

AN ASSESSMENT OF THE RELIABILITY OF POLLING DATA AHEAD OF FUTURE PRESIDENTIAL ELECTIONS IN THE UNITED STATES

Mohammed Kabeer Garba (Ph. D)

ECOWAS Parliament, Abuja

Email: kabirmashi@gmail.com; +234 803 659 0281

DOI:<https://doi.org/10.5281/zenodo.15730612>

Abstract: The accuracy of polling data has been a subject of intense debate in recent years, particularly in the wake of high-profile misses in the 2016 and 2024 United States of America's presidential elections despite polls indicating a high likelihood of a Hilary Clinton and Kamala Harris victory, Trump won in both elections. This study provides an assessment of the reliability of polling data ahead of presidential election polls in the United States of America. Contemporary polling faces numerous challenges, including declining response rates, the rise of non-probability internet samples, and increasing concerns about survey error. This study employs a mixed-methods approach, combining quantitative analysis of polling data from the 2016 presidential election to the 2024 presidential election with qualitative insights from expert interviews as both elections provide opportunities to examine the reliability of polling data in a more contemporary context. The quantitative analysis assesses the accuracy of polling data in predicting election outcomes, while the qualitative component explores the challenges and limitations of polling in the contemporary media landscape. The study adopted structural functional theory as developed by sociologists such as Émile Durkheim (1893) and Talcott Parsons (1949), emphasizes the interconnectedness of social structures and institutions in maintaining social order. The study concludes that there is the need for caution when interpreting polling data and underscores the importance of continued research into the challenges facing contemporary polling. The study recommends that the public should be cautious when interpreting polling data and recognize the potential limitations and biases of polling estimates.

Key words: Polling data. Polling reliability. Presidential elections. United States of America

Introduction

The history of political polling dates back to the early 19th century, but it became more scientific with advent of George Gallup's techniques in the 1930s. This technique involved random sampling, significantly improved the accuracy of polls. Over the decades, polling methodologies have evolved, integrating phone surveys, online polling, and now even social media sentiments analysis, (NIWA Report, 2024). The reliability of polling data has been a subject of interest in the United States of America (USA), predominantly in the context of presidential elections. The history of polling in America dates back to the 19th century, with the first presidential poll conducted in 1824. However, it wasn't until the 20th century that polling became essential in American politics.

The introduction of scientific polling methods in the 1930s transformed the field, with pioneers like George Gallup and Elmo Roper developing methodologies that are still used today, (Gallup, 1972). Despite these developments, polling has faced numerous challenges, including criticisms of bias, methodology, and accuracy. In recent years, the reliability of polling data has come under scrutiny, particularly in the wake of the 2016 presidential election. Many polls failed to predict the outcome of the election, leading to widespread criticism of the polling industry, (Silver, 2016). This has sparked a renewed interest in assessing the reliability of polling data, particularly in the context of future presidential elections. This study aims to contribute to this discussion by examining the historical background of polling in America, the methodologies used, and the challenges faced by the industry. The study will also assess the reliability of polling data ahead of future presidential elections, with a focus on identifying best practices and areas for improvement.

Throughout the 19th and 20th centuries, the United States experienced significant economic growth, technological advancements, and social change. The country played a major role in both World War I and World War II, emerging as a global superpower in the latter half of the 20th century. The United States operates under a federal system of government, with a president serving as head of state and government. The country has a mixed economy, with a combination of private enterprise and government regulation. The United States is a global frontrunner in terms of economic output, technological innovation, and cultural influence. However, the country also faces significant challenges, including income inequality, racial and ethnic disparities, and environmental degradation, (Kish, 1965).

