought to be between 5 and 10 percentage points

lean margin: an election whose outcome based on polling models is thought to be between 1 and 5 percentage points

tossup: an election whose outcome based on polling models is thought to be between 0 and 1 percentage points

state-level polls: a poll whose sample represents a state, as opposed to a locality or a country

incumbent: a person holding an office or a position that is up for re-election

Summary

Journalism has been changing so quickly that the accuracy of polling dominates headlines during election season, necessitating the accuracy of polls to be nearly perfect. However, for some yet unexplained reason, polling has become less predictive during the past few U.S. presidential election cycles. This study seeks to determine if any of the studied variables could be at the forefront of this lessened accuracy.

Time, money, careers, and policy decisions could be at risk because of polling inaccuracies. Pollsters spend their resources trying to obtain a representative sample, thereby increasing the accuracy of their polls. This, in turn, informs the public and increases the firm’s prestige and the public’s confidence in it. If polls continue to be too inaccurate, there will be a decrease of trust in journalism and consequently a decline in democracy. High-stakes polling must be more accurate—too much is on the line.

CHAPTER 2

Introduction

Before digging into the study at hand, it is important to discuss several parts of polling, some of which lie directly within the control of pollsters and others that do not. First, a discussion of differences between polls, aggregates, and models will take place, setting the table for what polls can do well in singularity and as a group. Following that will be two related sections on the recent history of polling and some ideas about what polling has done well and poorly. Interactions among polls, the media, and the candidates highlight the implications of recent polling problems. Finally, a short section on what has been tried, a failure or not, to make polls more consistent, and what polls face in the future.

Political Polling Modeling and Aggregation

It is likely apparent to any American who pays a modicum of attention to politics that polls provide information to several groups: candidates, the media, and the public itself. But a poll reveals a statistic that simply shows a relative snapshot of a moment in time. However, aggregators take data from several polls and turn them into a forecast or model, similar to say, meteorology. Combining these forecasts can also be quite powerful and more predictive than a single poll (Graefe, 2023). The idea of polling aggregation, polling models, and polling methodology makes this possible.

Polling numbers receive influence from dozens of factors in the population—most specifically from the population itself. Political polling deals with a nonrandom sample because a pollster does not know what the upcoming electorate will be; they make a best-guess decision as to whom it will include. One of the main issues pollsters face today, and they will always face, is the issue of the voter who says they will vote in an upcoming election, but then does not vote, or vice versa, because respondents to polls almost always overestimate their likelihood to vote (Clinton et al., 2022). This alone makes the sample frame and identifying the population impossible. In general terms, political polls do not represent their population, they just try their best to emulate it (Jackson et al., 2020).

But that does not mean polls are useless, and in fact, scientists do care about the actual credibility of polls (Dawson, 2023). Aggregators can use credible polls to make a forecast that has much more power and use than a single poll. However, if a systematic or one-directional bias exists, modelers will have a tough time making a good forecast, which happened in 2016 (Jackson et al., 2020). Running counterpoint to that, forecasters put their eyes on hundreds or maybe even thousands of polls during a presidential election cycle. They have a full-forest look at the broad election picture, whereas a single pollster might not see the mistakes they are making. So, an aggregator, such as FiveThirtyEight—also known as 538—will correct for these individual biases and might be able to see a systematic bias as well (Barnett and Sarfati, 2023). Additionally, aggregators who develop models have slightly better results than a simple aggregator that averages a period of recent polls, although forecasters have their own share of problems, such as undervaluing incumbent advantage (Rothschild, 2009).

When comparing the 538 model to a simple mean of polls such as RealClearPolitics, 538 performed better in all cases during the 2020 election than did RealClearPolitics. This comparison used the closest polls or model to Election Day. The farther away from the political center each state was compared to the rest of the country, the worse 538 did. The 538 model itself said it would predict 91% of states correctly, but outperformed its prediction, at 94%, with an average error of 1.9% per state. RealClearPolitics does not adjust for any biases and simply averages polls. Its polls-only numbers underestimated Trump's likelihood to win even more, specifically in tipping-point, or tossup, states. There was also more absolute error than in 538's model, and RealClearPolitics' mean gave Trump 77% of the vote share that he received, which shows a small systematic bias in its aggregated polls (Barnett and Sarfati, 2023).

Since 2004 a project called PollyVote has looked at polling forecasts, and it recently began to combine aggregators and forecasts to try to make an even more accurate representation of what single polls are trying to find. They use multiple aggregators and forecasts from different sources because a good forecast in the past might not perform as well today, and a poor forecast in the present might be refined enough to be the best predictor in a later election. Aggregating all possible forecasts and other sources lowers error (Graefe, 2023). Besides polling aggregators, other sources include expectations from betting markets, experts, and citizen forecasts (Armstrong and Graefe, 2021). This is particularly noteworthy because even though the average voter is not good at picking what a good forecast is (Soll and Larrick, 2009), they are generally good at predicting who will win the election, regardless of who they want to win (Lewis-Beck and Tien, 1999). Experts fare much better, as 62% of experts knew the direction of the error in the presidential election from 2004 to 2016, but they generally predicted a higher error than what existed (Graefe, 2018). In pre-election modeling of PollyVote during August 2020, its own aggregation using its mixed-models forecast had a less than 0.1% error compared to four other types of models (Armstrong and Graefe, 2021).

The reason to aggregate the models and markets follows the same logic as the aggregation of polls. The accuracy of a combined forecast model will always be at least as accurate as randomly choosing just one (Larrick and Soll, 2006). This happens because aggregators want to have negative correlations among forecasts so that errors cancel. However, in election polling the error is usually positively correlated because most pollsters use similar techniques (Graefe, 2023).

Furthermore, the aggregation of forecasts, markets, and predictions can be refined over time to make this “super aggregation” even more attuned. For example, PollyVote completely eliminated historical data in its 2020 aggregation as it seemed to perform negatively as an indicator. The PollyVote model ran, starting 100 days from the election, with the same tools as used in elections from 2004 to 2016 and with the new 2020 model included. It also calculated the 2020 model alone, with a side-by-side comparison being made. The 2020 alone model reduced error by 8%, adding to evidence that the “super aggregation” can be refined just like a regular aggregation or a poll (Greafe, 2023).

But why did PollyVote use 100 days from the election as a starting point? A large amount of literature supports this, including Graefe’s (2023) own work that states polls tend to be less accurate the farther away they are from the election. Polls conducted at the same time can vary widely, but differences in individual polls often cancel when aggregating. Campbell (2022) adds that survey technique did not matter when polling the week before the survey and that accuracy improves as the election draws near.

Election forecasters performed well, even 2 months before the election, but only on the national scale. They did not fare as well in the electoral college, which aligns with state polling (Jackson et al., 2020).

Graefe (2023) concludes, for PollyVote and for other aggregators, “when evaluating forecast accuracy, decision-makers should focus on longer forecast horizons.” He adds that while PollyVote was not the most accurate in the 2020 election, it did much better than many other models and aggregators. He finishes by stating that historical accuracy within polls or even models is fleeting and that aggregators should focus on a diversity of sources to reduce bias and error. So, what exactly have polls looked like recently, and what have they done to combat error?

Recent Polling History

Recency bias would have one believe that polls used to be perfect and predict elections with sparkling accuracy. And while the golden age of polling exists in the past, polling has a history with many misses.

In the weeks after the 2020 presidential election, discourse about polling dominated the headlines. And while some election-eve polls had Biden winning by a 10% or 12% margin, he only won by a healthy margin of 4.5% (Campbell, 2021). But unlike 2016, 2020 polling predicted the winner correctly. Polls in 2012 missed worse than in 2016—so much so that one of the most prolific pollsters in the U.S., Gallup, discontinued horse-race polling after that election (Campbell, 2022). But do not forget about the most historic miss: the election night headline in 1948 of “Dewey defeats Truman” despite an eventual 100 electoral college vote victory for Truman.

Disregarding the history of nearly a century ago, 2016’s presidential election of

Hillary Clinton versus Donald Trump remains one of the worst nights for polling since 1980. More than half of the national election-day polls in 2016 were outside of the margin of error, where normally only about 1 of 11 would be expected. State-level results proved far more inaccurate with a bias for Clinton at about 5% on average. Polling in battleground states mostly underestimated Trump, in 13 of these 15 states, with most underestimating by more than 4%. Part of this can be explained by national exit polls as they showed 13% of voters decided for whom to vote within a week of Election Day and about a quarter within the last month. These late deciders broke for Trump by 3 and 8 points respectively. In the important Great Lakes states of Michigan, Pennsylvania, and Wisconsin, those voters broke for Trump between 17% and 29% depending on the state (Jackson et al., 2020).

Finding respondents to take polls has increasingly become more difficult as response to polls settled in 2016 after declining for 20 years. Nonresponse will be discussed in a future section of this chapter, but it is important to note that the larger the nonresponse percentage is, the more impact it will have on the final estimate (Durand and Johnson, 2021).

Recent studies in nonresponse have centered on how to get the most representative sample possible while still knowing that the sample obtained is nonrandom because pollsters do not know who the population contains. As such, modes of polling have encountered major transitions during the early 21st century—from telephone polls in the early aughts to IVR, online panels, text to SMS, and mixed-methods polling in recent years (Hillygus, 2011).

Some statistics in a 2021 study from Durand and Johnson indicate trends that might help pollsters obtain a sample that better represents who will vote in an upcoming election. One of the oldest tried-and-true methods, live caller polls, tends to have a larger nonresponse from conservative voters. IVR calls to cellphones have been banned since 2016 in the U.S., even though they can still be made to land lines; therefore, live caller polls must be made to cellphones. However, these live interviewer polls decreased from 89% of polls in 2008 to 14% in 2020. IVR peaked at 24% in 2012 despite findings that they produce the most accurate results, specifically compared to online-only or live caller-only polls.

Additionally, in 2008 and 2016 IVR caught trends in future voting preferences, including a reduction in Hillary Clinton's vote share as Election Day neared. Additionally, in 2020 IVR and telephone polls were able to catch other trends that online polls did not. However, because online polls accounted for almost 80% of polling, most aggregators overestimated the online sector, which showed more positive results for Biden. From 2004 to 2020, IVR polling produced more accurate state-wide results, other than 2012. IVR also exhibited lower results for Democrats than other modes, which has been one of the largest downfalls for pollsters in the past two elections (Durand and Johnson, 2021). However, a 2018 midterm election study showed that IVR did tilt toward more Democrats with both random-digit dialing and registration-based polls (Clinton et al., 2022).

Even though IVR seems to be the most accurate mode, it must be used sparingly due to how difficult it is to conduct. What can pollsters do to filter the voters that IVR would usually capture? Maybe not much. But other methods have been tried to combat nonresponse and sampling problems and have worked to varying degrees.

Why Has Polling Been Less Reliable?

In the previous section, the discussion suggested that polling mode might have a significant impact on the types of voters that get selected for a poll. Some recent changes such as the restrictions on IVR seemed to negatively affect polls, but mode alone did not make polls miss by the margins they did. And the science behind why remains unclear.