The assessment of the reliability of polling data ahead of future presidential elections in the United States of America is a crucial aspect of understanding the democratic process. Polling data has been a cornerstone of election forecasting, providing valuable insights into public opinion and voter behavior. However, the accuracy and reliability of polling data have been questioned in recent years, particularly in the wake of the 2016 presidential election. Historically, polling data has played a significant role in shaping public opinion and informing election strategies. The first scientific polls were conducted in the 1930s, with the establishment of the Gallup Organization. Since then, polling has become an integral part of the electoral process, with numerous organizations and media outlets conducting polls to gauge public opinion, (Silver, 2016). The 2016 and 2024 presidential elections highlighted the limitations of polling data, with many polls predicting a Hillary Clinton and Kamala Harris victory. The unexpected outcome of the election raised questions about the accuracy of polling data and the methodologies used to collect it.

Theoretical Framework

Assessing the reliability of polling data ahead of future presidential elections in the United States of America can be understood through the lens of structural functional theory. This theoretical framework, developed by sociologists such as Émile Durkheim (1893) and Talcott Parsons (1949), emphasizes the interconnectedness of social structures and institutions in maintaining social order. In the context of polling data, structural functional theory suggests that the reliability of polls is influenced by the social structures and institutions that govern the

polling process. For instance, the media plays a crucial role in disseminating polling data to the public, while political institutions, such as parties and interest groups, can influence the polling process through their campaigns and advertising efforts.

Structural functional theory postulates that social stability is maintained through the functioning of social institutions and structures. In the context of polling data, social stability refers to the accuracy and reliability of polls in reflecting public opinion. Secondly is concept of collective functioning which refers to the synergy between pollsters, media outlets, and political institutions to produce accurate and reliable polling data. Thirdly is social evolution. Structural functional theory suggests that social change occurs through the adaptation of social structures and institutions to changing social conditions. In the context of polling data, social evolution refers to the evolution of polling methods and techniques in response to changing social and technological conditions. (Smith, 1997)

Literature review

The electoral system in the United States is a function of the country's history, constitution, and political culture. At its core, the system is designed to elect the President and Vice President of the United States through an indirect process, involving the Electoral College. The US electoral system is a unique and complex process that has been in place since the Country's founding. (GeeksforGeeks, 2023) The Electoral College is a group of 538 electors chosen by each state to cast votes for President and Vice President. The number of electors is equal to the number of members each state has in the House of Representatives plus the number of Senators. The District of Columbia also has three electors, thanks to the 23rd Amendment. (Lipset, 1996)

The composition of the Electoral College is quite complex. In the first stage, each state appoints its electors, usually through a popular vote. It consists of 538 electors, chosen by each state to cast votes for the President and Vice President. Each state appoints its electors using legal procedures determined by its legislature. Key features that determine victory is a winner-takes-all system. The second stage sees that the electors meet in their respective state capitals to cast their votes for President and Vice President. The third stage is that the candidate who receives the majority of the electoral votes (at least 270 out of 538) wins the presidency. Fourth and final stage is that if no candidate receives a majority, the House of Representatives chooses the President, and the Senate chooses the Vice President. (Groves, & Couper, 1998)

The US electoral system has faced criticisms and controversies over the years. Some argue that the system is obsolete and can lead to a candidate winning the presidency without receiving the most popular votes. Others argue that the system gives too much power to swing states and can lead to electoral deadlocks, (Lipset, 1996). In most states, the candidate who wins the most votes in that state gets all of its electoral votes, except in Maine and Nebraska which allocate their electoral votes proportionally. Although electors are pledged to vote for a specific candidate, some may choose to vote differently, potentially affecting the outcome of the election. The Electoral College system can lead to a circumstance where the winner of the popular vote does not win the presidency, as happened in 1824, 1876, 1888, 2000, and 2016. (Silver, 2016)

While the Electoral College system has been in place for over two centuries, it has been criticized for several reasons. Apart from creating a situation where the winner of the popular vote does not win the presidency, critics also argue that the Electoral College system is outdated, undemocratic. Additionally, the system has been criticized for favoring swing states and disenfranchising voters in states that are not considered competitive. The US electoral system has undergone significant changes since its inception. (Cochran, 1977) The 12th Amendment, ratified in 1804, changed the procedure for electing the President and Vice President. The amendment requires electors to cast separate ballots for President and the Vice President, rather than casting two votes for President. There have been several efforts to reform the Electoral College system over the years. One of the most notable efforts is the National Popular Vote Interstate Compact (NPVIC), which aims to ensure that the winner of the popular vote wins the presidency. The compact would require states to award their electoral votes to the winner of the national popular vote, rather than the winner of the state's popular vote. Although the system faces criticisms and controversies, it still remains an essential part of American democracy. (Groves et al. 2024)

Assessing the reliability of polling data is crucial in understanding its role in predicting presidential election outcomes in the United States. Polls typically rely on random sampling techniques to select respondents, but these methods can be flawed hereby threatening the validity and reliability of electoral outcomes. It occurs when a sample is selected in a way that systematically distorts the representation of the population, leading to inaccurate conclusion. (Groves et al. 2024) Consequences of sampling bias are inaccurate conclusions, lack of generalizability and waste of resources, (Lynn & Clarke, 2002).

Another challenge facing polling data is the issue of non-response bias. Many polls experience low response rates, which can lead to biased samples. For example, polls may over represent individuals who are more engaged in politics or have stronger opinions on certain issues. Respondents may not be interested in the survey topics or may not see the relevance of the survey, (Goyder, 1987). Another reason for non-response is when the survey is lengthy in time, content and complexity. This can lead to respondents fatigue and decreased response rates. Different models of data collection, for example, online, phone, mail, can affect responses rates and lead to bias. (Groves, & Couper, 1998)

Survey mode refers to the method used to collect data from respondents, such as online, phone, mail, or in-person interviews. The choice of survey mode can significantly impact data quality, response rates, and the overall success of a survey, (Dillman, 2000). This discussion will delve into the concept of survey mode, its types, advantages, and disadvantages, as well as strategies for selecting the most appropriate mode. Types of survey modes include, online surveys which are web-based surveys that use email invitations or online panels to collect data, (Couper, 2000). Another type of survey mode is phone surveys relying on massive telephone-based data from direct discussions with respondents using random digit dialing (RDD) or listed numbers to collect data, (Lavrakas, 1993). Also, mail surveys which are paper-based surveys that use postal mail to collect data, (Dillman, 2000). Finally is the in-person surveys. This is a physical Face-to-face interviews that use intercept or door-to-door methods to collect data, (Sudman, 1980). The choice of survey mode can significantly impact data quality,

response rates, and the overall success of a survey. By considering the target population, survey content, budget, and resources, as well as data quality requirements, researchers can select the most appropriate survey mode for their study.

Another challenge of data reliability is Question wording and design. These are crucial aspects of survey research, as they can meaningfully impact the quality and correctness of responses, (Schuman & Presser, 1981). The way questions are phrased and designed can influence respondents' answers, leading to biased or inaccurate data, (Tourangeau et al., 2000). This discussion will delve into the concept of question wording and design, its impact on survey responses, and strategies for optimizing question design. Components of question wording and design such as leading questions that suggest a particular answer or contain emotionally charged language can lead to biased responses, (Schuman & Presser, 1981). Also double-barreled questions that ask two or more things at once can confuse respondents and lead to inaccurate answers (Converse & Presser, 1986). Finally, negative wording can lead to different responses than those that use positive wording, (Tourangeau et al., 2000). Therefore, question wording and design are critical aspects of survey research, with the potential to significantly impact the quality and accuracy of responses. By understanding the impact of question wording and design on survey responses and employing strategies for optimizing question design, researchers can increase the validity and reliability of their findings, (GeeksforGeeks. 2023).

External issue is also a challenge of assessment of reliability of polling data ahead of future US presidential elections. Political climate can influence polling data, as respondents may be more or less likely to partake or answer truthfully based on their political beliefs. Secondly is social and cultural trends. Shifts in social and cultural attitudes can affect polling data, particularly if certain groups are more or less likely to participate or respond honestly. Economic factors, such as employment rates and income levels, can influence polling data, as respondents' financial situations may impact their opinions and behaviors. Technological advancements is also a major element of external influences. (Fowler, 2013) The rise of online polling and social media has presented new challenges and opportunities for polling data, including concerns about representation and bias.