The consensus best guess from research regarding the 2020 presidential election results suggests that pollsters largely underrepresented right-leaning voters (Campbell, 2022). Along with 2016's election polling problems, many have posited that education was weighted inaccurately, specifically in the blue wall states. Here, a disproportionate amount of White, noncollege-educated voters was left out of polls, the exact voter base that led Trump to victory (Clinton, et al., 2021). Additionally, after 2016, pollsters believed they had found the answers to make 2020 presidential polling more accurate, but many of these changes did not help (Campbell, 2022). Polls in the 2018 midterms closely resembled the outcomes of the election, which begs the question: Could it be that polling misfires connect specifically with Trump's presence on the ballot (Keeter, 2021)?

What other ideas have pollsters had regarding the 2020 miss? Many have called the 2020 errors systematic (Barnett and Sarfati, 2023) or claimed a systematic bias because the polling error was one-directional (Noble, 2021). Noble continues that weighting variables is unlikely to have large enough strength to overcome bias. Thus, he claims, there must be a systematic bias, and this bias is not an underrepresentation of right-leaning voters. This line of thinking runs in opposition to many other studies;

however, would possibly be in sync with a nonresponse bias. Therefore, polling systematically underestimated Trump (Jackson et al., 2020).

Barnett and Sarfati (2023) agree that there might be a systematic nonresponse bias because a post-mortem 2020 election report states that it seemed college-educated voters were not overrepresented as they were in 2016 (Clinton et al., 2021). This means some other variable must have affected the polls. Jackson et al. (2020) state that this problem stems from the complete impossibility of obtaining a random, representative sample. Their research suggests that even examining past voter lists and calling those respondents likely voters does not fix the random sampling problem, and both quota and de facto sampling raise a gamut of issues with obtaining a good sample.

With so many ideas about what did not happen, there exists little information about what did happen. However, a 2022 study by Clinton et al. provides quite a bit of insight into a possible mixed-weights approach to adjusting poll results:

> multiplying post-stratification weights by the inverse of the partisan cooperation rate to equalize the cooperation rates across partisan groups of voters reduces the average polling error on the final certified margin by 4 percentage points in the six states where a sizable pre-election phone poll of the electorate was conducted. (249)

So, the bias or error was in part due to a lack of cooperation during the polling process from Republican and independent voters.

An analysis of phone polling showed that, as stated earlier, statistically weighting the sample was too weak to solve the problems that the polls had. National election polling data of 12 states showed that likely Republicans were 3% less likely and independents 6% less likely to complete a poll than likely Democrats, even after controlling for other variables (Clinton et al., 2022). And although this percentage seems small, in larger samples this would omit between 50 and 100 voters, a majority of whom would likely have voted for Trump. Independents affect the bottom-line data less than Republicans, but it nonetheless leaves a sizable difference.

Clinton et al.'s method, while possibly useful and easily implemented, does have limitations. The largest problem with these missing voters remains the fact that pollsters do not know much about them, mostly because they refuse to cooperate. Due to demographic data, researchers do know that the refusal to cooperate has no correlation to age, gender, race, or rural and urban status; however, there is no mention of the relation to education level. This could be a major connection as the weighting technique that proved effective in some states was not effective in others. For example, even after this correction, whiter and less-educated states still had uncorrectable error. In this case, post-poll weighting and even stratification will not completely fix the problem (Clinton et al., 2022).

Before diving deeper into education, nonresponse and the possible ties between them, there must have been a reason why these two characteristics of polling keep popping up. Part of it relates to how polling firms, candidates, voters, and the media interact with one another.

Pollsters, the Public, and the Media

Political polling importance can boil down to several factors, one of which is how the public interacts with and perceives polls, usually vis a vis the media, but also the pollsters themselves. Of course, public opinion polling's credibility relies completely on the public's belief in the accuracy and fairness of the polls and the willingness of the public to participate in the institution. At this point, change defines best how the public interacts with pollsters and absorbs polling information through the media.

Voters have to decide whether to trust the polls—but trust does not mean truth. For example, individuals can view polls as credible if the results seem to be true to them. Several recent studies have suggested that polls conducted by media with a political slant similar to the respondent in question seem to be more credible (Searles et al, 2018). This has been shown in many countries, from Turkey and its multi-party system (Dawson, 2023) to the two-party stranglehold in the United States (Kuru et al., 2020). This becomes a problem when a member of a political party expects their candidate to win because the polls they trust have said so. This is an echo chamber and hurts democracy (Price and Stroud, 2006). However, Dawson (2023), suggests that polls with large leads in favor of one's own party can decrease credibility, seeming "too good to be true."

Doubling back to the study conducted in the United States, several other important key features were determined based on participants' own responses to what a scientifically good poll looks like. This might run in contrast to what the participant wants the poll to say, regardless of poll quality. In the study, two polls were compared, sometimes with matching quality, sometimes with differing quality, and within this comparison the margin of lead was varied. When both polls were high quality, almost 7 of 10 people said both polls were accurate; but if one was poor quality, 6 in 10 still said they were equally accurate. And while Democrats were more likely to say Clinton would win over Republicans saying Trump would win—based on the polls—when education was controlled those that thought the Trump poll was more credible overwhelmingly said that Trump was going to win. The good news for democracy? A comparison of two equally high-quality polls greatly reduced this bias (Kuru et al., 2020). Overall the public would benefit from high-quality polling and these high-quality polls might reduce the negative interactions that plague the industry.

Part of the negative feeling that some of the public has might have stemmed from the candidates themselves, particularly Donald Trump, during the past two presidential election cycles. Celebrity candidates have run for office in the past and won, think of Ronald Reagan and Arnold Schwarzenegger, but few have put on a carnivalesque attitude as Trump has (Mohammed and Trumpbour, 2021). The “fake news” media storm started with Sarah Palin in the presidential election of 2008 and bled over into tea party rhetoric in the 2010 midterms (Noble, 2021). However, Trump ramped this up during his campaigns; more than half of media coverage was about Trump’s carnivalesque attitudes, rather than issues. This gives a possibility for important political issues to be covered only on a surface level (Mohammed and Trumpbour, 2021). The implications for polling might be enormous.

Currently, one of the most notable problems for polling is contact nonresponse, where a possible respondent to a poll does not want to take place in the polling despite being contacted. This could affect data, and it could be tied to Trump, as this, spoken by Trump (Smith-Schoenwalder, 2020) might indicate:

> We have poll numbers in Wisconsin where we’re up one, and yet I see ABC comes out, ABC-Washington Post, of course they had us 12 down last time . . . and we ended up winning . . . It’s a shame they can get away with it. If you think about it, it’s almost like a campaign contribution to the DNC. The good news is our people understand it. They understand it very well.

Trump’s allusion to mistrust of polls is clear, and Noble (2021) agrees rhetoric like this is why Trump voters tend to distrust public institutions and the mass media.

Only 31% of low-trust voters believed that no matter who wins, they will accept the defeat (Noble, 2021). Before 2016, there was no difference in nonresponse between partisan groups in at least 20 years, which adds evidence to the tie of the nonrespondent to Trump (Keeter et al., 2017). Furthermore, this might not generalize beyond the 2020 election because Trump specifically told his voters not to cooperate, and it might be less salient in midterm elections (Clinton et al., 2022); however, there is also slight evidence that people did not want to do a poll if their party was portrayed negatively in the media at the time (Gelman et al., 2016). A task force studying the 2020 polling results could not pin one direct reason why polls missed, naming nonresponse as the most likely culprit (Campbell, 2022). “Without knowing how nonrespondents compare to respondents, we cannot conclusively identify the primary source of polling error” Clinton et al. (2021). Nonresponse creates a huge problem for polling firms.

But herein lies the overlap between media, polls, and voters. Polling provides a forecast to the public about who the most important person in American politics will be, thus they can cause a bandwagon effect and a turnout effect (Barnett and Sarfati, 2023). Polls stating that Clinton leading by a large margin of 4 to 5 points the day before the election might have depressed Democratic turnout (Westwood et al., 2020). Close elections have higher turnouts and sometimes pollsters use a lower-turnout methodology if elections are not deemed as close (Barnett and Sarfati, 2023), with all these interactions exacerbated by nonresponse.

Several studies have stated that it might not be a net negative if polls and media are incorrect and might benefit the public in the long run. Barnett and Sarfati (2023) explain that political polling gives the public a horse-race view rather than a more valuable issues-based election coverage. They also counterpoint themselves, saying that close races do give candidates a reason to speak about policy differences (Barnett and Sarfati, 2023). Stopping pollsters from releasing results near Election Day might also benefit the public. France bans polling a day before the election and Italy does so 15 days before (Jackson et al., 2020). Still others argue that polling late in an election cycle is more or less useless: “Apart from providing entertainment and informing last minute voters, Election Eve forecasts are of limited practical value” (Graefe, 2023).

Despite polls being more accurate in 2016 than in 2012, outcry and discourse arose from the media and the public in 2016 because polls did not predict the correct winner. “Polls that get the answer right, but still have considerable error, are considered ‘okay.’ Polls with small amounts of error that miss the results are considered bad” (Jackson et al., 2020). As America prepares for a presidential election in 2024, what are some ideas pollsters have to close the error gap?

The Future of Polls Heading into 2024

Polling seems to be at an all-time low point, or if not, seemingly nestled into a valley with diminished public trust and myriad problems accounting for this. However, this downturn in the industry has led to more creativity and less resting on laurels. While there have been a few trends to emerge in the past 10 years, there has not been a consistent answer, hence this study. But experts have talked in depth about what might happen in the coming decades.

Accurate polling will always have a place in American democracy (Campbell, 2021); however, pollsters must learn to evolve. Experimentation must be the next stage in polling, using many techniques or methods to refine what pollsters currently use. For example, in the past pollsters went door-to-door, then used phones, and now they are using the internet. All of this was evolution, and this evolution happened because, as pollsters have stated in the past, the best way of finding the public's opinion on something is to ask them a question (Campbell, 2022).

Polling evolution could mean using aggregators as a secondary check. It is probably better to think of polls as a raw good that aggregators or models can use to refine these polls into a usable product. Polls can be deemed successful if an aggregator or model can properly use them to predict an election. But aggregators must rely on at least somewhat accurate polls to develop a model. Aggregators have started to evolve with polls and did a better job in predicting presidential election results in 2020 than they did in 2016, even though both pollsters and aggregators still underestimated Trump. Accounting for positive momentum, this does not mean that internal polling bias or systematic bias has been corrected. However, as of 2024, polling, and specifically aggregators, should not be considered unreliable (Barnett and Sarfati, 2023).

One of the possible sticking points in polling accuracy is how to solve the existence of nonresponse bias, which has grown in recent years with the advent of cell phone usage, which has also made recruitment and randomness into a unique issue (Durand and Johnson, 2021). Some studies suggest that the systematic nonresponse bias cannot be weighted out; others say that there might be ways to do so.

According to Clinton et al. (2022), cooperation from nonresponse is a new and different problem that post-poll weighting cannot solve as many demographic weights are barely tied to party affiliation. Recent nonresponse can also be tied to distrust in media, which some voters are proud of, thereby making polling worse. (Noble, 2021)

Even systematic bias weighting in the study using nonresponse proportioning still left large errors in swing states; only a third to half of the total polling error was reduced. This systematic bias weighting also does not have broad implications as the study focused only on telephone interviews (Clinton et al., 2022). On the other hand, another study by Jackson et al. (2020) says that nonresponse bias weight should be low.