Weighting and adjustments are crucial components in ensuring the reliability of polling data, particularly in the context of US presidential elections. The primary goal of weighting is to account for the differences between the sample demographics and the actual population, thereby preventing biases and inaccuracies in the polling results. There are various types of weighting methods, including demographic weighting. This involves adjusting the sample data to match the demographic characteristics of the population, such as age, gender, income, and education levels. Another type of weighting method is probability weighting. This method is used when the probability of selection for each respondent is not equal, and it adjusts for these differences to ensure that each individual's response is represented in proportion to their likelihood of being chosen. Finally is post-stratification weighting method. This involves adjusting the sample data to match the population in terms of various groups or categories, such as age groups, education levels, and party affiliation. (Christian, & Dillman, 2004)

In the context of US presidential elections, pollsters have increasingly combined factors such as education, past vote, and party affiliation into their weighting schemes. For instance, weighting to education levels has become more common, with 83% of national polls in 2024 using this method, compared to just over half in 2016. (Sances, 2024) However, the effectiveness of these adjustments in improving polling accuracy is still a topic of debate. Some argue that weighting to party affiliation or past vote can introduce new sources of error, such as respondents misreporting their past vote or party affiliation. Others contend that these adjustments can help mitigate biases and improve the overall accuracy of polling data. (Silver, 2016)

While there are various methods and debates surrounding their effectiveness, the goal of weighting remains the same: to ensure that the sample data accurately represents the population, thereby providing a reliable snapshot of public opinion. Non-probability samples play a major role in the task of reliability of polling data ahead of US presidential elections. Unlike probability sampling, where every member of the population has an equal chance of being selected, non-probability sampling relies on subjective judgment, convenience, or other non-random strategies to pick participants. Types of non-probability sampling techniques include convenience sampling. Researchers choose members based on their accessibility and proximity, making it convenient to collect information. Secondly is the purposive sampling technique. Researchers handpick individuals who they consider are the most applicable or knowledgeable about the study topic. (Lessler, & Kalsbeek, 1992)

Quota sampling is another technique. Dividing the population into subgroups or strata and then putting a quota for each subgroup. Finally is snowball sampling. A researcher begins with some initial individuals and then asks them to refer others who match the study's criteria. Some of its advantages are cost-effectiveness, time-efficiency and flexibility as it allows researchers to adapt to changing circumstances. However, non-probability sampling also has several limitations, including bias and lack of generalizability. (Lessler, & Kalsbeek, 1992)

Non-probability sampling can have significant implications for polling data, predominantly in the context of US presidential elections. While non-probability sampling can provide valuable insights into public opinion, it can also introduce bias and affect the reliability of the results. Therefore, researchers and pollsters must carefully consider the advantages and limitations of non-probability sampling when designing their studies and interpreting their results. By doing so, they can increase the validity and reliability of their findings and provide more accurate insights into public opinion. Finally, certain demographic groups such as young people, low-income earners may be less likely to respond to surveys due to their low interest. (Kish, 1965)

Despite these challenges, polling data remains a vital tool for understanding public opinion and predicting election outcomes. To address the trials facing polling data, many polling organizations are turning to new methodologies and technologies. For example, some polls are using online platforms to collect data, which can provide more representative samples. Others are using advanced statistical techniques, such as weighting and stratification, to adjust for sampling bias. (Kay, & Yang, 2024)

Discussion of Major Research Findings

Unconfirmed opinions of voter fraud can significantly erode confidence in election outcomes in the US. Research has shown that media coverage of vote fraud allegations can increase public concerns about voter fraud, even when there is no evidence to support such claims. Consequences of unsubstantiated claims of voters trust include erosion of trust. Recurrent claims of voter fraud can lead to a decline in trust in the electoral process, making voters question the legitimacy of election outcomes. Unconfirmed claims of voter fraud can further polarize the electorate, with each side becoming more entrenched in their beliefs. Decreased voter turnout can also ruin electoral credibility. If voters believe that the electoral process is rigged or vulnerable to fraud, they may be less likely to participate in elections.

Polling has long been a vital tool for predicting electoral outcomes, with successes and failures shaping public trust in its reliability. Historically, polls have been accurate in evaluating overall trends, but key challenges have emerged. Increasing difficulty in reaching representative samples due to declining response rates, particularly with the rise of mobile phones and internet surveys. Some voters may not disclose their true preferences, leading to systematic underrepresentation of certain groups or candidates. Predicting who will vote remains one of the greatest challenges, with demographic shifts, voter suppression efforts, and enthusiasm gaps playing pivotal roles. Partisan polarization has aggravated the challenges of reliable polling. Polarized voters may be more likely to question the neutrality of polling organizations or refuse participation. Additionally, the rise of partisan-affiliated polling groups has introduced potential biases into aggregate results. Polling in polarized environments tends to amplify small biases, as voters' preferences align increasingly along ideological lines. Partisan nonresponse bias was identified in the 2016 and 2020 elections, where late shifts in voter sentiment were not adequately captured due to the timing of polls.

To combat these challenges, polling organizations have adopted innovative methodologies, including online and hybrid polling, that is, the use of internet panels combined with traditional methods to reach younger and more tech-savvy demographics. Integration of machine learning and voter file data to enhance predictive accuracy. Combining phone, online, and in-person polling to address accessibility issues. Public confidence in polling data is critical for its influence on voter behavior and campaign strategies. However, major errors in recent election cycles, such as 2016 and 2020, have eroded trust. Coverage errors, particularly among rural and non-college-educated voters, have been identified as key factors undermining trust.

The future reliability of polling data hinges on addressing these systemic issues while adapting to emerging trends. Accounting for increasing diversity in the electorate, mitigating the impact of disinformation and echo chambers on voter sentiment and safeguarding polling data from manipulation or external influences. While polling remains a valuable tool for understanding electoral trends, its reliability is increasingly challenged by methodological, technological, and societal shifts. Future improvements must address nonresponse bias, incorporate innovative technologies, and restore public trust to maintain its relevance in predicting presidential election outcomes.

Recent polling data for the 2024 US presidential election showed a tight race between Kamala Harris and Donald Trump. However, it's crucial to note that polling accuracy can be affected by various factors, including the margin

of error and the methodology used. Key challenges in polling accuracy such as selection bias, changing media consumption and social media biases must be tackled in order to increase more chances of accuracy in polls reliability. Best practices for evaluating polling data include transparency, considering use of multiple sources and watching for potential biases in polling data, including selection bias and social media biases.

Media coverage plays a crucial role in shaping public perceptions of voter fraud. Sensationalized reporting can create a false narrative that voter fraud is widespread, even when evidence suggests otherwise. The 2016 US presidential election saw numerous unsubstantiated claims of voter fraud, which were later debunked by fact-checkers and election officials. Despite this, many voters continued to believe that the election was rigged, leading to decreased confidence in the electoral process. In conclusion, unsubstantiated claims of voter fraud can have significant consequences for election outcome confidence in the US. It is essential for media outlets, politicians, and election officials to prioritize accuracy and transparency in their reporting and communications to maintain the integrity of the electoral process.

Conclusion, Summary and Recommendations

Conclusion

The reliability of polling data in predicting US presidential election outcomes has been a subject of debate. While polls provide valuable insights into public opinion, their accuracy can be influenced by various factors, including sampling bias, non-response bias, and external factors such as major events or economic shifts. Recent elections have highlighted the challenges of polling. For instance, the 2016 and 2024 US presidential elections saw most polls predicting a victory for Hillary Clinton and Kamala Harris respectively, while Donald Trump ultimately won in both. Similarly, the 2020 election witnessed discrepancies between state polls and actual outcomes.