Even though a low weight on response rate should be sufficient, this tactic might be the best way to correct error, purely because systematic errors tend to be one-directional (Jackson et al., 2020). Response rate increase, in practice, was shown in British polls to correct systematic error very well (Sturgis et al., 2016). An American National Election Study where pollsters went door-to-door increased the response rate almost tenfold, but even those respondents voted in lower numbers than they indicated during polling (Jackson et al., 2020).

Other methods of polling to find out who will win an election seem a bit more underhanded, such as asking voters who they think will win the election rather than who they will support on Election Day. Using American National Election Study data, citizens asked “who will win the election?” correctly predicted nine of eleven elections from 1956 to 1996, missing only close contests in 1960 and 1980 (Lewis-Beck and Tien, 1999). This question also worked exceedingly well at the state level in eight of the nine elections leading up to 2012. Moreover, these state data values were drawn from a national sample and then subdivided into states, so there would be no additional cost for the data (Murr, 2016). However, forecasting an election in this manner would be a malpractice in polling. It could be an extra question after the main survey though, as forecasters often use other data than polls alone (Jackson et al., 2020).

Sampling remains the main problem for pollsters, even with those who do respond. It is impossible to get a real sample frame for an election yet to occur because the pollster does not know the population. In these polls, there are yet more sources of polling error outside of the margin of error, such as polling people who say they are likely to vote, but do not vote. Think of this as the opposite of an unlikely voter but whose impact is like that of an undecided voter. One way to combat this might be to release several poll results from the same poll with different definitions of, and sample sizes for, likely voters. For example, there might be three released results from a poll, one result for people who said they were “7 or higher” likely to vote, a second for “9 or higher” and a third for “10, definitely.” Pollsters would publish these results and methods with transparency to the public (Jackson et al., 2020).

Conclusion

Although polling has had its ups and downs and has been severely panned in the past decade, aggregators and the media can make good use of raw polling numbers. Even if one single poll is incorrect, there are others that might be using slightly different techniques that would yield better results, which helps aggregators and the media. Furthermore, the problems of today might be solved in time for the next election, or the election thereafter. The following quote shows that although problems plague the current polling landscape, polls will still play an important role in the future.

> Not only is polling deeply engrained in American political life and American media, it has survived acute embarrassments of the past. It may be a platitudinous observation, but if election polling outlived the ‘Dewey defeats Truman’ fiasco of 1948, it certainly will survive the high-profile failures of 2016 and 2020. (Campbell, 2022)

CHAPTER 3

Introduction

From the previous chapter, it is clear that recent polls and election results have not interacted the way that pollsters would prefer. This study has one clear objective—researchers will investigate whether certain characteristics of polls and demographics affect the absolute difference between polling margin and election margin. What follows in this chapter outlines various features and qualities of the study such as the data sources, study design, procedures and other considerations.

Research Questions, Objectives, Hypothesis

Before discussing how the study will be conducted, it is important to reiterate the main question of the researchers and to state the research hypothesis. The researchers are interested in the following question: Are margin of error, percent of undecided voters, age, race, education, days from election, and poll type significant predictors of the absolute difference between a poll's predicted margin and the election margin?

H_0: There is no relationship between the absolute difference of a poll's predicted margin and the election margin regarding the predictors of poll margin of error, percent of undecided voters, difference in age from poll to voter demographic, difference in education from poll to voter demographic, difference in race from poll to voter demographic, days from election, and type of poll.

H_a: At least one of these variables has a relationship with the absolute difference of a poll's predicted margin and the election margin.

Data collection, Procedures and Context

Because the data comes from previously conducted surveys, the study was Institutional Review Board exempt. The data being analyzed in this study was collected in three batches: Sept. 30, 2012, to Oct. 30, 2012; Oct. 2, 2016, to Nov. 1, 2016; and Sept. 27, 2020, to Oct. 27, 2020. Additionally, there were dozens of pollsters who conducted these polls; therefore, reviewing the instrument used in each of the polls, or by each of the pollsters, would be a nearly impossible task. It was important that the researchers used poll mode as one of the variables as polls are conducted in several different manners. Research from chapter 2 indicated that it could be crucial to note this.

Data used during this study did not include sources who polled during the week before the date of the election because of the fear of poll herding. Herding happens most frequently when several polls from different firms are all released on or near the same data, in this case, Election Day, and pollsters want to avoid having their outlier be incorrect. According to the American Association for Public Opinion Research [AAPOR]: "To avoid raising questions regarding the accuracy of their results, some political pollsters adjust their findings to match or closely approximate the results of other polls . . ." (Herding, 2023). In no way did the researchers believe that any pollster used herding; the date range being used was simply precautionary.

Data used in this study came from states whose final margin was within 10% on Election Day in its respective year. This was chosen as the cut-off margin because the definition of a safe state is any state whose margin is greater than 10%. The research solely focused on states that were winnable for each candidate in the given election. In

2012 these states included Arizona, Colorado, Florida, Iowa, Michigan, Minnesota, Missouri, Nevada, New Hampshire, North Carolina, Ohio, Pennsylvania, Virginia, and Wisconsin. In 2016, this included all states from 2012 with the omission of Missouri and the additions of Georgia, Maine, New Mexico, and Texas. In 2020, this included all states from 2012 with the omissions of Colorado, Missouri, and Virginia and with the additions of Maine and Texas. The state of Georgia had two polls from 2012 that should have been included; however, no exit polling occurred in that state in 2012.

Using these states and years, the researchers obtained 37, 86, and 82 data points for the elections in 2012, 2016, and 2020, respectively, after deleting polls with missing data. Summing these gave 205 total polls that could be used in the study. A priori power analysis using G*power (Faul et al., 2009) stated that 205 eclipsed the 103 necessary for .80 power. This power analysis was completed using a standard effect size of 0.15, an α-value of .05, and seven predictors: margin of error, percent of undecided voters, age, race, education, days from election, and poll type.

Validity, Reliability, and Ethical Considerations

Now that the data has been clearly stated, it becomes important to note that this study does not itself occur in a bubble. Researchers across the globe should be able to generalize the procedures in this study and replicate it in the U.S. using other variables, elections, or data sets, and then report those results. Because data collection occurs from an outside source, anyone with a statistical software package, such as R, should be able to replicate it. Additionally, other countries could use this study's methodology for their presidential or prime ministerial races, where applicable.

The reliability of the data could be a slight issue within the study. Some pollsters do not publish their full methodologies, and thus they might be giving inaccurate numbers based on ineptitude alone. Others who are working with a political party might have a reason to do several polls and then publish the poll that best fits the narrative the party wants to push. There are prominent pollsters in this category on both sides of the aisle; however, this is noted by the researchers and is not to be implied as underhanded toward any pollster, partisan or otherwise. Additionally, news media and other polling organizations could also participate in these kinds of tactics for their own gains, such as what was mentioned above regarding herding.

The analysis used on the data should be as reliable as the data itself, which, as stated above, has no reason to be considered unreliable. The researchers are testing several demographic variables from voters, which are as reliable as the voters are truthful, whether during the poll itself or during exit polling as a comparison. The nondemographic variables of margin of error, days from election, and poll type are less likely to be unreliable. They are just matters of fact, whereas the variable percent of undecided voters seems as reliable as the age, race, and education questions that voters would have answered.

Finally, the researchers themselves have no ethical considerations to note. Data collection occurred several years before this analysis, and the researchers have no ties to any of the pollsters. The researchers strive to maintain honesty and truth in this process by only analyzing the data and discussing the findings. Mistakes in data entry will be minimized as much as possible and any mistakes are purely accidental.

Analysis Procedures

This study will rely on a multiple regression method of analysis of the seven predictor variables. Literature from chapter 2 suggests that some of these variables could be correlated with a large margin of error in polling, specifically education level, undecided voters, and proximity to Election Day. Other key predictors that might have correlation have also been included as possible explanations of the polling error and Election Day vote difference. Because these variables have been most often cited as possible reasons connected to polling misfires, it is important that they are examined closely. The only other possible reason experts have posited is nonresponse bias, which has a much larger scope than what is possible in this study. The researchers have additionally chosen multiple regression to link key variables, something that would be much more difficult to do using other methods. In chapter 5, there will also be a brief discussion of the differences between elections; however, that is not the primary focus of the study.

At the conclusion of the study, researchers hope that the results will yield a clearer idea of what factors, if any, have the largest impact on recent presidential election polling error. If there are any factors that seem to be relevant, it could help pollsters adjust methodology in future elections. As stated in chapter 1, this would benefit several parties and help better allocate resources for campaigns and pollsters if the polls have better accuracy.

Data analysis takes place within the statistical software R (R Core Team, 2021). Within R, the seven predictor variables will be tested using multiple regression techniques. Other studies on election behaviors have used multiple regression techniques including a 2016 rural-urban study by Ambrosius and a race study in 2021 by Buyuker, et al.

As far as data goes, only one categorical variable is tested, poll mode, which is coded as PollType, while the other six variables tested are numerical variables. A full example of a case is given in appendix A with a definition for each predictor. Because three predictors use the difference between the poll and the actual election data using demographics, exit poll demographics are used. This is consistent with a 2019 study by Bracic et al. about sex, race, and the interactions between them during the 2016 election. For example, the researchers define the education predictor, EduDiff, as exit polling data versus actual poll data, percentage polled no college degree minus exit polls no college degree.

Summary

In short, researchers have an interest in how several key predictors influence poll results as compared to Election Day results. A multiple regression technique will be used during the study to compare these variables, one of which is categorical, with the remainder numerical. This methodology has been used in other research and should be easily repeatable given the ease of securing data. Additionally, the study explores the three most recent presidential elections and could have realistic use during polling in the near future and possibly in other countries that have similar polling methods and election systems to the U.S.

CHAPTER 4

Introduction

The data presented in the following section are meant to determine if a link exists between the absolute difference of poll margin percentage and state vote percentage and several predictors. A multiple linear regression analysis method will determine if these links exist. The research question being examined in this analysis is “Are margin of error, percent of undecided voters, age, race, education, days from election, and poll type significant predictors of the absolute difference between a poll’s predicted margin and the election margin?”

Materials and Methods

A standard multiple regression was performed between absolute difference of poll margin percentage and state voter percentage (Diff), which was the dependent variable, and the following predictors: poll margin of error (MoE), undecided voter percentage (Undecided), days from election (DaysOut), poll mode (PollType), age (AgeDiff), race (RaceDiff), and education (EduDiff).

For clarity, Diff is the absolute difference of poll margin percentage and the state’s vote percentage in the election. AgeDiff, RaceDiff, and EduDiff are all demographics-based predictors, and all of them compare state exit polling data to the demographics data from each specific poll. AgeDiff takes the percentage of polled voters aged 18–44 minus the percentage of voters in exit polls aged 18–44. A positive AgeDiff indicates that more younger people were polled than what the exit polls found. RaceDiff is the percentage polled of White voters minus the percentage of voters in exit polls that were White. A positive value indicates that more White people were polled than what exit polls found. EduDiff compares the percentage of people polled with no college degree minus what the exit polls found had no college degree. A positive value states that more people without a college degree were polled than what exit polls indicated. Finally, PollType is a categorical variable consisting of three polling modes: phone, online, or mixed methods. A phone poll is a poll taken via a live interview on a cell phone or taken via live interview or IVR on a landline. An online poll is a poll taken by SMS-to-Web, email-to-survey, or an online panel. Finally, a mixed-methods poll can take any part of a phone poll and any part of an online poll and use them to obtain a sample. An example of the data obtained from one poll can be found in appendix A.

The initial sample size was n = 205. Because the number of cases per predictor exceeds 15 (Field et al., 2013), there was no concern with adequate sample size. A priori power analysis using G*power (Faul et al., 2009) stated that 205 eclipsed the 103 necessary for .80 power. This power analysis was completed using an effect size of 0.15, and an α-value of .05. Analysis was performed using R (R Core Team, 2021).

Results of Shapiro's test for normality revealed that none of the predictors other than EduDiff appeared to come from a normal population; however, because of the restrictions placed on the data, this was mostly expected. Examination of outlier cases, high standardized residuals, and influential cases led to the deletion of six cases, resulting in a total sample size of 199 cases, on which analysis was performed.

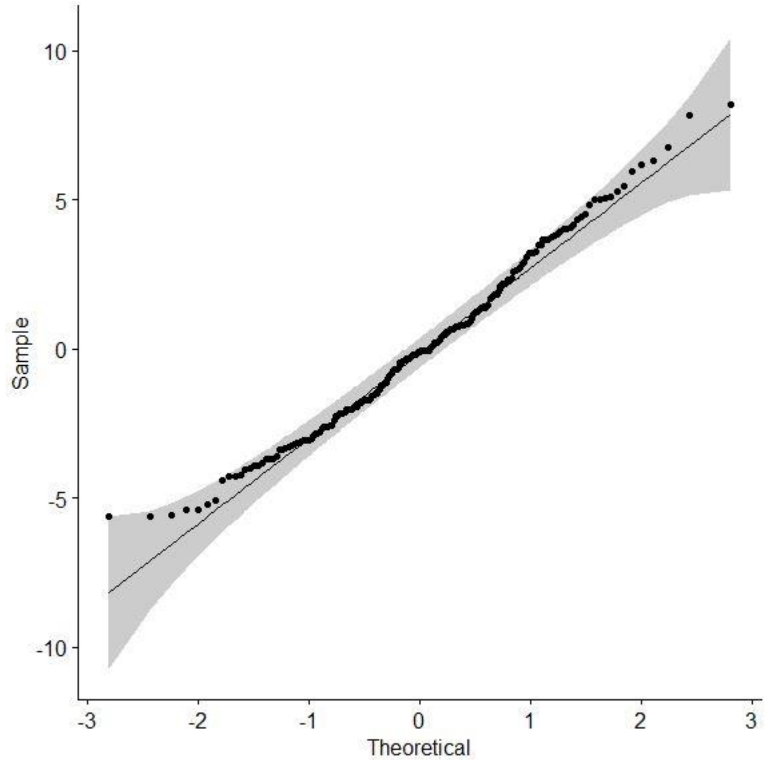
Table 1 displays the descriptive statistics of the dependent variable and each of the six numerical predictors being studied. For the categorial predictor PollType, there were three categories, with the amount preceding each type: 134 phone, 24 online, and 41 mixed method, accounting for 67.3%, 12.1%, and 20.6% of cases, respectively.

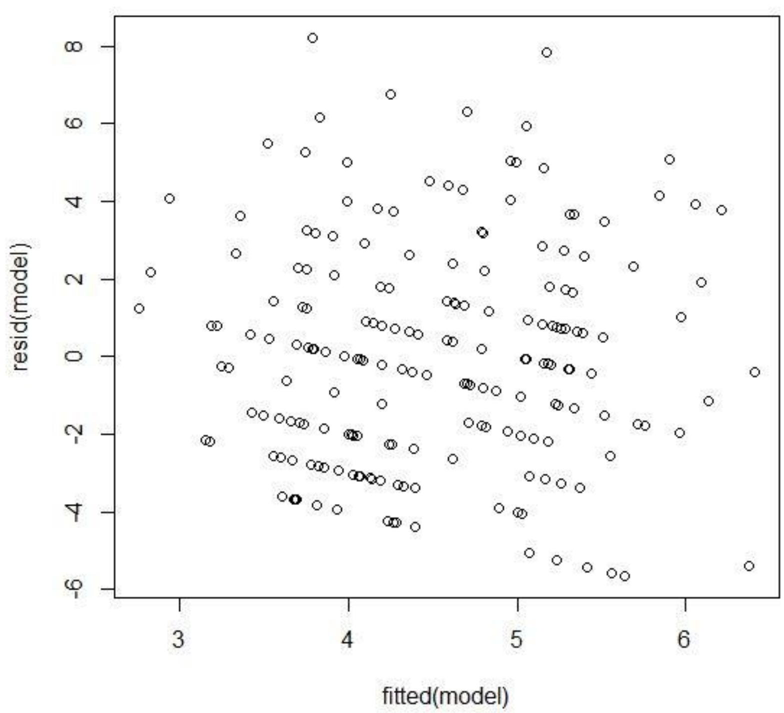
Table 1. Descriptive statistics

Variable	**Mean**	**Standard deviation**
Diff	4.53	2.97
MoE	3.89	0.65
Undecided	3.98	2.32
DaysOut	20.12	9.42
AgeDiff	-1.40	5.39
RaceDiff	1.25	4.34
EduDiff	0.52	5.67

Results of the evaluation of assumptions indicated some concerns with model assumptions. The Durbin-Watson test showed violation of the independence assumption; D-W statistic = 1.22, $p < .01$. Shapiro's test for normality of residuals revealed a slight concern; $W = 0.986$, $p < .05$. Additional plots that were used to check the normality of residuals and equal variance assumptions are presented in Figure 1. The equal variance assumption was not violated. The plot for the normality of residuals assumption shows less of an issue than the Shapiro test might indicate. Multicollinearity was examined using variance inflation factors, which ranged from 1.03 (DaysOut) to 1.25 (PollType), suggesting no issues with multicollinearity. Despite some issues with assumptions, the researchers felt comfortable proceeding with the data as a multiple linear regression study, specifically because the normality of residuals was very close to meeting the .05 threshold.

Figure 1. Plots for Normality and Equal Variance Assumptions





Data Analysis

Table 2 displays the correlations between the variables, unstandardized regression coefficients, and the adjusted R^2. A test of the full model against the intercept-only model was not significant, $F(8, 190) = 1.721$, $p = .096$. The set of predictors in combination contributed to approximately 2.8% of the variance in Diff. Only DaysOut emerged as significant in the model, at the $p < .05$ level: DaysOut ($t = 2.411$, (0.01, 0.10)). This means that controlling for all other predictors, for each day farther away from Election Day that a poll concludes, the absolute difference between polling margin and election result increases by 0.05%. Another, more useful way to phrase this would be to say that for every 20 days closer to Election Day that a poll concludes, the absolute difference between poll margin and election result decreases by 1%.

Table 2. Correlation for n = 199 cases

Variables	MoE	Undecided	DaysOut	AgeDiff	RaceDiff	EduDiff	PollMixed	PollOnline	B	SE
Diff	0.11	0.02	0.18	-0.04	0	-0.07				
MoE	---	0.10	0	0.17	0	-0.05			0.54	0.33
Undecided	0.10	---	0.12	-0.02	0.03	-0.12			-0.03	0.09
DaysOut	0	0.12	---	-0.07	0.07	0.03			0.05*	0.02
AgeDiff	0.17	-0.02	-0.07	---	-0.19	-0.13			-0.04	0.04
RaceDiff	0	0.03	0.07	-0.19	---	-0.08			0	0.05
EduDiff	-0.05	-0.12	0.03	-0.13	-0.08	---			-0.05	0.04
PollMixed							---		-0.88	0.55
PollOnline								---	-0.15	0.69
Adjusted R^2 = 2.8%										
$F(8, 190) = 1.721$, $p = .096$										

Note: *, significant at the .05 level

Because so many predictors were shown to not be significant at the $p = .05$ level, a backward model was conducted. This model kept only MoE and DaysOut as predictors. However, this model was statistically significant when compared to the null model, $F(2, 196) = 4.415$, $p < .05$. The set of predictors in combination contributed to approximately 3.3% of the variance in Diff. Just as in the full model, only DaysOut emerged as significant in the backward model at the $p < .05$ level: DaysOut ($t = 2.531$, (0.01, 0.10), $B = 0.06$). Although MoE was not significant at the $p < .05$ level ($t = 1.560$, (-0.13, 1.12), $B = 0.50$, $p = 0.12$), it was still included in the final backward model.

A one-way analysis of variance was conducted between Diff and PollType. Bartlett's test of homogeneity of variances revealed no concerns with homogeneity, test-statistic = 0.03, df = 2, $p = .99$. As stated earlier in the chapter, the normality assumption of Diff was broken, $W = 0.96$, $p < .01$, but that was mostly to be expected with the restrictions placed on the data. The results from the ANOVA revealed no significant differences in the mean of Diff and PollType, $F = 1.60$, $p = .21$.

However, this study found no online-only polls conducted in 2012, so researchers used two separate ANOVA analyses for the 2016 and 2020 elections. Both analyses yielded the same results as the analysis with all three elections. There were no significant differences between the mean of Diff and PollType: for 2016 $F = 0.13$, $p = .88$; and for 2020 $F = 0.96$, $p = .39$.

A Predictor of Note

With the mixed-methods sampling approach, it would be remiss if the researchers did not mention that in the original 205 cases model, before the six case deletions, PollMixed returned a $B = -0.95$, $t = -1.67$, $p = .097$. If the significance level had been set at $p < .10$, then PollMixed would have been statistically significant. The B-value would have stated a reduction in the absolute difference between the poll margin percentage and the election night margin percentage by almost a whole percent when comparing mixed-methods sampling to phone-only polls. The relevancy of PollMixed will be discussed along with nonresponse in chapter 5.

Conclusion

Because of the various p-values obtained during data analysis, the answer to the research question is not exact. Are margin of error, percent of undecided voters, age, race, education, days from election, and poll type significant predictors of the absolute difference between a poll's predicted margin and the election margin? The answer would be, not all of them; however, some of them became important in the backward model. Additionally, DaysOut was a significant predictor of the absolute difference of a poll's predicted margin and the election margin. In the following chapter, a discussion about these variables and what else could be researched will take place.

CHAPTER 5

Introduction

Before discussing what the results from the previous chapter mean contextually, it is important to remember why this study is taking place. The main objective of the study is to research if seven characteristics of polling have a significant impact on the absolute difference between polling margin and election margin. This study is being conducted to investigate a relatively larger polling error than one would expect in recent elections. The impact of this covers many parts of politics, from media coverage to voter knowledge to campaign resources.

From this point, a research question arose: Are margin of error, percent of undecided voters, age, race, education, days from election, and poll type significant predictors of the absolute difference between a poll's predicted margin and the election margin? The null hypothesis stated that there was no relationship between any of the seven predictors used in the study with this difference, while the alternative hypothesis claimed that at least one of the variables had some statistically significant relationship with the absolute difference of a poll's predicted margin and the election margin.

To summarize chapter 4's results, the alternative hypothesis has been accepted. The variable "DaysOut" has a statistically significant effect on the absolute difference between poll and election margins, while the other predictors did not have a statistically significant effect. According to the results, the variable "DaysOut" has a positive linear relationship with "Diff" with a coefficient of 0.05. In other words, for each day farther from the election that a poll ends, its inaccuracy compared to election results goes up by 0.05%. This means that polls conducted nearer to the election are more accurate.

Major Findings

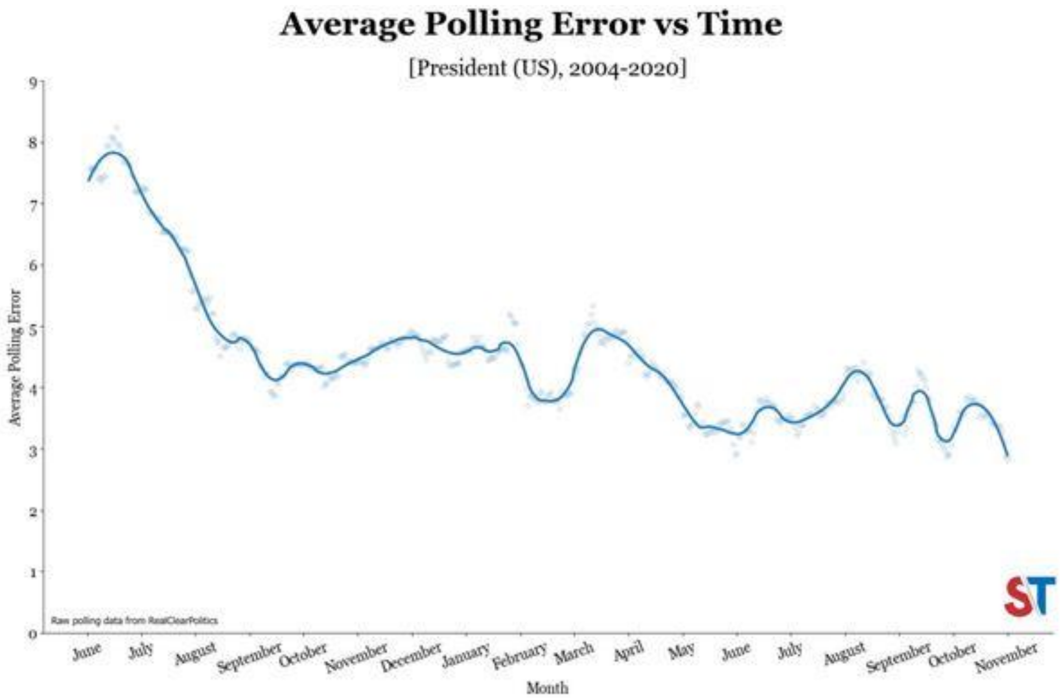
After the completion of the study and evaluation of the data, the researchers have concluded that the nearness of the end of the poll's field survey to the election date has a statistically significant linear relationship. However, at the $p < .05$ level, poll demographics have no significant impact on the absolute difference between poll results and election results, also called Diff in this study. This included RaceDiff, or what percent of White people were polled compared to exit polls; AgeDiff, or what percent of people aged 18–44 were polled compared to exit polls; and EduDiff, or what percent of people without a bachelor's degree or higher were polled compared to exit polls. Additionally, the percentage of voters that claimed to be undecided has no significant effect on Diff, and neither does PollType, which is how the poll was conducted—by phone, online, or a mix of phone and online.

The DaysOut Predictor

To reiterate findings in chapter 4, the DaysOut predictor was the only of seven predictors in the study found to be statistically significant, which suggests that for each day farther away from Election Day that a poll concludes, the absolute difference between polling margin and election result increases by 0.05%. Or for every 20 days closer to Election Day that a poll concludes, the absolute difference between poll margin and election result decreases by 1%.

Several other sources have found this to be true. A 2023 article from Jain and Lavelle showed a sharp decrease in polling error in presidential elections in the roughly year-and-half leading up to the election date during this century. Polls did not even reach an average error below their own margins of error, a standard of 4%, for any appreciable time until April of the election year, or about 7 months away from the election. However, they conclude that after this point, there is not too much more reduction of error, and, as can be seen in the graphic below, only takes a sharp decline in the last month before the elections. This late reduction is similar to the findings of the current study. Jain and Levelle used polls retrieved from RealClearPolitics, as did the current study. Their analysis is shown in Figure 2.

Figure 2.



Note. Average Polling Error vs Time. Reprinted from *Split Ticket*, by L. Jain and H. Lavelle, Retrieved June 20, 2024, https://split-ticket.org/2023/06/20/how-much-does-early-presidential-polling-matter/. Copyright 2023 by Lakshya Jain and Harrison Lavelle. Reprinted with permission.

Additionally, a 2011 Greg Marx article discussed similar findings in data between 1952 and 2000 by Wleizen and Erikson (2004) that showed a fairly large adjusted-R^2 for predictability of polls based solely on nearness of polls to elections. For example, a fitted R^2 reaches about 0.5 around 160 days before the election and holds relatively steady before waffling upward until it reaches about 0.6 at 100 days before the election. However, in the final 100 days, the adjusted-R^2 grows to near 0.9 on Election Day. In comparing this to the current study and Figure 2 presented above, it seems that there is strong evidence that election proximity has a statistically significant impact on the absolute value of the poll margin minus the election margin.

This is good news for pollsters, as data from the AAPOR's 2020 election report (Clinton et al., 2021) shows that no 5-day period before the period of 20–25 days away from the election has more than 100 polls conducted. And of those final 5-day periods, four of them have more than 100 polls logged. This means that pollsters are saving their resources for when polls should be statistically more accurate.

The EduDiff Predictor

Much has been made of the hidden shy Trump voter in the 2016 election as being a noncollege graduate who would just say they are undecided rather than say they are voting for Trump. Although there is a small amount of evidence to this effect (Wozniak et al., 2019), Campbell (2022) claims that education could at least have some sort of impact, according to the AAPOR. Campbell states that he thinks that Trump supporters were simply not polled at the numbers indicated by the election's turnout, leading to nonresponse bias. This type of bias will be examined later in the chapter.

As far as weighting for education goes, it seems like those who did not do so as much in 2012 started to do it in 2016, and most polling firms weighted for education in 2020. Noble (2021) says that educational weighting was most likely done correctly, as all the error in the 2020 election was one-directional. As was mentioned in chapter 2, Clinton, et al. (2022) extensively studied the weighting of categories as a mechanism for lowering the difference between poll and election margins. They found that weighting proved sufficient in some states, but in others such as Pennsylvania, Wisconsin, and Michigan it lacked the power to overcome the difference in these margins.

What the current study has found, no statistical significance between EduDiff and Diff, would tend to agree with all these analyses. EduDiff had the smallest mean, with the largest standard deviation, among the three demographic predictors. These weighted poll values were less than a fraction of a point away from being equal to exit polls among the 199 polls in the study, indicating that it was weighted almost exactly as it should have been. Mean Diff was several times larger. For EduDiff to actually be significant, a very small change in EduDiff would have to have a large impact on Diff, and it simply did not.

The demographic of education has not been overly analyzed, which was the main reason for its inclusion in this study. Several pollsters have started to use a “White with no college degree” co-demographic in their crosstabs. In the future it could be meaningful to study this as a variable for two reasons. It could be that only White voters without college degrees are underrepresented in polling, and other races are overrepresented. If that is true, then other races, who tend to vote Democrat, might swing the education variable meaningfully into a less accurate place. For example, Cable News Networks’ 2020 exit polling stated that White, noncollege-educated voters preferred Trump to Biden by 35%, while noncollege-educated voters of color preferred Biden to Trump by 46%. Thus, one could compare the racial parts of these demographics to see if weighting the White voters without college degrees more heavily would improve accuracy. The second reason is to see if the education of White voters needs to be weighted versus itself, i.e. noncollege educated White voters are simply nonrespondents more than any other group, specifically college-educated White voters, who tend to be the easiest to reach.

Education remains in a tricky spot, as some pollsters choose not to weight by education at all, others do, and do so by a combination of race and education as one variable or as a covariate. The education weighting might be a symptom of Trump's candidacy and could fade into the background after he leaves politics, or it could continue to be a burden on pollsters for the next several rounds of elections. Research remains ongoing and is needed to pinpoint if education level has a serious impact.

The Undecided Predictor

The undecided predictor garnered quite a bit of interest from the researchers before the study began. Specifically from the 2016 election, did undecided voters break one way, for Donald Trump? Data analysis from chapter 4 suggests that they did not meaningfully do so, as the undecided predictor was not statistically significant in the analysis. Other studies run both counter to and support the results in the current study.

To start, there is exit data from the 2016 election that would point to evidence that undecided voter percentage does matter. Voters who decided on their choice in the month of October chose Trump over Clinton by a 14-point margin, and in the last week of the election, they chose Trump by an 8-point margin (CNN, 2016). In elections decided by less than 50,000 votes in 2016 and 2020 (NBC News, 2020), those margins seem extremely important to the naked eye. However, in 2020, voters who decided in the last few days before the election broke for Biden by 2%, while those deciding in the last month broke for Trump by only 1% (CNN, 2020).

This would make undecided voters seem less important; however, the volume of these undecided voters might matter. The mean percentage of undecided voters in this study was 3.98%. Exit polling data indicates that as many as 16% of voters did not truly decide until the last month before the 2020 election (CNN, 2020). This is a gap of 12% between what voters are telling pollsters and what they are doing in practice. A magnitude of undecided voters this large could make or break an election, especially with elections as close as the past two have been.

However, there has been a study dedicated exclusively to undecided voters. Spencer Kimball's 2020 study of undecided voters suggests that of three ways to split undecided voters, any of them would produce similar results when trying to accurately identify winning candidates. Undecided voters could be allocated evenly, a 50–40 poll with 10% undecided would become 55–45. They could be allocated proportionally, a 50–40 poll with 10% undecided would become 55.5–44.5. Or undecided voters could be asked how they lean and allocated, then remaining undecided voters could be redistributed by either method (Kimball, 2020). This method worked for house, Senate, and governor's polls, all of which have higher absolute error than presidential polls (Enten, 2018). Kimball (2020) additionally posited that adding political party to the polling question would reduce undecided voters. This is because many voters know the party they would choose, even if they were unfamiliar with the candidates.

Undecided voters generally tend to have a higher proportion in reality than they admit to in polling. If the percentage of these voters is large enough, perhaps they could swing an election. Literature tends to agree with the current study, though, that they do not significantly affect the absolute difference between poll margin percentage and the election night margin percentage.

The AgeDiff and RaceDiff Predictors

This study did not find a statistically significant effect of age or race on the Diff variable. To the researchers, this result remains the least surprising. The means of AgeDiff and RaceDiff were small, though not as small as EduDiff, suggesting that pollsters are mostly effective at weighting these variables. On average polls skewed slightly older and whiter than the actual voter when compared to exit polling data.

The reason that these variables have little impact on the absolute difference between poll margin percentage and the election night percentage can be seen as tandem in some cases and in opposition in some cases. For example, age and race have been weighted for decades, and can be done so by simply using voting-age census data. Thus, pollsters have long been able to put good weighting measures into practice, unlike with a newer variable such as education.

Jonge et al. (2018) note regarding the weighting of these variables that age is one of the most minor variables, even though it has been weighted properly. It barely explains any support for Clinton in their 2016 election analysis and ranked lowest among all variables in explaining Trump support. CNN exit polling from 2016 concurs as the net difference from the youngest vote range to the oldest vote range is 27 points—from Clinton plus-19 among those aged 18–29 to Trump plus-8 among those aged 45–64. A -1.40 mean AgeDiff in the current study could barely explain any of the change in Diff across the 199 polls.

On the other hand, race accounts for a much bigger difference. For example, that same CNN exit poll had a net swing in race more than three times as large as age. White voters preferred Trump by 20%, while Black voters preferred Clinton by 81%. However, of all voters in 2016, 71% were White, so, if polls were remotely close to surveying the correct percentage of White people in each state, there would not be as many repercussions in their data. In the current survey, White voters were over-surveyed by 1.25%, but since the difference in vote choice was only 20% among White voters, error was reduced drastically. That is not to say race is unimportant. In fact, Jonge et al. (2018) suggest that race explains the most variance in support between Trump and Clinton, with the cross between race and education explaining the third most.

Exit polling, history, and the 2018 study mentioned above all point to race being an important ingredient. But race and gender have long been on pollsters' radars. Weighting of these variables via the use of census data has been occurring for decades. As Campbell (2022) states, polls are in a tough place, and it is time that they innovate. Luckily for polling firms, age and race are two predictors they will not have to worry about much in the near future.

The PollType Predictor

Possibly one of the fastest-changing variables in this millennium, the PollType predictor did not come back as statistically significant in this survey. Since the ease of using the Internet has increased, and with the widespread use of cell phones, polling mode has changed drastically. In 2000, random-digit-dialing methods were by far the most popular mode (Prosser and Mellon, 2018). But 4 years later, even those older than 30 began to drop a traditional landline from their lives, and by 2008 almost 19% of

American households were landline-free (Mokrzycki et al., 2009; Blumberg and Luke, 2009). Pollsters tended to navigate the change to cellphones well, with the Pew Research Center noticing a skew toward John McCain when comparing landline-only surveys to mixed phone surveys during the 2008 U.S. presidential election (Keeter et al., 2008).

In the current study, which started with the 2012 election, the researchers identified methodologies that included not only landline-exclusive phone polls, but mixed phone polls, cell phone IVR polls, and even mixed-methods approaches using landlines with internet surveys. In 2016, there were even more methods with online-only polls entering the fray, along with IVR landline surveys as well. Mixed-methods approaches also diversified from the 2012 election cycle. How to weight the samples among the various methods has been an undertaking that pollsters have not quite figured out; otherwise, pollsters in 2020 would have all been using one mode. However, sampling methods are quickly changing.

According to a 2023 study by the Pew Research Center, more than three-fifths of public pollsters changed their sampling method between the 2016 election and the 2022 midterm election. Additionally, the incidence of mixed-methods sampling has climbed; polls using three or more sampling methods have risen from 2% in 2016 to 17% in 2022 (Kennedy et al., 2023). This could be a boon to finding the correct people to survey for the sample. As mentioned in chapter 2, Kimble and Holloway’s (2022) research suggests that landline-only polls tend to find older, conservative voters and online-only panels find younger, urban voters, with SMS-to-Web finding a somewhat middle ground. Incorporating these three sampling methods, plus other methods as necessary, could help to alleviate at least some sampling error, or nonresponse bias, which will be discussed shortly.

The mixed-methods sampling approach yielded some interesting results in the current study. The PollMixed predictor marginally missed being a statistically significant variable. If it had been significant, a mixed-methods sampling poll would reduce the absolute difference between poll margin percentage and the election night margin percentage by almost a whole percent compared to a phone-only poll. Because of this near-miss, it might be worth studying the PollType variable to see whether mixed-methods samples are better than online-only panels or phone-only surveys.

Generalizability and Threats to Study at Hand

As with any research that does not, or cannot, study the full population, there will be issues with generalizability. In this study, only states deemed to be competitive in presidential elections were included in the sample, so generalizing to less competitive states would not be reasonable. Additionally, some states had better polling in so much as polls had more easily attainable demographics data or more notable pollsters had done the sampling. Or a more difficult hurdle for generalizability, some states just simply had a higher quantity of polls taken.

For example, polls from North Carolina and Florida accounted for 51 of 199 polls included in this study, or more than a quarter of data points. Also very competitive and important states electorally, Michigan and Georgia accounted for just 23 polls. Some states, namely Minnesota, Maine, and New Mexico, had only about one poll average included per election. Because of Florida's high weight in the study, and geographic and demographic differences, it might be hard to generalize results to Maine.

Another reason it is hard to generalize the results is because of the nature of U.S. presidential elections, which might be more aptly named as threats to the study. In 2012 and 2020 there were incumbents, but not in 2016. Donald Trump's divisive personality was absent from the 2012 election. The coronavirus pandemic did not necessarily impact a person's ability to take a poll, but it might have restricted, or in some cases enhanced, their ability to vote. The polling error in 2012 skewed pro-Republican while the errors in 2016 and 2020 skewed pre-Democrat. Every election has its own unique variables and circumstances, and only some of the predictors measured in this study might have future implications. At this point, it is unlikely that anyone could predict which of them will.

Implications on Polling Aggregation

While the study used two voter theory models, the sociological and the psychosocial, for its theoretical framework, these will be discussed in the next section. A secondary theory used as a basis for this study was discussed in chapter 2 as poll aggregation. This idea generally means that as single polls are combined and aggregated, or modeled, they become quite powerful and more predictive than a single poll (Graefe, 2023).

To put this into perspective, one poll in the study had a Diff of 13, meaning the poll missed the election result by more than three times its margin of error. However, the mean Diff, or the simple aggregation of all polls, lowers that score to 4.53, which was only 0.64 higher than the mean margin of error. Additionally, the study used absolute value to calculate Diff; a study with a signed Diff would have a lower score than the margin of error. This clearly shows agreement with Graefe's statement.

Aggregation of polling models has more predictive power, or in other words, a poll of 1,000 likely voters pales in comparison to an aggregation of a dozen polls that might have a combined sample size of 12,000 likely voters, drastically lowering the margin of error.

Armstrong and Graefe (2021) name several predictors outside of horse-race polling such as betting markets, citizen forecasts, and political experts. These predictors can be combined with aggregators to possibly make even stronger forecasts than a simple average alone (Graefe, 2023). The results from the current study confirm this. The levels seen in Diff for any random single poll will never be more powerful than a mean of all polls and will certainly not be more predictive than a forecast constructed using aggregation and the other predictors that Armstrong and Graefe named.

Implications on Voter Theory and Current Practices

This study began on the framework that two voter theories maintain importance in the current electoral landscape, the sociological and the psychosocial. Shortly, the sociological voter theory pertains to the media and social groups with which a person interacts politically. The psychosocial voter theory explains a vote by the voter's own party affiliation and, in some cases, other lenses such as an economic lens or a current events lens (Lazarsfeld et al., 1968; Antunes, 2010). Though this study cannot comment on whether these inform a voter's ballot, these theories do lead to a current problem in electoral polling that is beyond the scope of this study, the nonresponse bias, or why certain groups might have higher nonresponse. Perhaps there could be a link between mixed-methods sampling and nonresponse bias such that mixed-methods samples reduce this bias either by chance or naturally.

Nonresponse, specifically in political polling, happens when a possible subject cannot be contacted, is contacted and refuses to take a poll, or leaves before the poll has been completed (ScienceDirect, 2024). Nonresponse bias has been a growing threat to polling for the past 25 years, or roughly since the wide use of caller identification. A lower response rate tends to bring the possibility of more error into the study, but it does not necessarily preclude accurate polling. Earlier in this chapter, the necessity of mixed-methods sampling in the polling world was briefly discussed. Data from the Pew Research Center supports this, as response rates in telephone surveys have fallen from 36% in 1997 to 6% in 2018 (Kennedy and Hartig, 2019). During data collection for the current survey, the researchers observed several published polls with a response rate below 2%. While this alone does not impair the poll's accuracy, it does allow for nonresponse bias.

Historically when nonresponse bias prohibited accuracy less often, weighting solved most of the pollsters' problems. If a pollster did not have quite enough Black voters or young voters, they would simply apply statistical techniques, including but not limited to simple proportions, to match the demographics of their polls to fit the demographics of the likely voter populations. However, demographic weighting has grown weaker with the rise of nonresponse (Mercer et al., 2018). Several studies (Gelman, 2021; Jackson et al., 2020; Clinton et al., 2022) have parroted this information and gone one step further, calling for something more than weighting. Clinton et al. (2022) achieved success in doing this to improve some, but not all, results.

A few demographic groups will be so hard to reach in any one way, whether by phone, text message, email, online, or other means, that they will be roughly excluded from the survey. Or if there is a small sample of that group, the weighting would have to be such that it might misrepresent the actual demographic group. Hypothetically, a landline-only phone survey would likely have a major problem reaching those under age 39. And while mixed-methods sampling can help this, nonresponse bias can rear its head in other ways. Jackson, et al. (2020) mention that this is a growing problem, and several other studies presented in chapter 2 agree that is either a piece of the puzzle or the full reason why polls missed so badly in 2016 and 2020 (Durand and Johnson, 2021; Barnett and Sarfati, 2023).

It has been noted several times during this study that 2016 and 2020 polling missed toward the left, giving Democrats a larger lead in polls than each election's outcome. Nonresponse bias could explain this. According to AAPOR's 2020 post-mortem election report, a disproportionate amount of White, noncollege-educated voters were left out of polls, the exact voter base that led Trump to victory (Clinton et al., 2021). A 2024 Pew Research Center report found that in the upcoming 2024 presidential election, there is a 20-point gap in favor of Biden for White, college-educated voters as compared to White noncollege-educated voters. For people of color, the gap was 5 points or less using the same comparison.

Other surveys have results that coincide with this. In 2022, 538 sponsored a tracking poll with Ipsos where the same voters were contacted monthly for 5 months. After 2 months, Republicans responded the least to recontact, and they remained lower than Democrats and independents throughout the whole survey. Additionally, the study found that it was the farthest-right Trump supporters and those who received their news mostly from social media who ghosted the poll most often, exactly the group who would be young, disenchanted, low-propensity voters (Feldman and Mendez, 2022).

The tracking poll ties into both the sociological and psychosocial voter theories. Why does this group not cooperate with polling, or are they just generally hard to reach? What can be done to combat this problem? Several pollsters have tried to reach out to these low-propensity voters, and results might be beginning to show.

In an interview with Axios (Saric, 2024), Pew Research Center's Courtney Kennedy said that she believes that because 2024 polls have shifted more toward Donald Trump in a rematch of two well-known candidates from 2020, maybe some of these issues have been fixed. Recent *New York Times* and Siena College surveys found that these low-propensity, registered voters might finally be showing up in polls. In their polling, voters from the 2020 election preferred Biden by 2%, while nonvoters from the 2020 election preferred Trump by 14% (Cohn, 2024). These are the kinds of voters that have been getting lost during the past two presidential elections. Now, has there been a concrete correction for this bias?

Research remains in its infancy, though a 2023 study from West and Andridge suggests that a new technique could detect nonresponse bias and correct for it. They do this by using a measure of unadjusted bias for a proportion. Using covariates of each person in the poll, they fit a regression model to determine a binary candidate selection. Other statistical techniques follow, boiling data down to a specific linear model based on the relationship between these sets of covariates. This technique was applied to nine polls in eight states from the 2020 U.S. presidential election, and it performed at the same level or better than weighting in every case, specifically doing well when more than just simple demographics of a person were available (West and Andridge, 2023).

If pollsters begin to take nonresponse bias very seriously and use new techniques such as those mentioned above, chances are polling could see a rise in accuracy. However, if these hard-to-reach voters start to show up because of redesigns to polling methods, it will be at a cost: three presidential elections of massive polling misses, pollsters going out of business, and eroded trust from the public.

Conclusion

At the outset of the study, the researchers wanted to know if specific predictors could influence the absolute difference between a poll's margin and the actual election margin. The answer is, put simply, some of them. The results suggest that there exists a strong tie between the proximity of the election and the accuracy of polling, and myriad other studies also assert this. Other variables in the study showed no statistically significant effect on the difference being studied; however, PollType became an interesting case.

When examining the PollType variable using all 205 cases, a nearly significant connection between increased accuracy and mixed-methods sampling arose when compared to phone-only or online-only samples. Because mixed-methods samples can reach more demographic groups in society purely by the nature of the medium, it led to a discourse about nonresponse bias—or why some people are systematically being left out of the polling process. Several studies indicate that more than any possible predictor, the way polling firms handle nonresponse bias versus weighting demographics could be the key to accuracy in modern polling. But, as AAPOR Task Force Chair Josh Clinton said, learning about people who might actively want to disengage or avoid the process is difficult (Keeter, 2021).

Some say polling has always been flawed, and though polls will be around for a long time, they need to change to avoid an early death (Prosser and Mellon, 2018). Others say that polling is alive and well and point to evidence that some pollsters did extremely well in the 2020 presidential elections (Silver, 2021). Even more confirm that sentiment with data from the 2022 midterms showing the most accurate polls since 1998 (Rakich, 2023).

After examining the heart of the problem, it is important to remember the stakes. The public wants as much information as possible, and it needs to be as accurate as possible. News media put their credibility on the line when they report polls and analyze them around the clock, news cycle after news cycle. Campaigns spend millions or even billions of dollars, use human capital, budget advertising time, and plan campaign trail stops based on a poll saying that a state and its electoral votes can be won.

Political experts say that polling needs to be better, and it does. It is highly improbable that a polling firm would put its livelihood at risk by using outdated methods. This study found some useful connections between variables. One of them, election proximity, is straightforward. The other, poll mode, and by proxy nonresponse bias, is in the infancy of its exploration. But by no means should it be assumed that polling is dead; it is evolving. In a 2021 editorial in *The Washington Post*, David Byler sums up polling and its current issues in an almost humane manner, relating the difficulties of the work with the everchanging landscape of American politics and media:

> Survey research is hard. Pollsters have to contact people who do not want to talk to them, talk about extremely personal, high-stakes issues and make a complex series of statistical decisions before presenting the public with something as unlikely as a consensus set of opinions. No person can perfectly navigate these obstacles all the time.

REFERENCES

Ambrosius, J. D. (2016). Blue City ... Red City? A Comparison of Competing Theories of Core County Outcomes in U.S. Presidential Elections, 2000-2012. *Journal of Urban Affairs*, *38*(2), 169–195. doi:10.1111/juaf.12184

Antunes, Rui. (2010). Theoretical Models of Voting Behaviour. *Exedra*, *4*. 145–170.

Armstrong, J.S., & Graefe, A. (2021). The PollyVote Popular Vote Forecast for the 2020 US Presidential Election. *PS: Political Science & Politics*, *54*(1), 96–98. doi:10.1017/S1049096520001420

Barnett, A., & Sarfati, A. (2023). The Polls and the U.S. Presidential Election in 2020 …. and 2024. *Statistics & Public Policy*, *10*(1), 1–11. doi:10.1080/2330443X.2023.2199809

Blumberg, S.J., & Luke, J.V. (2009). *Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July–December 2008.* National Center for Health Statistics. https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200905.pdf

Bracic, A., Israel-Trummel, M., & Shortle, A.F. (2019). Is Sexism for White People? Gender Stereotypes, Race, and the 2016 Presidential Election. *Political Behavior*, *41*(2), 281–307. doi:10.1007/s11109-018-9446-8

Buyuker, B., D'Urso, A.J., Filindra, A., & Kaplan, N.J. (2021). Race politics research and the American presidency: thinking about white attitudes, identities and vote choice in the Trump era and beyond. *Journal of Race, Ethnicity & Politics*, 6(3), 600–641. doi: 10.1017/rep.2020.33

Byler, D. (2021, July 22). Polling is broken. No one knows how to fix it. *Washington Post*. Retrieved June 25, 2024, from, https://www.washingtonpost.com/opinions/2021/07/22/polling-is-broken-no-one-knows-how-fix-it/

Campbell, W.J. (2021). About Those Dismal Pre-election Polls: Yes, They were Predictions. *Sociological Forum*, *36*(2), 560–561. doi:10.1111/socf.12708

Campbell, W.J. (2022). Misfires and Surprises: Polling Embarrassments in Recent U.S. Presidential Elections. *American Behavioral Scientist*, *0*(0). doi:10.1177/00027642221118901

Clinton, J., Agiesta, J., Brenan, M., Burge, C., Connelly, M., Edwards-Levy, A., Fraga, B., Guskin, E., Sunshine Hillygus, D., Jackson, C., Jones, J., Keeter, S., Khanna, K., Lapinski, J., Saad, L., Shaw, D., Smith, A., Wilson, D., & Wlezien, C. (2021). Task force of 2020 pre-election polling: An evaluation of the 2020 general election polls (Report of the American Association for Public Opinion Research). https://aapor.org/wp-content/uploads/2022/11/AAPOR-Task-Force-on-2020-Pre-Election-Polling_Report-FNL.pdf

Clinton, J.D., Lapinski, J.S., & Trussler, M.J. (2022). Reluctant Republicans, Eager Democrats? Partisan Nonresponse and the Accuracy of 2020 Presidential Pre-election Telephone Polls. *Public Opinion Quarterly*, *86*(2), 247–269. doi:10.1093/poq/nfac011

Cox, D. (2020, November 24). *Could Social Alienation Among Some Trump Supporters Help Explain Why Polls Underestimated Trump Again?* FiveThirtyEight. https://fivethirtyeight.com/features/could-social-alienation-among-some-trump-supporters-help-explain-why-polls-underestimated-trump-again/

Cohn, N. (2024, May 24). A Polling Risk for Trump. *New York Times*. Retrieved June 25, 2024, from, https://www.nytimes.com/2024/05/24/briefing/a-polling-risk-for-trump.html

Dawson, S. (2023). Perceptions of opinion poll credibility: The role of partisan bias. *Party Politics*, *29*(3), 594–599. doi:10.1177/13540688221098837

Durand, C., & Johnson, T.P. (2021). Review: What about Modes? Differences between Modes in the 21st Century's Electoral Polls across Four Countries. *Public Opinion Quarterly*, *85*(1), 183–222. doi:10.1093/poq/nfab006

Enten, H. (2018, November 19). *2018 Was a Very Good Year for Polls* | *CNN Politics*. CNN.
https://www.cnn.com/2018/11/19/politics/2018-midterm-elections-good-year-polls/index.html

Exit Polls 2016. (2024, June 21). CNN. https://www.cnn.com/election/2016/results/exit-polls

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. doi:10.3758/BRM.41.4.1149

Feldman, S., & Mendez, B. (2022, October 26). *Who Are The People Who Don't Respond To Polls?* FiveThirtyEight. https://fivethirtyeight.com/features/nonresponse-bias-ipsos-poll-findings/

Field, A., Miles, J., & Field, Z. (2013) *Discovering statistics using R*. Los Angeles: Sage.

Florida | 2016 Election Forecast | FiveThirtyEight. (2024, June 10). FiveThirtyEight. https://projects.fivethirtyeight.com/2016-election-forecast/florida/

Gelman, A. (2021). Failure and Success in Political Polling and Election Forecasting. *Statistics and Public Policy, 8*(1), 67–72. doi:10.1080/2330443X.2021.1971126

Gelman, A., Goel, S., Rivers, D., & Rothschild, D. (2016). The Mythical Swing Voter. *Quarterly Journal of Political Science*, *11*(1), 103–130. doi:10.1561/100.00015031

Graefe, A. (2018). Predicting elections: Experts, polls, and fundamentals. *Judgment & Decision Making*, *13*(4), 334–344. doi:10.1017/s1930297500009219

Graefe, A. (2023). Embrace the differences: Revisiting the PollyVote method of combining forecasts for U.S. Presidential elections (2004 to 2020). *International Journal of Forecasting*, *39*(1), 170–177. Doi:10.1016/j.ijforecast.2021.09.010

Herding. (2023, November 11). American Association for Public Opinion Research. https://www-archive.aapor.org/Education-Resources/Election-Polling-Resources/Herding.aspx/

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

Jackson, N., Lewis-Beck, M.S., & Tien, C. (2020). Pollster problems in the 2016 US presidential election: vote intention, vote prediction. *Italian Journal of Electoral Studies*, *83*(1), 17–28. doi:10.36253/qoe-9529

Jain, L., & Lavelle, H. (2023, June 20). *How Much Does Early Presidential Polling Matter? – Split Ticket*. Split Ticket. https://split-ticket.org/2023/06/20/how-much-does-early-presidential-polling-matter/

Jonge, C.P.K. de, Langer, G., & Sinozich, S. (2018). Predicting State Presidential Election Results Using National Tracking Polls and Multilevel Regression with Poststratification (MRP). *Public Opinion Quarterly*, *82*(3), 419–446. doi:10.1093/poq/nfy023

Keeter, S. (2021, July 21) *Q&A: A conversation about U.S. election polling problems in 2020*. Pew Research Center. https://www.pewresearch.org/short-reads/2021/07/21/a-conversation-about-u-s-election-polling-problems-in-2020/

Keeter, S., Dimock, M., & Christian, L. (2008). *Cell Phones and the 2008 Vote: An Update*. Pew Research Center. https://web.archive.org/web/20081001212101/http://pewresearch.org/pubs/964/

Keeter, S., Hatley, N., Kennedy, C., & Lau, A. (2017, May 25). What Low Response Rates Mean for Telephone Surveys. https://www.pewresearch.org/methods/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/

Kennedy, C. (2020, August 5). *Key things to know about election Polling in the United States*. Pew Research Center. https://www.pewresearch.org/short-reads/2020/08/05/key-things-to-know-about-election-polling-in-the-united-states/

Kennedy, C., & Hartig, H. (2019, February 27). *Response rates in telephone surveys have resumed their decline*. Pew Research Center. https://www.pewresearch.org/short-reads/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/

Kennedy, C., Popky, D., & Keeter, S. (2023, April 19). *How Public Polling Has Changed in the 21st Century.* Pew Research Center. https://www.pewresearch.org/methods/2023/04/19/how-public-polling-has-changed-in-the-21st-century/

Kimball, S. (2020). Allocating Undecided Voters in Pre-election Polling. *Tripodos*, 48, 69–84. doi:10.51698/tripodos.2020.48p69-84

Kimball, S., & Holloway, I. (2022). In the Mode. . .Text-to-Web Survey Data Collection: An Exploratory Study in Preelection Polling of the U.S. Presidential Election. *American Behavioral Scientist*, *0*(0). doi:10.1177/00027642221132801

Kuru, O., Pasek J., & Traugott M.W. (2020). When Polls Disagree: How Competitive Results and Methodological Quality Shape Partisan Perceptions of Polls and Electoral Predictions. *International Journal of Public Opinion Research*, *32*(3), 586–603. doi:10.1093/ijpor/edz035

Lazarsfeld, P.F., Berelson, B. & Gaudet, H. (1944). *The People's Choice: How the Voter Makes up His Mind in a Presidential Campaign*. New York: Columbia University Press.

Lazarsfeld, P., Berelson, B., & Gaudet, H. (1968). *The People's Choice: How the Voter Makes up His Mind in a Presidential Campaign* (3rd ed.). New York: Columbia University Press.

Lewis-Beck, M.S., & Tien, C. (1999). Voters as forecasters: a micromodel of election prediction. *International Journal of Forecasting*, *15*(2), 175–184. doi:10.1016/S0169-2070(98)00063-6

Mokrzycki, M., Keeter, S., & Kennedy, C. (2009). Cell-Phone-Only Voters in the 2008 Exit Poll and Implications for Future Noncoverage Bias. *Public Opinion Quarterly*, *73*(5), 845–865. http://doi.org/10.1093/poq/nfp081

Larrick, R.P., & Soll, J.B., (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Managing Science*, *52*(1), 111–127. doi:10.1287/mnsc.1060.0518

Marx, G. (2011, May 26). *More on Early Polls – Columbia Journalism Review*. Columbia Journalism Review. https://www.cjr.org/the_kicker/more_on_early_polls.php/

Mercer, A., Lau, A., & Kennedy, C. (2018, January 26). *1. How different weighting methods work*. Pew Research Center. https://www.pewresearch.org/methods/2018/01/26/how-different-weighting-methods-work/

Mohammed, S.N., & Trumpbour, R.C. (2021). Polls and Elections: “The Carnivalesque in the 2016 US Presidential Campaign.” *Presidential Studies Quarterly*, *51*(4), 884–903. doi:10.1111/psq.12658

Murr, A. (2016). The wisdom of crowds: What do citizen forecast for the 2015 British General Election. *Electoral Studies*, *41*, 283–288. doi:j.electstud.2015.11.018

National Results 2020 President Election Polls. (2024, June 21). CNN. https://www.cnn.com/election/2020/exit-polls/president/national-results

NBC News. (n.d.). Pennsylvania Results. NBC News. Retrieved November 20, 2023, from https://www.nbcnews.com/politics/2016-election/pa/

NBC News. (2020, November 3). Wisconsin Election Results 2020. NBC News. Retrieved November 20, 2023, from https://www.nbcnews.com/politics/2020-elections/wisconsin-results/

Noble, I. (2021). A Question of Trust: How and Why the Polls Underestimated Support for Donald Trump. *Sociological Forum*, *36*(2), 559–560.

Nonresponse Bias – an overview | ScienceDirect Topics. (2024, June 24). ScienceDirect. https://www.sciencedirect.com/topics/nursing-and-health-professions/nonresponse-bias

Pennsylvania Results 2016 – NBC News. (2024, June 21). NBC News. https://www.nbcnews.com/politics/2016-election/pa/

Opinion polling for the 2016 United States presidential election in Florida. (2024, June 10). Wikipedia. https://en.wikipedia.org/wiki/Opinion_polling_for_the_2016_United_States_presidential_election_in_Florida

Pew Research Center (2024, April 24). *Report: Support for Biden, Trump in 2024 rematch | Pew Research Center*. https://www.pewresearch.org/politics/2024/04/24/the-biden-trump-rematch/

Price, V., & Stroud, N.J. (2006). Public attitudes toward polls: Evidence from the 2000 U.S. Presidential Election. *International Journal of Public Opinion Research*, *18*(4), 393–421. doi:10.1093/ijpor/edh119

Prosser, C., & Mellon, J. (2018). The Twilight of Polls? A Review of Trends in Polling Accuracy and the Causes of Polling Misses. *Government and Opposition*, *53*(4), 757–790. doi:10.1017/gov.2018.7

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rakich, N. (2023, March 10). *The Polls Were Historically Accurate in 2022*. FiveThirtyEight. https://fivethirtyeight.com/features/2022-election-polling-accuracy/

RealClearPolitics - 2012 Election Maps - Battle for White House. (2024, May 18). RealClearPolitics. https://www.realclearpolitics.com/epolls/2012/president/2012_elections_electoral_college_map.html

RealClearPolitics - 2016 Election Maps - Battle for White House. (2024, May 18). RealClearPolitics. https://www.realclearpolitics.com/epolls/2016/president/2016_elections_electoral_college_map.html

RealClearPolitics - 2020 Election Maps – 2020 Electoral College Map. (2024, May 18). RealClearPolitics. https://www.realclearpolitics.com/epolls/2020/president/2020_elections_electoral_college_map.html

Rothschild, D. (2009). Forecasting Elections: Comparing Prediction Markets, Polls, and Their Biases. *Public Opinion Quarterly*, *73*(5), 895–916. doi:nfp082

Saric, I. (2024, June 15). *How 2024 pollsters are trying to avoid their 2020 mistakes*. Axios. https://www.axios.com/2024/06/15/2024-election-polls-trump-biden

Searles, K., Smith, G., & Sui, M. (2018). Partisan Media, Electoral Predictions and Wishful Thinking. *Public Opinion Quarterly*, *82*(Suppl.), 302–324. doi:10.1093/poq/nfy006

Silver, N. (2021, March 25). *The Death Of Polling Is Greatly Exaggerated*. FiveThirtyEight. https://fivethirtyeight.com/features/the-death-of-polling-is-greatly-exaggerated/

Smith-Schoenwalder, C. (2020, October 28). *Trump Attacks 'Suppression Polls' That Show Biden Ahead*. U.S. News & World Reports. https://www.usnews.com/news/elections/articles/2020-10-28/trump-attacks-suppression-polls-that-show-biden-ahead

Soll, J.B., & Larrick, R.P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 780–805. doi:10.1037/a0015145

Statewide opinion polling for the 2012 United States presidential election. (2024, May 24). Wikipedia. https://en.wikipedia.org/wiki/Statewide_opinion_polling_for_the_2012_United_States_presidential_election

Statewide opinion polling for the 2016 United States presidential election. (2024, May 24). Wikipedia. https://en.wikipedia.org/wiki/Statewide_opinion_polling_for_the_2016_United_States_presidential_election

Statewide opinion polling for the 2020 United States presidential election. (2024, May 24). Wikipedia. https://en.wikipedia.org/wiki/Statewide_opinion_polling_for_the_2020_United_States_presidential_election

Sturgis, P., Baker, N., Callegaro, M., Fisher, S., Green, J., Jennings, W., Kuha, J., Lauderdale, B., & Smith, P. (2016). *Report of the Inquiry into the 2015 British general election opinion polls*. Market Research Society and British Polling Council, London. https://eprints.ncrm.ac.uk/id/eprint/3789/1/Report_final_revised.pdf

West, B.T., & Andridge, R.R. (2023). Evaluating Pre-election Polling Estimates Using a New Measure of Non-Ignorable Selection Bias. *Public Opinion Quarterly*, *87*(Suppl.), 575–601. doi:10.1093/poq/nfad018

Westwood, S. J., Messing, S., & Lelkes, Y. (2020). Projecting Confidence: How the Probabilistic Horse Race Confuses and Demobilizes the Public. *Journal of Politics*, *82*(4), 1530–1544. doi:10.1086/708682

Wisconsin election results: Live results by county. (2024, June 21). NBC News. https://www.nbcnews.com/politics/2020-elections/wisconsin-results/

Wleizen, C., & Erikson, R.S. (2004, September 2). *The Fundamentals, the Polls, and the Presidential Vote* [Paper presentation]. Annual Meeting of the American Political Science Association, Chicago, IL, United States. https://web.archive.org/web/20110913184441/http://campus.murraystate.edu/academic/faculty/mark.wattier/wlezienerikson2004.pdf

Wozniak, K.H., Calfano, B.R., & Drakulich, K.M. (2019). A “Ferguson Effect” on 2016 Presidential Vote Preference? Findings from a Framing Experiment Examining “Shy Voters” and Cues Related to Policing and Social Unrest. *Social Science Quarterly*, *100*(4), 1023–1038. doi:10.1111/ssqu.12622

APPENDIX A

Example of a poll with predictors and predictor abbreviations

Diff	MoE	Undecided	DaysOut	PollType	State	AgeDiff	RaceDiff	EduDiff
5	4.6	8	23	Phone	MI	-1	-6	-6

Diff: Absolute difference of poll margin (%) and state vote margin (%)

MoE: Poll Margin of Error

Undecided: Percent of voters who declared themselves as undecided in that poll

DaysOut: How many days before Election Day that the poll concluded

PollType(categorical): Three categories consisting of Phone, which includes Interactive Voice Recognition (IVR); Online, which includes SMS-to-Web; and Mixed Methods (Mixed)

State(categorical): State postal code abbreviation, used for tracking purposes

AgeDiff: Exit polling data versus actual poll data, percentage polled 18–44 minus exit polls 18–44; a positive value indicates more younger people were polled than what exit polls indicated

RaceDiff: Exit polling data versus actual poll data, percentage polled White minus exit polls White; a positive value indicates more White people were polled than what exit polls indicated

EduDiff: Exit polling data versus actual poll data, percentage polled no college degree minus exit polls no college degree; a positive value indicates more people without a college were polled than what exit polls indicate

BIBLIOGRAPHY

Benjamin Andrew Schneider

Candidate for the Degree of

Master of Science Mathematics

Thesis: The Effect of Age, Race, Education, Margin of Error, Undecided Voters, Poll Type, and Election Proximity on Poll Margin Accuracy

Major Field: Mathematics

Education: Ohio University, B.S. Journalism, 2007; Ohio University, M.Ed. AYA Mathematics Education, 2010

Completed the requirements for the Master of Science in Mathematics, Portsmouth, Ohio, in July 2024.

7/22/2024

ADVISER'S APPROVAL: Dr. Douglas Darbro