Despite these challenges, polling remains a crucial tool for understanding public opinion. To improve the reliability of polling data, it is essential to address the factors that contribute to inaccuracy. This can be achieved by enhancing sampling methods, increasing transparency, and leveraging new technologies. Ultimately, the reliability of polling data depends on the ability of pollsters to adapt to changing voter behaviors and technological landscapes. By prioritizing transparency, innovation, and methodological refinement, polling can continue to play a vital role in understanding public opinion and predicting election outcomes.

Recommendation

The study provides the following recommendations to improve accuracy of polling data in US presidential elections

- **Address Non-Response Bias:** Implement strategies to minimize non-response bias, such as using multiple contact methods and offering incentives for participation.
- **Leverage New Technologies:** Utilize innovative methods like online polling, social media sentiment analysis, and data analytics to improve data collection and analysis.
- **Improve Question Wording and Design:** Ensure survey questions are clear, concise, and unbiased to elicit accurate responses.

- Regularly evaluate and refine methods: Continuously assess polling methodologies and refine them as needed to adapt to changing social and technological landscapes.
 - Encourage Collaboration and Data Sharing: Foster collaboration among pollsters, academics, and industry experts to share best practices, data, and methodologies.
 - Develop Advanced Statistical Models: Employ sophisticated statistical models, such as those used by FiveThirtyEight, to aggregate polls and provide more accurate predictions
- By implementing these recommendations, pollsters can improve the accuracy and reliability of polling data, providing better insights into public opinion and electoral outcomes.

References

- Christian, L. M., & Dillman, D. A. (2004). The influence of graphical and symbolic language manipulations on responses to self-administered questions. *Public Opinion Quarterly*, 68(1), 58-81.
- Cochran, W. G. (1977). *Sampling techniques* (3rd Ed.). John Wiley & Sons.
- Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Sage Publications.
- Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64(4), 464-494.
- Dahl, R.A. (2000). *On Democracy*. Yale University Press
- Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (2nd Ed.). John Wiley & Sons.
- Émile, D. (1893). Durkheim's doctoral thesis, where he explores the concept of social solidarity and the division of labor
- FiveThirtyEight. (2024). Do Pollsters' New Weighting Methods Make Their Surveys More Accurate?
- Fowler, F. J. (2013). *Survey research methods* (5th Ed.). Sage Publications.
- Gallup, G. (1972) Gallup opinion index. Reports No. 96. Princeton, NJ: Gallup Organization
- GeeksforGeeks. (2023). Non-Probability sample
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. John Wiley & Sons.

- Goyder, J. (1987). *The silent minority: Non respondents on sample surveys*. Westview Press.
- Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. John Wiley & Sons.
- Kish, L. (1965). *Survey sampling*. John Wiley & Sons.
- Lauderdale, B. E., Bailey, D., Blumenau, J., & Rivers, D. (2020). Model-based pre-election polling for national and sub-national outcomes in the US and UK. *International Journal of Forecasting*, 36(2), 399-413.
- Lavrakas, P. J. (1993). *Telephone survey methods: Sampling, selection, and supervision*. Sage Publications.
- Lessler, J. T., & Kalsbeek, W. D. (1992). *Non sampling error in surveys*. John Wiley & Sons.
- Lipset, S.M. (1996). *American Exceptionalism: A double-Edged Sword*. W.W. Norton & Company
- Lynn, P., & Clarke, P. (2002). Separating bias and variance in measurement error models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165(2), 169-184.
- NIWA Report, (2024) Precision or Predictions? Analyzing the Historical Accuracy of Top Political Polls and 538 (June, 30, 2024)
- Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, E., & Singer, E. (2004). *Methods for testing and evaluating survey questions*. *Public Opinion Quarterly*, 68(1), 109-130.
- Rosenbaum, P. R. (2010). *Design of observational studies*. Springer.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Academic Press.
- Silver, N. (2016). *Clinton-Trump Forecast*. FiveThirtyEight.
- Smith, A. (1997). *The wealth of Nations* (originally published in 1776). Penguin books.
- Sudman, S. (1980). Improving the quality of shopping center data. *Journal of Marketing Research*, 17(4), 453-461.
- Talcott, P. (1949). *A collection of essays showcasing Parson's development of structural functionalism*
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications.