https://doi.org/10.26219/heal.aueb.9743

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

ΜΕΤΑΠΤΥΧΙΑΚΟ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ
MSc IN APPLIED STATISTICS

Polling Misses: Causes and Treatment

By

Foteini Barali

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Applied Statistics

Athens, Greece
September 2025

https://doi.org/10.26219/heal.aueb.9743

Σφάλματα Δημοσκοπήσεων: Αιτίες και Αντιμετώπιση

Φωτεινή Μπαραλή

ΕΡΓΑΣΙΑ

Που υποβλήθηκε στο Τμήμα Στατιστικής

του Οικονομικού Πανεπιστημίου Αθηνών

ως μέρος των απαιτήσεων για την απόκτηση

Διπλώματος Μεταπτυχιακών Σπουδών

στην Εφαρμοσμένη Στατιστική

Αθήνα

Σεπτέμβριος 2025

Table of contents

Abstract

This thesis explores the recurring phenomenon of polling inaccuracies, or "polling misses," in election forecasting, examining the systemic, methodological, and behavioral factors contributing to these failures. It begins by establishing the historical context and evolution of election polling, highlighting its critical role in modern democratic processes, media narratives, and campaign strategies. Despite significant advancements in survey technology—from telephone-based to digital and multi-mode platforms—recent elections such as the 2016 U.S. presidential election, the Brexit referendum, and the 2018 Quebec provincial vote have demonstrated notable inaccuracies that challenge the reliability and legitimacy of polls.

Central to the analysis is an investigation of the structural vulnerabilities inherent in polling methods, including sampling errors, nonresponse bias, coverage gaps, and inadequate weighting procedures. It underscores the challenges posed by rapidly evolving communication habits and demographic shifts, illustrating how these factors systematically exclude or misrepresent key voter segments, thus skewing poll results. Additionally, the thesis identifies psychological phenomena such as social desirability bias, the "shy voter" effect, late-decider volatility, and the "bandwagon effect," emphasizing their roles in distorting polling accuracy.

Through detailed case studies—including notable polling failures in the United States, the United Kingdom, Quebec, and Australia—the thesis demonstrates that polling misses rarely result from isolated errors but rather from a complex interplay of methodological shortcomings and dynamic voter behaviors. It critically assesses contemporary methodological innovations designed to mitigate these errors, such as Multilevel Regression with Post-stratification (MRP), hybrid sampling designs, adaptive fieldwork, and real-time weighting adjustments.

The research ultimately advocates for a dual approach: continual methodological refinement paired with heightened transparency and ethical standards. By integrating rigorous statistical techniques with an understanding of voter psychology and behavior, pollsters can better navigate the complexities of modern electorates. This thesis contributes valuable insights and recommendations aimed at enhancing the accuracy, credibility, and utility of public opinion polling, ensuring it remains a vital and trusted component of democratic discourse and decision-making.

Περίληψη

Η παρούσα διπλωματική εργασία διερευνά το φαινόμενο των δημοσκοπικών αστοχιών (polling misses) στις εκλογικές προβλέψεις, εξετάζοντας τους συστημικούς, μεθοδολογικούς και συμπεριφορικούς παράγοντες που οδηγούν

σε αυτά τα λάθη. Αρχικά παρουσιάζεται το ιστορικό πλαίσιο και η εξέλιξη των εκλογικών δημοσκοπήσεων, υπογραμμίζοντας τον κρίσιμο ρόλο τους στις σύγχρονες δημοκρατικές διαδικασίες, την ενημέρωση των ΜΜΕ και τις προεκλογικές στρατηγικές. Παρά τις σημαντικές τεχνολογικές εξελίξεις στις δημοσκοπήσεις, από τις τηλεφωνικές έρευνες έως τις διαδικτυακές και τις πολυτροπικές προσεγγίσεις, πρόσφατες εκλογικές αναμετρήσεις όπως οι προεδρικές εκλογές των ΗΠΑ το 2016, το δημοψήφισμα για το Brexit και οι εκλογές του Κεμπέκ το 2018 ανέδειξαν σημαντικές ανακρίβειες, θέτοντας υπό αμφισβήτηση την αξιοπιστία και τη νομιμότητα των δημοσκοπήσεων.

Στον πυρήνα της ανάλυσης βρίσκεται η διερεύνηση των δομικών αδυναμιών στις μεθόδους δειγματοληψίας, συμπεριλαμβανομένων των σφαλμάτων κάλυψης, της μεροληψίας λόγω μη απόκρισης, των ελλείψεων στην αντιπροσωπευτικότητα του δείγματος και των ανεπαρκών διαδικασιών στάθμισης. Ιδιαίτερη έμφαση δίνεται στις δυσκολίες που προκύπτουν από τις ταχύτατες αλλαγές στις συνήθειες επικοινωνίας και τις δημογραφικές μεταβολές, οι οποίες συστηματικά αποκλείουν ή παρουσιάζουν λανθασμένα σημαντικά τμήματα του εκλογικού σώματος, προκαλώντας στρεβλώσεις στα δημοσκοπικά αποτελέσματα. Επιπλέον, η εργασία αναδεικνύει ψυχολογικά φαινόμενα όπως η κοινωνική επιθυμητότητα, οι «σιωπηλοί ψηφοφόροι» (shy voters), η αστάθεια στις επιλογές των αναποφάσιστων και το φαινόμενο του «bandwagon effect», τονίζοντας τον ρόλο τους στις δημοσκοπικές αστοχίες.

Μέσα από λεπτομερείς μελέτες περιπτώσεων, συμπεριλαμβανομένων γνωστών αστοχιών στις ΗΠΑ, το Ηνωμένο Βασίλειο, το Κεμπέκ και την Αυστραλία, η εργασία αποδεικνύει ότι τα δημοσκοπικά σφάλματα σπάνια προέρχονται από μεμονωμένους παράγοντες, αλλά είναι αποτέλεσμα ενός πολύπλοκου συνδυασμού μεθοδολογικών αδυναμιών και δυναμικών μεταβολών στη συμπεριφορά των ψηφοφόρων. Παράλληλα, αξιολογούνται κριτικά σύγχρονες μεθοδολογικές καινοτομίες για τον περιορισμό των σφαλμάτων αυτών, όπως η πολυεπίπεδη παλινδρόμηση με μεταστρωματοποίηση (MRP), οι υβριδικές δειγματοληπτικές μέθοδοι, οι προσαρμοστικές έρευνες πεδίου και οι προσαρμογές στάθμισης σε πραγματικό χρόνο.

Τέλος, η έρευνα προτείνει μια διπλή προσέγγιση: συνεχή μεθοδολογική βελτίωση σε συνδυασμό με ενισχυμένη διαφάνεια και ηθικά πρότυπα. Συνδυάζοντας αυστηρές στατιστικές μεθόδους με την κατανόηση της ψυχολογίας και της συμπεριφοράς των ψηφοφόρων, οι δημοσκόποι μπορούν να ανταποκριθούν καλύτερα στις προκλήσεις του σύγχρονου εκλογικού περιβάλλοντος. Η παρούσα διπλωματική εργασία παρέχει πολύτιμες γνώσεις και προτάσεις, με στόχο τη βελτίωση της ακρίβειας, της αξιοπιστίας και της χρησιμότητας των δημοσκοπήσεων, ώστε αυτές να συνεχίσουν να αποτελούν σημαντικό και αξιόπιστο εργαλείο της δημοκρατικής διαβούλευσης και λήψης αποφάσεων.

Chapter 1: Introduction

Public opinion polling has become one of the most recognizable features of modern electoral politics, providing scholars, journalists, policymakers, and citizens with a continuous stream of quantitative insights into the public mood. Especially in the context of United States presidential elections, polls have served a dual purpose: on the one hand, they have advanced academic understanding of voter behavior; on the other, they have guided the strategic planning and execution of political campaigns. From its modest and largely U.S.-centered origins in the 1930s to its present-day global application, election polling has shaped both the empirical study of political behavior and the practical decisions of political actors. The increasing accessibility of fast and low-cost survey technologies, ranging from telephone surveys in the post-war era, to online panels in the 1990s, and to mobile and mixed-mode platforms in the present, has helped solidify public opinion polling as an indispensable tool for campaign management, agenda setting, and media reporting. The increasing reliance on polling data across media and political institutions reflects a broader societal trust in quantitative measurements of sentiment. However, this trust is frequently undermined by the persistent and widely publicized failures of polls to accurately forecast election outcomes.

The problem is not occasional, statistically expected error but rather a pattern of misses that transcend national and institutional boundaries. The shock results of the 2016 U.S. presidential election, the 2016 Brexit referendum, the 2015 UK general election, the 2018 Quebec provincial vote and, most recently, the 2019 Australian and 2020 U.S. contests all featured disparities that exceeded conventional margins of sampling error. These failures, far from being isolated anomalies, point to recurring patterns of misestimation, often the result of methodological deficiencies such as nonresponse bias, unrepresentative sampling, and inadequate adjustment for key demographic and attitudinal variables. Such patterns suggest deeper, systemic issues that transcend national boundaries and challenge the legitimacy of polling as a reliable forecasting instrument.

The recurrence of these failures raises fundamental questions about the reliability of polls as forecasting tools and, by extension, about the media narratives, campaign tactics and public expectations that polls so powerfully influence. While a certain degree of error is intrinsic to any probabilistic sampling technique, the consistency, magnitude, and impact of recent polling inaccuracies demand serious reconsideration.

Persistent inaccuracy does more than embarrass pollsters; it risks eroding democratic trust. When forecasts mislead parties and voters alike, they can alter resource allocation, discourage turnout or trigger bandwagon and boomerang effects that reshape the very outcomes polls attempt to measure. Because polling translates statistical samples into sweeping claims about the popular will, systematic bias threatens not merely technical validity but the perceived legitimacy of political processes. For this reason, the debate over polling error has moved from methodological journals to mainstream commentary, provoking calls for transparency, better data stewardship and alternative modes of gauging public sentiment.

This thesis treats polling misses as signals of deeper structural, behavioural and technological challenges rather than as isolated forecasting glitches. Its first objective is diagnostic: to catalogue and analyse the primary mechanisms—sampling, weighting, mode effects, behavioural biases and late swings—that repeatedly generate error across diverse electoral contexts. The second objective is prescriptive: to evaluate the efficacy of emerging corrective strategies, from model-based poststratification and turnout calibration to hybrid probability–non-probability designs, river and intercept sampling, passive digital trace analysis and adaptive, rolling fieldwork. By comparing how these interventions perform across a set of emblematic case studies—1948 U.S., 1992 UK, 2016 U.S. and Brexit, 2019 Australia, 2020 U.S.—the study illuminates both their promise and their limitations.

In pursuing these aims, the thesis integrates perspectives from survey methodology, political psychology and communication research. Concepts such as social desirability bias, affective polarisation and the spiral of silence help explain why certain voter subgroups systematically evade measurement or misstate their preferences. At the same time, advances in computational statistics, notably Multilevel Regression and Poststratification (MRP) and Bayesian hierarchical forecasting, offer avenues for mitigating some—but not all—of these blind spots. The analysis therefore emphasises an interdisciplinary approach, contending that a full account of polling error must join statistical refinement with behavioural insight and contextual awareness.

Ultimately, the study argues that safeguarding the credibility of electoral polling in an era of fragmented media ecosystems and volatile electorates requires a dual commitment: methodological innovation coupled with ethical and communicative transparency. Polling will remain an influential social institution so long as it can demonstrate both technical competence and representational fairness. By critically assessing past failures, comparing corrective initiatives and outlining best practices, this thesis seeks to contribute to a more resilient architecture for measuring—and understanding—public opinion in modern democracies.

Chapter 2: Literature Review and Theoretical Framework

2.1 Development of Election Polling

Public-opinion polling is now an integral part of democratic life, offering researchers, journalists, policy-makers, and citizens a steady flow of data on the electorate's preferences. Since its first large-scale application during the 1936 U.S. presidential election, polling has evolved from an early experimental approach into a global industry that shapes campaign tactics, news agendas, and policy debates. Yet, even as each new survey technology promises broader reach, lower cost, and finer-grained insight, the accuracy of polls has remained uneven, revealing the deep methodological challenges involved in measuring public opinion (Jennings & Wlezien 2018).

Comparative research confirms that polling errors are neither recent nor confined to any single country. High-profile misses have punctuated electoral history, from the United Kingdom's 1992 and 2015 general elections to the 2016 U.S. presidential race and the Brexit referendum of the same year (Sturgis et al. 2016; Kennedy et al. 2018; Durand et al. 2001). Jennings and Wlezien (2018) show that the magnitude of these errors fluctuates with sampling frames, turnout models, and voter volatility, underscoring that no polling method, nor any electoral system, is immune to sizable misestimates.

Advancements in technology have facilitated progress while simultaneously creating new weaknesses. Over the decades, polling technology has evolved from the use of telephone surveys in the post-war era to mobile and web panel surveys in the twenty-first century. With every new innovative polling technology, there has always been forecasting failures, creating doubts about the behavioral fundamentals and assumptions upon which the polls are grounded.

These concerns have become particularly acute in the wake of recent election shocks. The 1992 UK general election, the 2016 U.S. presidential contest, Brexit, the 2019 Australian federal vote, and the 2020 U.S. swing-state results all generated polling errors well beyond conventional margins, rivalling earlier surprises such as the Dewey–Truman upset (1948), Canada's 2011 Conservative victory, and Québec's CAQ landslide in 2018. A closer look at these episodes points to a common set of pressures—from collapsing response rates and mobile-only households to late campaign swings and social-desirability effects—that standard designs struggle to absorb.

The infrastructure of polling itself is now in flux. Kennedy, Popky, and Keeter's (2023) Pew report documents a steep decline in landline response rates—from roughly 30 percent in 2000 to under 8 percent in 2022—and the rapid migration to mobile-only communication. With more than 70 percent of U.S. adults reachable only by cell phone, landline-centred RDD has become untenable, prompting a wholesale shift toward address-based sampling (ABS) and "push-to-web" invitations. Since 2021, every major Pew national survey has begun with mailed recruitment letters and routed respondents online (or, when necessary, to paper), signalling a decisive break from the telephone era.

In parallel, probability-recruited online panels such as Pew's American Trends Panel, Ipsos KnowledgePanel, and NORC's AmeriSpeak have emerged as partial remedies to rising costs and coverage gaps. While these panels preserve known selection probabilities, they still face cumulative response rates below five percent, and their benchmark errors—though smaller than those of opt-in panels—rarely disappear without aggressive weighting. Kennedy et al. (2023) underscore that advanced adjustments such as multilevel regression and post-stratification can dampen, but seldom eradicate, bias in non-probability samples. The lesson is clear: the future of polling depends on blending probability recruitment with flexible mixed-mode collection and increasingly sophisticated weighting.

Finally, the Pew authors chart a ten-point drop in public confidence that polls are conducted fairly between 2012 and 2022. In response, the industry has moved toward stricter disclosure norms—the AAPOR Transparency Initiative 2.0, Pew's "Methods Explorer," and routine publication of design effects and weighting schemes. Kennedy, Popky, and Keeter argue that methodological innovation alone will not restore credibility; pollsters must also communicate limitations candidly and engage in continuous experimentation to keep pace with changing technologies and respondent behaviour. Together, these developments map a field undergoing rapid, multifaceted transformation—one that is re-engineering its sampling frames, embracing digital collection modes, and redefining professional standards to remain a reliable barometer of public sentiment.

2.2 The history of polling misses

The 1936 U.S. Presidential Election

The 1936 U.S. presidential election is historically significant, particularly due to the infamous polling error committed by the magazine *Literary Digest*. According to Squire (1988), the election featured incumbent President Franklin D. Roosevelt, a Democrat seeking re-election amid the challenging conditions of the Great Depression, against Republican candidate Alf Landon. Despite Roosevelt's broad popularity, the *Literary Digest* poll predicted a decisive victory for Landon. The magazine's survey was extensive, based on around 10 million ballots mailed out to potential voters identified from sources such as automobile registrations, telephone directories, and subscription lists (Squire, 1988).

The results published by *Literary Digest* forecasted that Landon would win with 57 percent of the vote against Roosevelt's 43 percent, which sharply contradicted the actual electoral outcome. In reality, Roosevelt secured a resounding victory with approximately 61 percent of the popular vote, one of the most substantial wins in U.S. electoral history. This glaring discrepancy between poll prediction and actual results was due largely to the sampling method used by the magazine. As Squire (1988) explains, the magazine's sampling frame disproportionately represented wealthier voters who could afford telephones, cars, or magazine subscriptions, thus systematically excluding significant portions of the electorate who were more likely to support Roosevelt's policies addressing economic hardship.

This notable polling failure of 1936 had substantial implications for survey methodology, initiating a wave of professional and academic discussions focused on understanding the reasons behind such an unprecedented discrepancy. It highlighted critical flaws inherent in non-random sampling techniques and underscored the importance of achieving representative samples in polling practices. Consequently, George Gallup, who employed smaller but methodologically rigorous samples, correctly predicted Roosevelt's victory, thereby establishing credibility for modern scientific polling. The lessons drawn from the *Literary Digest* episode significantly advanced the field of public opinion research by demonstrating the necessity of random, representative sampling as a cornerstone for accurate electoral forecasts (Squire, 1988).

Early analyses conducted by leading pollsters and scholars attributed much of the error to nonresponse bias, suggesting that voters favoring Roosevelt were less likely or willing to return their ballots, which significantly skewed the results toward Landon (Crossley, 1937; Gosnell, 1937, as cited in Squire, 1988). Additionally, scholars such as Katz and Cantril (1937, as cited in Squire, 1988) highlighted the fundamental sampling flaws in the poll's approach. They argued that the *Literary Digest's* reliance on affluent voters drawn from subscription lists, telephone directories, and automobile registration records systematically excluded significant segments of the voting population, especially those economically disadvantaged voters who strongly supported Roosevelt's economic reforms. Consequently, this case underscored the critical necessity of employing random, representative sampling methods and helped establish the foundations of modern, scientifically rigorous polling practices (Squire, 1988).

The 1948 U.S. Presidential Election

The 1948 U.S. presidential election is recognized as an early and influential instance of polling failure, when leading organizations like Gallup, Roper, and Crossley confidently forecast a decisive win for Republican Thomas E. Dewey over incumbent Harry S. Truman. In a major methodological misstep, Gallup and others stopped polling weeks before the election, mistakenly assuming voter preferences would remain static. This resulted in a historic upset as Truman triumphed, surprising nearly all major analysts and media.

Recent scholarly research identifies several core reasons for this polling failure. First, pollsters did not adequately account for late shifts in voter sentiment; many undecided or wavering voters moved toward Truman in the campaign's final weeks—a phenomenon well documented in studies of late-deciding voters. Second, the reliance on quota sampling rather than probability sampling led to systematic biases, disproportionately under-representing some social groups more likely to support Truman. Third, the assumption that opinions were "frozen" weeks before the election overlooked the dynamism of public attitudes and the crucial importance of continuous polling through to election day. Finally, the industry's overconfidence, exemplified by the premature cessation of polling, left them ill-equipped to detect and model these late shifts.

In the aftermath, the Social Science Research Council convened a major review (Mosteller et al., 1949), which prompted significant reforms: improved sampling methods, more sophisticated turnout predictions, and the practice of polling as close to election day as possible in later cycles. The 1948 episode established the necessity for methodological rigor, humility, and adaptability in interpreting polling data—lessons that continue to inform the field today.

The 2015 United Kingdom General Elections

In more recent times, the UK has experienced repeated polling surprises. The 2015 UK general election stands as a particularly striking instance of polling error, highlighting significant methodological challenges in contemporary electoral forecasting. In the 2015 general election, most major polls suggested a tight race between the Conservative Party and the Labour Party. Leading up to the election, prominent polling organisations regularly published results indicating a highly competitive race, with many surveys showing a slight advantage for the Labour Party or predicting a tight, almost evenly matched contest between Labour and the Conservative Party (Mellon & Prosser, 2016). Instead, the Conservatives achieved a clear majority. These predictions shaped media coverage, public expectations, and party strategies, creating widespread anticipation of either a Labour-led coalition or another hung parliament.

Nevertheless, the actual results deviated notably from these predictions, presenting a dramatically different picture. On election night, the Conservative Party secured an unexpected outright majority, gaining substantially more parliamentary seats than anticipated, whereas Labour fell significantly short of the polling forecasts (Mellon & Prosser, 2017). This discrepancy between forecast and reality triggered extensive debate among researchers, pollsters, and journalists, prompting urgent calls for introspection within the polling industry to understand what had gone fundamentally wrong. A post-election inquiry by the British Polling Council found that the error stemmed primarily from unrepresentative samples, especially the under-representation of older voters who were more likely to vote Conservative (Sturgis et al., 2016).

In their detailed analysis, Mellon and Prosser (2017) pinpointed the core cause of this polling misjudgment as stemming from a critical issue in sample representativeness combined with incorrect modelling of voter turnout. Their study revealed that respondents who participated in pre-election polls tended to be disproportionately politically engaged individuals—voters more inclined

https://doi.org/10.26219/heal.aueb.9743

to support Labour. Simultaneously, less politically active and less frequent voters, who were more likely to support the Conservatives, were significantly underrepresented or absent altogether from the samples used by pollsters. This imbalance created a systematic bias, leading pollsters to overestimate Labour's electoral prospects and underestimate the Conservatives' appeal among the broader electorate.

Ultimately, the 2015 UK election underscored the vital importance of methodological rigour in polling practices. Mellon and Prosser’s findings reinforce that robust and representative samples, combined with precise adjustments for turnout likelihood, are indispensable elements for accurate electoral forecasting. Their approach not only corrected past mistakes but also set a precedent, highlighting the necessity of continual methodological innovation within election polling research to maintain credibility, improve accuracy, and restore public confidence in opinion polling (Mellon & Prosser, 2017).

Brexit (2016)

The 2016 Brexit referendum exposed notable challenges in polling accuracy, with most surveys predicting a narrow Remain victory but the UK ultimately voting 52% to 48% to leave the European Union. Several academic analyses attribute these polling misses to multiple factors. A key issue was the divergence between telephone and online polls, with internet-based surveys tending to capture the Leave lead more accurately, while telephone polls overestimated Remain support (Clarke, Goodwin & Whiteley, 2016). The polling errors were also linked to the difficulty of modeling turnout, as the referendum occurred with an unusually high turnout, particularly among voters who had historically been less politically engaged and who disproportionately supported Leave (Cardiff University, 2016). Some polling methods did not adequately adjust for this demographic shift, leading to under-sampling of Leave voters, especially in certain regions (Curtice, 2016).

Moreover, the late shifts in voter opinion leading up to polling day, together with potential social desirability bias causing some Leave supporters to withhold their true preferences, contributed to inaccuracies (British Polling Council, 2016). Pollsters' decisions about handling undecided voters and weighting final results also played a role in overestimating Remain's position. Collectively, these methodological challenges highlight both the complexity of referendum polling and the systemic issues in conventional polling that were revealed by Brexit, calling for improved survey techniques and turnout modeling in future political polling (Clarke, Goodwin & Whiteley, 2016; Cardiff University, 2016).

The 2016 and 2020 U.S. Presidential Elections

Following several strong forecasts in the national polls, the 2016 U.S. presidential election became a high-profile example of polling failure (Kennedy et al., 2018). Most media outlets had the Democratic nominee, Hillary Clinton, in the lead, yet state-level surveys routinely underestimated Donald Trump's support in key battlegrounds such as Michigan, Wisconsin, and Pennsylvania. The unexpected Electoral College victory for Trump surprised both pundits and the public alike. After states certified their results, researchers immediately set out to analyze the discrepancies and identify the methodological factors that had led polls astray.

Donald Trump's victory in the 2016 presidential election came as a shock to pollsters, political analysts, reporters, and pundits, including those inside Trump's own campaign (Jacobs and House 2016). Leading up to the election, three types of information widely discussed in the news media indicated that Democratic nominee Hillary Clinton was likely to win. First, polling data showed Clinton consistently leading the national popular vote, which is usually predictive of the winner (Erikson and Wlezien 2012), and leading, if narrowly, in Pennsylvania, Michigan, and Wisconsin—states that had voted Democratic for president six elections running. Second, early voting patterns in key states, particularly in Florida and North Carolina, were described in high-profile news stories as favorble for Clinton (Silver 2017a). Third, election forecasts from highly trained academics and data journalists declared that Clinton's probability of winning was about 90 percent, with estimates ranging from 71 to over 99 percent (Katz 2016).

Following the certification of vote totals in the aftermath of the 2016 United States presidential election, researchers commenced a systematic evaluation of polling performance. It became evident that a combination of factors contributed to the perception that the polling errors were more severe than they may have actually been, particularly in certain respects. Notably, the divergence between the popular vote, won by Hillary Clinton, and the Electoral College outcome, secured by Donald Trump, intensified scrutiny. While such a split outcome is not unprecedented, it remains statistically uncommon in the broader context of U.S. electoral history. With regard to polling accuracy, pre-election estimates indicated a more competitive Electoral College landscape than was often conveyed in media narratives (Trende, 2016; Silver, 2017b). Specifically, eight states comprising over one-third of the electoral votes required to win the presidency had polling margins within three percentage points (Trende, 2016). According to Trende, his organization's polling averages for battleground states placed Clinton ahead in the Electoral College by a narrow 272 to 266 margin, suggesting that Trump was within striking distance of an electoral victory. Additionally, the outcomes in three key Upper Midwest states—Pennsylvania, Michigan, and Wisconsin—which unexpectedly shifted in favor of Trump, were determined by exceptionally narrow margins. Collectively, over 13.8 million votes were cast in these states, yet Trump's total margin of victory amounted to just 77,744 votes, or approximately 0.56%. In such tightly contested races, even the most methodologically rigorous polls may be unable to predict outcomes with consistent accuracy.

Despite these contextual factors, the 2016 election raised significant questions regarding the reliability of polling, particularly the consistent underestimation of Trump's support in state-level, and to a lesser extent, national polls. The precise causes of this pattern were not immediately apparent, but identifying them remains critical to minimizing bias in future polling efforts. Furthermore, disparities in results between different polling methodologies—such as online surveys versus live telephone interviews—prompted inquiries into their relative accuracy and the reasons behind such variation. More broadly, these developments underscored the need to assess how the performance of 2016 pre-election polls compared with those of previous electoral cycles.

Polling organizations later attributed the miss to several factors, including under-representation of non-college-educated white voters, who favored Trump, and incorrect assumptions about voter turnout (Kennedy et al., 2018). Moreover, some voters may have decided late or concealed their intentions due to social stigma—a phenomenon sometimes labelled the “shy Trump voter” effect.

The 2020 U.S. presidential election stands out prominently among recent examples of substantial polling misses, generating renewed skepticism regarding polling accuracy and methodological robustness. Despite considerable advancements in polling methods following the polling inaccuracies of 2016, the 2020 election demonstrated persistent biases and significant predictive failures, particularly at the state level. According to Panagopoulos (2021), pre-election polls consistently overestimated support for Democratic candidate Joe Biden while underestimating Republican incumbent Donald Trump. These inaccuracies were widespread, impacting key battleground states where Biden's projected margins far exceeded his actual electoral performance, leading pollsters and researchers alike to question fundamental assumptions and practices within contemporary polling methodologies (Panagopoulos, 2021).

Several factors contributed to these persistent inaccuracies, as outlined by Panagopoulos (2021). A critical issue was the continued underrepresentation of specific voter demographics—particularly white voters without college degrees, rural voters, and voters less inclined to participate in polls, groups known from prior elections to disproportionately favor Donald Trump. Moreover, the polls appeared to systematically overrepresent demographic groups typically aligned with the Democratic party, including younger, college-educated, and urban respondents. Despite attempts by pollsters to correct these imbalances through advanced weighting techniques and adjustments, the underlying sampling biases remained largely unresolved, leading to consistently skewed results. This pattern indicates that sampling frames and survey modes remain insufficiently adjusted to capture the full complexity and evolving dynamics of the U.S. electorate.

Finally, the 2020 polling errors underscore deeper structural challenges that extend beyond technical or statistical adjustments. Panagopoulos (2021) emphasizes that nonresponse bias remained a key contributor to polling inaccuracies, suggesting that supporters of Donald Trump may have been systematically less willing to participate or respond accurately to polls, possibly reflecting broader social or psychological dynamics at play. Furthermore, the unusual context of the 2020 election, conducted amidst the COVID-19 pandemic, introduced unprecedented variables, including dramatic shifts to mail-in and early voting, changes in voter engagement levels, and heightened political polarization. These contextual influences complicated pollsters' efforts to model voter behavior accurately and to anticipate late swings in voter sentiment, highlighting the need for continued innovation and adaptation in polling methodologies to better reflect the contemporary electoral landscape (Panagopoulos, 2021).

Quebec Elections 2018.

The provincial election of 2018 stands out as one of the most striking polling surprises in modern Canadian politics. In the final week of the campaign, reputable firms—including Ipsos, Léger and Mainstreet—projected a photo-finish between the Coalition Avenir Québec (CAQ) and the Parti libéral du Québec (PLQ). Instead, the CAQ won by 12.6 points. Media outlets and academics quickly branded the episode a "major polling disaster," yet the evidence suggests a *polling miss* rather than a *polling failure*. Following Durand and Blais's distinction, a miss occurs when late-breaking voter movements create a gap between poll and result, whereas a failure points to deeper methodological flaws that would have prevented accurate estimation even if voters had stood still (Durand & Blais, 2020).

A post-election audit reveals that most Québec pollsters met basic design standards: samples were weighted for key demographics, question wording was conventional, and headline figures were internally consistent. The discrepancy emerged because a sizeable bloc of late-deciding and initially non-disclosing voters swung toward the CAQ during the campaign's closing days—particularly in two regions outside Montréal and Québec City. Re-contact evidence, including a follow-up survey by Léger 360, confirms that many respondents who had expressed uncertainty or softness in their vote intention ultimately broke disproportionately for the CAQ. In short, voter behaviour rather than sampling design generated the bulk of the error, although persistent under-performance on the Island of Montréal hints that some methodological issues—possibly language-group coverage or weighting of highly diverse ridings—also played a role (Durand & Blais, 2020).

The Québec experience echoes a broader pattern seen in other high-profile misses (U.K. 1992, Brexit 2016, U.S. 2016): polls struggle when turnout or preference shifts accelerate after the final fieldwork window, or when undecided voters coalesce late behind a single option. It also underscores how regional heterogeneity can magnify last-minute swings; missing only a handful of ex-urban ridings can shift a provincial vote share by a couple of points—enough to invert the narrative of a "tight race" (Durand & Blais, 2020).

Practical lessons follow. First, fieldwork should extend as close to Election Day as logistics allow, supplemented by rolling or tracking samples that can detect momentum. Second, pollsters and news organizations must communicate the growing volatility of late deciders, framing results as probabilistic snapshots rather than fixed forecasts. Third, researchers should couple polls with post-election re-contacts to disentangle behavioral change from design error, mirroring the multi-wave approach used by Sturgis et al. (2016, 2018) in the U.K. Finally, the public should embrace a healthy skepticism: election surveys are accurate most of the time, but their precision is always conditional on stable turnout and preferences, conditions that, in the era of voter volatility, cannot be assumed (Durand & Blais, 2020).

Australian Federal Elections 2019

 https://doi.org/10.26219/heal.aueb.9743

In the lead-up to the 2019 Australian federal election, national opinion polls—which had in previous cycles proved reasonably accurate, converged on a forecast of a Labor victory. This pervasive narrative shaped expectations among journalists, political actors and the public, so much so that the Coalition's eventual win came as a genuine surprise.

Mansillo and Jackman (2020) fit a Bayesian "state-space" model to every national two-party-preferred poll conducted between 2016 and 2019, explicitly estimating (a) the underlying trajectory of voting intentions, (b) "house effects" or systematic biases unique to each pollster, and (c) a discrete jump in public support coinciding with Malcolm Turnbull's replacement by Scott Morrison in August 2018.

The principal source of polling error in 2019 was a consistent underestimation of Coalition support coupled with overestimation of minor-party votes—most notably by YouGov Australia. Although some bias echoes the under-Labour, over-Greens pattern observed before 2016, the magnitude of new errors between 2016 and 2019 was far greater than anything seen in prior campaigns.

What made the 2019 miss especially stark was that errors in first-preference estimates did not "wash out" once converted to two-party-preferred figures. As a result, the final aggregated polls delivered a two-party-preferred estimate that was larger—and wronger—by historical standards, directly leading to the (mis)prediction of a Labor victory and triggering a crisis of confidence in polling.

Finally, while Mansillo and Jackman reaffirm that national two-party-preferred polls generally outperform seat-specific surveys in reliability, this time even the national aggregates failed to capture the late-campaign swing to the Coalition. In contrast, seat-level polls remain, on average, lower quality and far more volatile

2.3 Lessons from Repeated Failures

An examination of prominent polling failures reveals several consistent themes. Firstly, polling errors tend to be systematic rather than purely random,

 https://doi.org/10.26219/heal.aueb.9743

often rooted in identifiable demographic, methodological, or behavioral factors. Common recurring errors include inaccurate turnout modeling, sampling biases, and the handling of undecided voters, indicating underlying structural challenges within polling methodologies. Secondly, as electorates have grown increasingly complex due to factors such as political polarization, societal fragmentation, and rapid technological shifts, the difficulties facing pollsters have intensified accordingly.

These persistent failures have driven critical reflection and innovation within the field, prompting advancements in survey design, weighting techniques, and data collection methods. However, they have also contributed to growing public skepticism toward polling results and a corresponding demand for increased transparency and methodological rigor.

The history of polling inaccuracies demonstrates that no organization is immune to errors, even those with extensive experience and resources. While some degree of error is inevitable in sampling-based research, the recurrence and scale of these failures underscore the urgent need for the scientific evolution of polling practices to maintain their relevance and reliability. Appreciating these historical patterns is crucial for diagnosing present-day challenges and guiding the development of more robust and adaptive approaches to electoral forecasting, a focus of the subsequent chapters.

Chapter 3: Challenges faced by Sampling Methods

Election polling now confronts unprecedented headwinds. Response rates have fallen to single digits, eroding the once-secure foundations of probability-based designs (Baker et al., 2010; Wu, 2022). In response, many organizations have turned to opt-in, non-probability panels that violate the core requirement that response propensity be unrelated to political attitudes, thereby introducing non-ignorable bias (Wang et al., 2015; Meng, 2018).

 https://doi.org/10.26219/heal.aueb.9743

Weighting can correct known demographic gaps only when selection is ignorable; in opt-in surveys that condition fails, so post-stratification may even amplify error when willingness to participate co-varies with vote intention (Meng, 2018). Meng's "data-defect correlation" framework reminds us that tiny deviations from randomness, once scaled to a nation-sized population, can yield strikingly large errors.

The wholesale shift toward mobile and online interviewing layers on additional coverage gaps. Electorate segments that remain offline, phone-averse, or otherwise hard to reach simply vanish from the frame, systematically skewing estimates (Wu, 2022; West & Andridge, 2023). When low participation, imperfect weighting, and uneven coverage intersect, the result is a triple threat that demands a paradigm able to model non-ignorable selection rather than relying on ad-hoc, post-hoc fixes.

Sampling remains the conceptual foundation of every pre-election poll, yet history is replete with cases in which sampling schemes failed to capture the electorate's demographic and political diversity. Classic pitfalls—selection bias, coverage error, and non-response—interfere with representativeness and, by extension, predictive accuracy.

Complicating matters further is the elusive target population. The ideal universe contains only those who will actually cast a ballot, but that electorate has not fully materialised at interview time. Surrogate frames such as registered voters or self-reported likely voters introduce fresh uncertainty and systematically exclude some demographic or ideological strata.

Taken together, these structural vulnerabilities surface as "polling misses", errors that routinely exceed what random sampling variability would predict. Coverage deficiencies, turnout misclassification, and demographic underrepresentation all conspire to widen the gap between poll and outcome.

The persistent nature of polling errors, as documented also by Jennings and Wlezien, highlights the critical need to explore their underlying theoretical mechanisms. These scholars identify several methodological challenges, such as declining response rates, nonrepresentative sampling, late swings in voter sentiment, and social desirability biases, which collectively contribute to polling inaccuracies (Jennings and Wlezien, 2018; Curtice and Sparrow 1997; Shy and elusive voter theories discussed by Durand et al. 2001). By systematically addressing and examining these theoretical dimensions, researchers can better understand the limitations of polling and subsequently enhance the methods employed to improve future electoral predictions.

The following sections go through each mechanism individually, showing how it interferes with the lens of statistics, revealing methods that can minimize the effects, and detailing how each mechanism captures the essence of the problem. Addressing this particular issue will need creativity in sampling methods, clear reporting of calibration models, and anticipatory models of population and behavioral changes. Only after can pre-election opinion polling regain its reliability and usefulness as an indicator of public opinion.

3.1 Sampling Frame & Coverage

Election polls begin with a deceptively simple question—*who* should be asked—but recent shifts in technology and social behavior have made that choice unusually perilous. Classical probability designs once relied on random-digit-dial telephone samples that could plausibly reach most households, yet response rates have fallen into single digits (Baker et al., 2010). As landline ownership shrank and caller-ID screening grew, entire pockets of the electorate drifted outside the frame. Younger, mobile-only voters; renters who change numbers frequently; and rural residents in cellular “dead zones” now slip past the interviewer’s net, producing systematic gaps that weighting alone cannot mend (Wu, 2022; West & Andridge, 2023).

These challenges were vividly illustrated in the 2016 U.S. presidential election, as Jackson, Lewis-Beck, and Tien (2020) point out. Many polls failed to adequately cover demographic groups that strongly supported Donald Trump, particularly working-class voters in rural or semi-rural communities, and voters without college degrees. Polling methodologies often overlooked these groups because traditional sampling frames, such as landline-based random-digit dialing, systematically missed areas with lower telephone penetration rates, as well as communities that were culturally or geographically distant from mainstream urban centers.

Pollsters have reacted by stitching together multiple frames—address-based lists, voter-registration files, and opt-in web panels—but each patch introduces its own imperfections. Voter rolls contain duplicates, movers, and people who will never cast a ballot, while online panels overrepresent tech-savvy, politically engaged citizens (Meng, 2018). When these divergent sources are merged, the resulting mosaic can look comprehensive yet still miss low-income households without reliable internet or first-generation immigrants who distrust unsolicited surveys. Auxiliary data help locate the blind spots: early-voting returns, census micro-data on education and language use, and consumer databases that reveal where telephone or broadband coverage is thin (Wang et al., 2015).

Coverage errors are not merely academic; they ripple through turnout models and seat projections. A handful of missing precincts in ex-urban counties, or an undercount of apartment dwellers in fast-growing cities, can tilt statewide estimates by the one or two points that decide close contests. For that reason, leading pollsters now run pilot waves weeks before fieldwork begins, mapping where recruitment falters and adjusting incentives or mode mix to fill the holes. No sampling frame can ever be perfect, but a deliberately layered design—tested, benchmarked, and constantly audited—offers the best chance to ensure that every geographic and demographic niche of the electorate is at least *within reach* of the interviewer's call.

3.2 Non-Response & Weighting

Even a perfectly defined sampling frame cannot guarantee that the selected voters will actually pick up the phone, click the survey link, or complete the questionnaire. Over the past three decades, contact and cooperation rates have plunged from above 60 percent to well below 10 percent on most telephone surveys (Baker et al., 2010). The drop is not random: citizens who are time-pressed, politically disengaged, or distrustful of institutions are markedly less likely to respond, while retirees, habitual news-consumers, and partisans with

 https://doi.org/10.26219/heal.aueb.9743

strong opinions tend to stay on the line. The resulting non-response bias means that a small, eager subset of the electorate speaks on behalf of everyone else.

Weighting offers a partial remedy by giving more influence to under-represented groups, but its success depends on two fragile assumptions: first, that response is "ignorable" once we control for observed demographics, and second, that those demographics are measured and benchmarked accurately (Meng, 2018). Both assumptions have shown cracks. In 2016, many state-level U.S. polls omitted education from their weighting schemes; college-educated voters, who were both more likely to answer surveys and more inclined toward Hillary Clinton, were therefore over-represented, leading to a systematic understatement of Donald Trump's support (Kennedy et al., 2018).

Even when key variables are included, respondents do not always report them reliably. Past vote recall fades or becomes "retrospectively partisan," income is rounded or withheld, and self-identified ideology can shift with question wording. Each misreport propagates through the weights, adding noise or amplifying bias. Sophisticated approaches such as multilevel regression with post-stratification (MRP) promise finer calibration by modelling vote intention within thousands of demographic-geographic cells, yet MRP inherits the same vulnerability: if the auxiliary data used for post-stratification are wrong or out-of-date, the elegant model simply redistributes error with greater mathematical flourish (Wang et al., 2015).

Modern best practice therefore couples weighting with aggressive diagnostics. Pollsters now cross-check weighted estimates against early-vote tallies, administrative turnout files, and high-quality probability benchmarks. They run split-sample experiments, applying, say, education weights in one half and education-plus-internet-access in the other, to gauge how sensitive results are to the chosen scheme. When discrepancies persist, targeted re-contact campaigns can nudge hard-to-reach subgroups into the sample, reducing the need for extreme weights. Even so, the lesson of recent cycles is clear: weighting can correct for what we see, but it struggles with what we cannot observe, and each election brings fresh patterns of participation that pollsters must learn, test, and re-learn anew.

3.3 Mode & Interviewer Effects

How a survey is administered—and by whom—shapes what respondents are willing to say. Decades of research show that answers obtained by a human interviewer can differ markedly from those typed into an anonymous web form (Tourangeau & Yan, 2007). Telephone or face-to-face modes place respondents in a social situation where courtesy, time pressure, and a desire to avoid judgment encourage brisk, socially acceptable replies. Web questionnaires, in contrast, offer privacy and unlimited reflection time, often producing higher reports of sensitive behaviours but also attracting younger, more internet-savvy participants (Kreuter, Presser, & Tourangeau, 2008).

Jackson, Lewis-Beck, and Tien (2020) emphasize that these mode-related discrepancies were evident in the 2016 U.S. election polls. Specifically, mode effects impacted how openly respondents expressed support for controversial candidates, such as Donald Trump. Interviewer-administered polls, particularly telephone surveys, likely elicited more socially desirable responses, underreporting support for Trump due to the stigma perceived by respondents. Conversely, online surveys captured more candid support for such candidates but risked over-representing demographics comfortable with digital technology.

Interviewer presence introduces yet another layer. Respondents routinely adjust their answers to cues such as the interviewer's voice, gender, or perceived ethnicity. In U.S. pre-election polls, for example, Trump support rose when the interviewer was a white male and dipped when the interviewer was female or non-white—an asymmetry attributed to social-desirability pressure (Fisher et al., 2011). Similar patterns have appeared in European surveys on immigration or welfare, where socially disapproved opinions are voiced more freely to interviewers perceived as "in-group" (Holbrook, Cho, & Johnson, 2006).

Mixed-mode designs promise broader coverage but complicate comparability. A Likert scale read aloud on the phone may not map neatly onto a five-point grid on a smartphone screen; respondents scrolling quickly can skip instructions or misclick, while telephone respondents may settle on mid-points to shorten the call (Couper, 2017). If modes are blended without calibration, the resulting dataset can conflate real opinion differences with mode-induced measurement error, especially when key subgroups (older voters by phone, younger voters online) line up with partisan divisions.

Pollsters use several defences. First, they standardise wording and response options across platforms, then test equivalence in split-sample experiments. Second, they randomise mode assignment—contacting part of the sample by phone and part online—to quantify mode effects directly. Third, they model interviewer characteristics as post-stratification factors, re-weighting responses when, say, female interviewers recorded fewer "socially risky" answers than their male counterparts. Finally, they monitor paradata—call length, break-offs, device type—to flag systematic deviations in real time.

Even with these safeguards, mode and interviewer effects remain a stubborn source of error. They remind us that polling is not simply about *what* questions are asked but *how* and *by whom* they are delivered—a human encounter that can amplify or muffle the electorate's true voice.

3.4 Shy voters and social desirability bias

Partisan polarization has transformed vote choice from a simple policy preference into a badge of social identity. When the partisan divide maps onto cultural fault lines, declaring support for a polarizing figure such as Donald Trump can feel like crossing a bright line between "us" and "them." The cost is not just a spirited argument about policies; it can be social ostracism, awkward glances at the office coffee machine, or a flood of disapproving emojis on social media. Small wonder, then, that some citizens think twice before revealing their preference to a stranger wielding a clipboard or staring back through a video call.

This reluctance animates the so-called "Shy Tory" or "Shy Trump" effect, in which voters mask their leanings to dodge moral judgment or reputational damage (Fisher et al., 2011). The effect is most acute in mixed-mode surveys that rely on live interviewers: a subtle shift in tone, an inquisitive eyebrow, or the interviewer's perceived demographic traits can whisper that a response will be weighed—and perhaps found wanting. Faced with this social calculus, respondents retreat behind the safer banners of "undecided," "no preference," or, on rare occasions, the opponent's name.

Media coverage magnifies the problem. Headlines trumpeting "silent Trump voters" and post-mortems about hidden support teach citizens that certain choices are taboo. This narrative acts like a hall of mirrors: the more respondents believe their view is stigmatized, the more they hide it, seemingly confirming the story for the next news cycle.

Survey researchers have tried to detect camouflage. One strategy is to compare answers across interviewer demographics. If Trump support surges in interviews conducted by white men but dips when the interviewer is female or non-white, the asymmetry points to social-desirability distortion. The method is hardly foolproof, some voters stay guarded even with a demographically "comfortable" interviewer, but persistent gaps sound a useful alarm.

In response, pollsters experiment with anonymity-enhancing techniques: self-administered web questionnaires that lower the stakes, randomized-response formats that scramble individual answers beyond recognition, and list experiments that let respondents signal sensitive views obliquely. Each tool grants relief at a price, added complexity, higher respondent burden, or thinner coverage of offline populations. Grasping the psychological and sociological roots of concealment is therefore indispensable. Only by acknowledging and, where possible, easing the climate of apprehension can surveys hope to capture the electorate's genuine voice rather than its carefully curated public mask.

3.5 Differences between reported vote intention and actual vote ("late swing")

Elections are fluid contests, not static snapshots. Breaking news, debate performances, policy U-turns—or a single scandal—can flip public opinion in the campaign's final days. When polls conclude fieldwork before these late tremors, they freeze a moment that may vanish overnight, misleading both forecasters and the public.

This "late-swing" problem begins the instant a survey goes dark and only reveals itself once ballots are counted. Early-terminated polls create an illusion of stability, tempting analysts to believe voter sentiment has plateaued when it is still in motion. Forecasts anchored in yesterday's intentions can therefore diverge sharply from Election-Day reality.

History offers cautionary tales. In 1948 roughly 15 percent of Americans chose a candidate in the final fortnight, propelling Harry Truman to an upset win. Seven decades later, late-deciding voters in key Rust Belt states nudged the 2016 U.S. election toward Donald Trump, movement most early polls missed. Similar last-minute reversals colored the Brexit referendum, the 2019 U.K. campaign, and several European parliamentary races. Even meticulously weighted samples cannot compensate for the temporal gap between polling and voting.

Pollsters have responded with adaptive strategies. Rolling or tracking polls refresh a slice of the sample each day, producing a moving average that absorbs late tremors. Extending fieldwork closer to Election Day narrows the mismatch, though at a higher logistical cost. Real-time reweighting, fed by early-vote returns or turnout models, nudges estimates as fresh signals arrive. Bayesian dynamic forecasting and multilevel regression with post-stratification weave time directly into their equations, allowing predictions to bend when shocks land. None of these remedies can abolish uncertainty, but by keeping data collection as close as possible to the finish line, and by modelling the volatility explicitly, pollsters can shrink the gap between projected and actual outcomes.

3.6 Undecided Voters and Late Deciders

Another significant source of polling error comes from citizens who remain genuinely undecided until the eleventh hour. Unlike shy voters, they are not concealing a preference; they simply do not have one yet. Political ambivalence may reflect low interest, limited information, or a deliberate wait-and-see strategy that postpones commitment until voting becomes unavoidable.

For pollsters, the immediate problem is what to do with these respondents during the estimation process. Some organizations drop them from the sample, others redistribute them in proportion to decided voters, and still others use statistical models to infer where they might land. Each strategy carries risk. Assuming proportional allocation can backfire if undecided voters lean heavily in one direction—as happened in the 2015 UK general election and the 2016 U.S. presidential race, where late deciders in key regions moved decisively toward the Conservative and Republican candidates, respectively (Sturgis et al., 2016; Silver, 2017).

Because even a two- or three-point break among late deciders can flip marginal constituencies, timing matters immensely. Polls that finish fieldwork several days before the vote freeze preferences that are still in flux, giving a false sense of stability. Historical post-hoc analyses show that undetected movement in the final 48–72 hours can explain much of the gap between forecast and outcome.

Capturing this volatility requires both measurement and modelling. Leaning questions ("If you had to choose today, which way are you leaning?") coax soft signals from fence-sitters, while daily tracking polls trace the contraction of the undecided share as Election Day nears. Advanced approaches—Bayesian dynamic forecasting, multilevel regression with post-stratification, real-time reweighting using early-vote returns—explicitly incorporate uncertainty about where uncommitted voters will ultimately break. Yet no technique can abolish uncertainty entirely. The very fluidity that keeps some citizens undecided is also what makes them powerful agents of last-minute change, ensuring that every forecast remains, at best, a probabilistic bet rather than a definitive prediction.

3.7 Geographic Heterogeneity

Geographic heterogeneity refers to differences in voter preferences that vary across regions, creating challenges for accurate polling. Because political attitudes often cluster geographically—urban voters might differ significantly from rural ones—polls that don't carefully capture this diversity can produce misleading estimates. This issue becomes especially pronounced when pollsters rely on sampling methods that disproportionately reach certain regions or communities, unintentionally excluding harder-to-reach or lower-participation groups.

Prosser and Mellon highlight several recent instances where geographic heterogeneity contributed significantly to polling errors. For instance, polls might adequately capture voter sentiment in large metropolitan areas but fail to represent rural areas or smaller towns adequately, where political preferences can differ sharply. These regional imbalances can swing election outcomes, especially in closely contested states or districts. Thus, ignoring geographic diversity not only distorts polling results but also can significantly affect political strategies, misdirecting campaign efforts based on incomplete or biased data. Pollsters must therefore carefully design samples that reflect the full mosaic of voter preferences across geographic boundaries.

3.8 Herding

Another subtle but impactful source of polling error is known as "herding." Herding happens when polling organizations, consciously or unconsciously, adjust their results or methodologies to match what other pollsters are reporting. Rather than independently capturing voter sentiment, pollsters might feel pressure to produce results similar to those already published, either to protect their credibility or to avoid criticism for being an "outlier." While this might temporarily give the appearance of accuracy and consensus, it ultimately masks the genuine uncertainty that exists within the electorate (Prosser & Mellon, 2018).

One problematic consequence of herding is that it creates an artificial sense of stability and certainty. Polls might converge around an incorrect estimate simply because pollsters prefer safety in numbers. This was notably seen in high-profile polling misses such as the 2016 U.S. presidential election, where many organizations produced strikingly similar—and ultimately misleading—predictions. Herding thus serves as a cautionary tale: when polls look too consistent to be true, it might indeed indicate that their accuracy has been compromised by collective caution rather than genuine consensus (Prosser & Mellon, 2018).

3.9 "Bandwagon effect".

Since the earliest systematic surveys, scholars have suspected that polling trends can shape, not just reflect, electoral outcomes (Gallup & Rae 1940; Simon 1954). The core hypothesis is that when people see their social or political in-group gaining visible momentum in successive poll releases they shift their own preferences to ride that wave—a dynamic labelled the *bandwagon effect* (Simon 1954; Marsh 1985). A steady upward signal can thus become a self-fulfilling prophecy: voters join the apparent winner both to share the satisfaction of majority status and to avoid the discomfort of backing an underdog (Rothschild & Malhotra 2014). Several psychological mechanisms reinforce the shift—search for group-based confidence and happiness (Mondak 2010; Brady & McNulty 2011; Erikson & Stoker 2011; Gerber et al. 2011), faith in the "wisdom of crowds" (Lau & Redlawsk 2001; Hardmeier 2008), and the impulse to minimise cognitive dissonance through anticipatory attitude adjustment (Mutz 1997; Kay et al. 2002)—so a modest lead in the polls can swell far beyond what underlying, stable preferences alone would predict.

The paragraphs that follow will explore how this effect materializes in practice, how media coverage and campaign responses can reinforce bandwagon dynamics, and why overlooking its presence risks overestimating the durability of poll-driven surges.

The "bandwagon effect" describes the tendency of voters to shift their support toward the candidate or party that appears to be leading in the polls simply because that option seems popular or likely to win. When surveys show one campaign ahead, an increasing number of individuals may decide to back that apparent victor. They do so out of a desire to be on the winning side, to feel part of a majority or to avoid the social discomfort associated with supporting an underdog. In this way what began as a modest lead can swell into a much larger one, not because more people truly prefer that candidate, but because the polls themselves have altered perceptions of who is favored. Pollsters who measure support during this surge risk overestimating the candidate's genuine strength since some of the recorded support reflects voters' reactions to the polls rather than stable political preferences.

This effect interacts with the wider campaign environment, further complicating efforts to obtain accurate forecasts. Campaign teams that see their numbers rising or falling in successive polls often react by shifting resources, for example by investing in regions where they think momentum is strongest or launching late advertising pushes to exploit their perceived advantage. Media outlets then amplify these shifts by highlighting dramatic poll swings and promoting narratives of growing popularity. Voters who wish to align with what they believe is the prevailing consensus may switch their allegiance not because of new policy information or personal conviction but simply because they want to be part of a trend. As a result polls taken amid this feedback loop may confuse temporary enthusiasm with lasting support and thus report an inflated level of confidence in the front-runner.

Moreover, the bandwagon effect can obscure underlying pockets of resistance and conceal important last-minute changes in voter sentiment. As more people rally to the seemingly leading campaign those who continue to support trailing options become less visible in survey samples. This creates the false impression of a one-way surge and may cause pollsters to overlook small but strategically significant shifts that occur late in the campaign. In some cases, the very publication of a commanding poll lead can discourage supporters of the underdog from turning out or prompt undecided voters to commit early to the presumed winner. These dynamics mean that understanding the origins of the bandwagon effect the way it feeds into media coverage and campaign tactics and its power to mask genuine voter intentions is essential for interpreting poll data accurately and for designing survey methods that guard against reporting a misleading sense of momentum.

Conclusion

Polling misses rarely trace to a single flaw; they arise from the interaction of methodological weak spots and a political landscape that refuses to stay still. Non-response bias mutes the voices of younger, lower-income, and politically detached citizens, while coverage gaps emerge as random-digit-dial and address-based frames lag behind shifting communication habits and residential churn. Weighting errors can magnify these distortions when benchmarks are outdated or crucial variables—education, language, urban-rural residence—are missing. And turnout and undecided-voter models add a further layer of uncertainty, resting on behavioural assumptions that may fail in a volatile campaign season.

Recent innovations—multilevel regression with post-stratification, hybrid probability/opt-in designs, real-time reweighting—have narrowed some gaps, yet each new model introduces parameters that demand validation. No single technique can fully erase bias; public opinion research remains part statistics, part adaptive fieldwork, and part human psychology.

Notably, deliberate misreporting by "shy" voters explains only a small slice of error. Far more often, misses reflect failures of representativeness, late surges of previously undecided voters, or bandwagon dynamics in which visible poll trends nudge citizens and even pollsters toward the perceived winner. Capturing these last-minute swings requires rolling samples, rapid-response weighting, and transparent disclosure. By pairing methodological rigour with operational flexibility, pollsters can shrink the gap between eve-of-election estimates and the eventual count—improving reliability without claiming impossible precision.

These persistent challenges highlight the insufficiency of traditional polling and sampling approaches to fully capture today's complex electoral dynamics. As a result, the subsequent chapter examines promising methodological advancements—including address-based and hybrid sampling strategies, river sampling, and advanced statistical modeling techniques like multilevel regression with post-stratification (MRP), that seek to enhance forecast accuracy and reduce the prevalence of polling misses going forward.

Chapter 4: Treatment of Polling Misses

Election forecasting has traditionally depended on representative polling, wherein individuals are randomly sampled and queried regarding their voting intentions. Although this method has historically produced reliable outcomes, it entails substantial costs in terms of both time and financial resources. Furthermore, response rates to such surveys have steadily declined over recent decades, thereby diminishing the statistical reliability that representative sampling once offered. When electoral forecasts are based exclusively on conventional sampling techniques and survey design, there exists a substantial risk of drawing inaccurate or misleading conclusions. This risk is particularly pronounced when these methods fail to adequately capture the heterogeneity and complexity inherent in voter behavior across different demographic and geographic groups. The underlying premise of representative polling—that every individual within a target population, such as registered or likely voters, possesses an equal probability of selection—is increasingly challenged in practical applications due to issues such as nonresponse bias and coverage error.

In response to these challenges, researchers and practitioners have stressed the importance of rethinking sampling approaches and adapting them to contemporary social and technological realities. The steady decline of traditional methods has created space for the development of new strategies designed to improve representativeness and minimize systematic sources of error. While no single approach can fully eliminate the difficulties associated with electoral polling, the refinement of sampling techniques and the integration of more flexible designs have been recognized as central steps toward restoring the credibility of polls. Such innovations aim to balance statistical rigor with the practical need to capture the diversity of modern electorates, thereby providing a more accurate foundation for electoral forecasting.

This study demonstrates that, when accompanied by appropriate statistical adjustments, non-representative polling can serve as a viable alternative for generating accurate electoral forecasts. Notably, such forecasts can often be produced more efficiently and at a reduced cost relative to those derived from traditional survey methodologies. Model-based approaches provide a rigorous and adaptable framework for overcoming the limitations associated with conventional sampling methods. These approaches are particularly advantageous in their capacity to account for known biases in non-representative data and to incorporate auxiliary information that enhances predictive accuracy. Moreover, model-based methods are well suited for identifying and analyzing subnational trends, such as those at the state level, which are often obscured in aggregate polling data. The findings presented herein suggest that non-representative polling, when methodologically reinforced through statistical modeling and combined with adjustments in sampling practices, holds significant potential not only for electoral forecasting but also for measuring public opinion on a wide spectrum of political, social, economic, and cultural issues.

4.1 Changes in sampling designs

4.1.1 Address-Based Sampling with Push-to-Web Invitations

The first remedy highlighted in recent literature is the shift from traditional telephone frames to address-based sampling (ABS). In this design, researchers draw a random sample of household addresses from the national postal registry, then mail recruitment letters that invite residents to complete the survey online—or on paper if they lack internet access. This strategy reaches both landline-absent and mobile-only households while reducing the cost of fieldwork relative to live telephone calls. Although ABS now represents the gold standard for probability sampling, it still faces low response rates, often under ten percent, and therefore requires extensive weighting to correct resulting imbalances and preserve representativeness (Kennedy, Popky & Keeter, 2023).

4.1.2 Probability-Based Online Panels

This method involves building probability-recruited online panels such as Pew's American Trends Panel, Ipsos KnowledgePanel, and NORC's AmeriSpeak. Panelists are selected through ABS poll methods or dual-frame random digit dialing (RDD) and subsequently complete recurring web surveys, maintaining known selection probabilities even as data-collection costs drop sharply. Kennedy et al. (2023) report cumulative response rates below five percent and provide benchmark errors after raking to illustrate both the promise and the limits of this model. To counter demographic and behavioural skews that emerge as panels age, researchers apply multilayered weighting on an expanded set of variables. These panels offer stability and longitudinal depth, but they must be continually recalibrated to keep pace with population shifts and digital divides.

4.1.3 Hybrid Designs Combining Probability and Opt-In Samples
A third practice gaining traction in commercial settings blends a small probability sample (drawn via ABS or RDD) with a much larger opt-in online sample. The probability component serves as an “anchor” for calibration, while the voluminous opt-in data provide speed, granular detail, and lower cost. Advanced techniques such as propensity-score weighting and Multilevel Regression with Poststratification (MRP) are used to correct the self-selection bias inherent in the opt-in portion. Hybrid designs deliver flexibility and affordability, yet their quality hinges on the size and integrity of the probability subsample and on how effectively statistical adjustments compensate for unobserved differences between volunteers and non-volunteers (Kennedy, Popky & Keeter, 2023).

4.1.4 Multimode and Responsive Sampling Strategies
Finally, recent studies emphasize the increasing importance of multimode and responsive survey designs. Pollsters combine telephone interviews, web questionnaires, SMS prompts, and, when necessary, face-to-face visits, choosing for each demographic segment the contact mode most likely to elicit participation. During fieldwork, responsive-design protocols monitor sample composition in real time; segments showing under-response receive additional outreach or a different contact mode. This approach boosts coverage and mitigates nonresponse bias but also increases logistical complexity, costs, and the need for sophisticated data integration and weighting procedures (Kennedy, Popky & Keeter, 2023).

4.1.5 River Sampling

River sampling—defined as the real-time interception of anonymous web users through banners, pop-ups, or embedded in-article prompts—has emerged as a highly innovative response to persistent coverage gaps that contribute to modern polling errors. Murray-Watters, Zins, Silber, Gummer, and Lechner (2023) position this approach at the most "organic" end of the non-probability sampling spectrum: respondents enter the survey flow without email invitations, prior panel membership, or monetary incentives. By casting a wide digital net, river sampling can rapidly accumulate thousands of completions even among populations notoriously difficult to reach via traditional frames, such as mobile-only young adults, transient renters, or multilingual browsing communities.

In a German case study, the authors achieved a target of 5,000 interviews in under forty-eight hours—less than one-third of the time and cost required by an address-based probability panel. This speed equips election forecasters with the ability to track late swings in voter sentiment or monitor local campaign shocks typically missed by telephone or mail methods.

While critics have long warned that self-selection inflates bias in open-link surveys (Baker et al., 2013; Couper, 2017), Murray-Watters et al. acknowledge these concerns but demonstrate that river sampling's main shortcomings are not in measurement quality—break-off and straight-lining rates were comparable to opt-in panels—but rather in representativeness. Their river sample over-represented heavy internet users and college-educated adults, while under-representing older and lower-income groups. However, after applying multilevel regression with post-stratification (MRP) and propensity-score weighting—methods standard in high-quality non-probability research—the residual bias for key vote-intention variables fell to within two percentage points of benchmark probability samples. This finding echoes Callegaro and Yang's (2020) conclusion that rich auxiliary data can meaningfully reduce non-probability error, though rarely eliminate it entirely.

Where river sampling particularly excels is as a complement rather than a stand-alone substitute for probability surveys. Murray-Watters et al. propose a hybrid “anchor-and-calibrate” model: a small, well-designed probability subsample is fielded first, supplying ground-truth margins for age, gender, region, and past vote. A concurrent river inflow then expands case counts dramatically, providing granular detail on subgroups and real-time trend detection. Statistical techniques—such as Bayesian benchmarking, dynamic post-stratification, and Jackknife Variance Estimation—integrate the two data sources, yielding estimates that approach probability accuracy while preserving river sampling’s speed and cost advantages. Tourangeau et al.’s (2017) cautionary note about “professional respondents” is addressed by the river sample’s widely distributed recruitment network, which lowers the risk that a few frequent survey-takers dominate the data.

Finally, practical considerations are crucial for deploying river sampling effectively. Questionnaire length must be tightly controlled, as longer versions saw surges in break-offs. Landing page design should emphasize legitimacy to mitigate curiosity-driven clicks. Ethical transparency is paramount: respondents must be informed of the voluntary and anonymous nature of participation, and researchers should disclose weighting variables and design effects in accordance with AAPOR’s latest transparency standards. When these conditions are met, river sampling offers pollsters a nimble instrument to fill demographic blind spots, capture late-cycle volatility, and stress-test probability estimates—thereby reducing the likelihood and magnitude of future polling misses.

4.2 Model-Based approaches for more accurate predictions.

4.2.1 Multilevel Regression and Post-Stratification (MRP)

Multilevel Regression and Post-Stratification (MRP) is a powerful statistical methodology that addresses key limitations of traditional polling techniques by combining detailed hierarchical modeling with population-level weighting to produce accurate and consistent forecasts of public opinion and electoral outcomes. The approach, as explicated by Lauderdale et al. (2019), involves decomposing the electoral prediction problem into three fundamental components: the conditional voting distribution (the probability that an individual with certain demographic and political characteristics votes for a particular option), the conditional turnout distribution (the probability that this individual participates in voting), and the overall population distribution of voter types (the demographic makeup of the electorate).

Operationally, MRP begins with the estimation of a multilevel regression model that captures how voting behavior varies with individual-level predictors (age, gender, education, region, past voting behavior) and contextual factors. This hierarchical model pools information across groups to generate stable predictions even in sparsely sampled subpopulations. Subsequently, these modeled probabilities are weighted using post-stratification based on external population data (such as census statistics or voter registries) that describe the joint distribution of demographics across relevant geographic units, allowing researchers to aggregate individual-level predictions into accurate estimates of vote intention at both national and subnational levels, such as electoral districts or states.

This method contrasts with conventional polling approaches which typically apply weighting adjustments directly to survey respondents without explicit modeling of interactions or turnout heterogeneity. By modeling outcomes first and then aggregating according to known population structures, MRP provides more internally consistent and granular predictions, better capturing the heterogeneity across demographic and geographic strata (Lauderdale et al., 2019).

Empirical applications of MRP have demonstrated its superior performance in major elections, including the 2016 UK EU referendum, the 2016 US presidential election, and the 2017 UK general election. In these cases, MRP-based forecasts exhibited lower error metrics, such as Root Mean Squared Error, and stronger correlations with actual results compared to traditional weighted polls. Importantly, MRP's capacity to produce robust subnational estimates addresses a core challenge in electoral forecasting, since political outcomes are often determined by local contests rather than national aggregates alone.

Despite its strengths, Lauderdale et al. (2019) acknowledge methodological challenges such as potential attenuation bias—whereby predicted vote shares tend to understate highly polarized outcomes—and difficulties in modeling turnout dynamics that may fluctuate markedly between elections. Furthermore, the accuracy of MRP is contingent upon the availability and quality of population-level data used for post-stratification and the correctness of the hierarchical modeling assumptions.

Building upon this foundation, Broniecki, Leemann, and Wüest (autoMrP) introduce an enhanced MRP variant that integrates machine learning techniques to automate and improve the selection of context-level variables, functional form specification, and regularization, collectively addressing key modeling complexities. Traditional MRP involves subjectively choosing covariates and specifying linear or additive models, which might inadequately capture nonlinearities and complex interactions inherent in electoral behavior.

The autoMrP approach employs an ensemble of supervised machine learning classifiers—including Lasso regression for feature selection and shrinkage, principal component analysis, gradient boosting machines, and support vector machines—combined through Bayesian model averaging. This data-driven framework flexibly captures complex spatial and demographic patterns without manual intervention or reliance on expert-specified variables. Applied to a large-scale U.S. survey dataset, autoMrP substantially reduces prediction error, outperforming standard MRP and individual classifiers by between 12% to 31% in mean squared error, indicating enhanced predictive precision.

Overall, autoMrP represents a significant advancement by automating core modeling decisions and enabling robust, scalable, and accurate small-area estimation in electoral and public opinion research (Broniecki et al., 2023). This methodological sophistication further strengthens the advantage of model-based approaches over traditional polling, offering greater adaptability to complex and evolving electorate structures.

In sum, MRP and its machine-learning-augmented extensions provide a rigorous and flexible framework for election forecasting. By explicitly modeling voting behavior and turnout conditional on detailed demographics and geographic context, and by integrating rich auxiliary data, these methods resolve key shortcomings of conventional weighted polling. They produce more internally consistent, localized, and accurate electoral predictions, thereby reducing risks of polling misses and advancing the reliability of pre-election polling science.

4.2.1 Bayesian Model

This section introduces an alternative model-based approach that is widely used at both national and state levels to improve the accuracy of election outcome forecasts: Bayesian modeling of pre-election polls. Many researchers have developed Bayesian frameworks that leverage polling data, often combined with political and contextual information, to produce more precise predictions.

It is important to note that a common trait among many forecasting methods is that their primary unit of analysis, and hence the level at which inferences are made, typically lies at the aggregate state or national level. This macroscopic focus, which is often deliberate in structuralist models, allows researchers to avoid direct reliance on survey data that may introduce various forms of estimation error into predictive models. However, methods operating primarily at this macro-level remain vulnerable to the ecological fallacy — the mistaken inference about individual-level behavior from aggregated data — particularly when trying to predict outcomes driven by individual voter decisions based solely on aggregate variables such as national economic indicators (Kramer, 1983). This limitation therefore underscores the analytical value of survey-based models, which have undergone considerable refinement in recent years. Among these advancements are strategies aimed at addressing nonresponse bias and systematic measurement error, frequently through the incorporation of reported past voting behavior and detailed sociodemographic controls.

Survey-based approaches, however, face their own challenges. Growing political polarization and social phenomena such as the spiral-of-silence effect impose significant obstacles to accurately capturing voter intent. The spiral-of-silence suggests that some individuals may conceal their true electoral preferences due to apprehension about social exclusion, especially when they believe their views to be in the minority within their social groups. Empirical research provides evidence of this dynamic: for example, Camatarri, Luartz, and Gallina (2023) demonstrate, using survey data alongside county-level election results from the 2020 U.S. presidential election, that Republican voters exhibit a notable tendency towards reticence. This pattern aligns with earlier findings reported by Dinas, Martínez, and Valentim (2024) and Urquizo-Sancho (2006), which suggest that right-leaning voters are less likely to disclose their political preferences in contexts perceived as hostile. By contrast, Democratic respondents tend to be more forthcoming, possibly due to higher levels of psychological openness — which correlates with a greater willingness to share personal opinions and participate in political discourse regardless of societal pressures (Gerber et al., 2010; Mutz, 2002). Importantly, Republican voters are not homogeneous in this regard: while some may refrain from voicing their preferences due to social pressures, others, particularly those with nonconformist tendencies, may feel freer to express dissenting opinions within their social networks (Kushin, Yamamoto, & Dalisay, 2019).

While traditional Bayesian forecasting methods anchored at national or state levels offer parsimonious, survey-free baselines, their macro-level focus risks oversimplification and the ecological fallacy. To overcome this, Chen, Garnett, and Montgomery (2023) propose a dynamic hierarchical Bayesian forecasting model that excels in capturing micro-level polling fluctuations while still drawing strength from broad structural anchors. Their model treats macro-level Bayesian estimates as informative hyper-priors within a Dirichlet–Gaussian-process framework. This allows high-frequency polling data to update, but not completely override, these structural fundamentals as Election Day approaches. In doing so, the model achieves a synthesis whereby structural elements provide stability when data are sparse, while hierarchical smoothing absorbs local shocks and volatility. This results in a unified, multi-resolution forecasting approach that is more resilient to typical polling errors.

At the core of Chen et al.'s (2023) model lies a hierarchical Dirichlet-regression layer which accounts for the compositional nature of vote shares by modeling Democratic, Republican, and "Other" vote percentages as a Dirichlet-distributed vector. The parameters of this layer are anchored on fundamental predictors such as incumbency and each state's Partisan Lean. On top of this, Gaussian process priors are imposed on latent support trajectories for each state-party combination, smoothing daily fluctuations and allowing states with sparse or early polling to borrow strength from states with abundant polling data through partial pooling.

Polling observations enter the model as noisy measurements of these underlying support trajectories. The model explicitly estimates house effects and sampling variance, while the fundamentals act as informative priors whose influence wanes as the election date nears. Because uncertainty propagates through each layer of the hierarchical structure, the model provides well-calibrated predictive intervals that remain reliable even when polling data are sparse or irregularly spaced. Retrospective evaluations of the 2018 and 2020 election cycles show that this combined framework of polls, contextual information, and temporal smoothing significantly reduces average state-level forecast errors, often reaching accuracy comparable to high-quality probability samples. This approach offers a practical and statistically principled means of mitigating polling misses.

4.2.3 Leveraging Twitter-Derived Signals in a Polling Time-Series Framework

Recent research by Scarano et al. (2024) provides a critical examination of election-related polls conducted on social media platforms, particularly Twitter, during the 2016 and 2020 U.S. presidential election campaigns. Their extensive analysis of nearly two thousand Twitter polls reveals significant prevalence patterns and notable biases that directly affect the reliability and interpretability of polling data derived from social media.

Scarano and colleagues document that election polls on Twitter are predominantly authored and participated in by older male users, indicating a demographic skew that raises questions about representativeness. Moreover, their findings show that Twitter polls demonstrate a pronounced bias favoring former President Donald Trump, contrasting with traditional mainstream surveys, which tended to display a relative bias toward Democratic candidates. For example, in 2020, the median support for Trump in Twitter polls was approximately 17 percentage points higher than in mainstream polls, which themselves slightly underestimated his support compared to actual election outcomes. These discrepancies underscore the systematic nature of biases inherent in social-media polling (Scarano et al., 2024).

Their study further investigates sources of bias, identifying inauthentic behaviors including bot participation and vote manipulation techniques. They provide evidence suggestive of purchased votes influencing poll outcomes, as well as inconsistencies between publicly visible vote counts and those accessible solely to poll authors. Additionally, Twitter accounts engaged in election polls were more likely to be automated or semi-automated bots, raising serious concerns about the validity of social media polling as an accurate reflection of public opinion (Scarano et al., 2024).

These findings have direct implications for the integration of social-media polling data into forecasting frameworks. While social media offers the advantages of immediacy and high frequency, its demographic distortions and vulnerability to manipulation necessitate sophisticated correction methods. Techniques such as demographic inference, bias-correction terms, and the application of structured priors within multilevel regression and post-stratification (MRP) models—as employed in PoSSUM and other hybrid approaches—can help mitigate these distortions and render social-media-derived data more informative and reliable (Cerina, 2025; Cerina & Duch, 2023).

Understanding the specific mechanisms of bias and distortion in social-media polls, as detailed by Scarano et al., is essential for improving their incorporation into contemporary electoral forecasting pipelines. This underscores the importance of transparency, robust bot detection, and careful sample weighting when interpreting social media signals, reaffirming that although socially sourced polls hold promise, they must be treated with methodological rigor and caution.

This methodology reframes traditional opinion polls as merely one input among several within a dynamic forecasting system. Rather than relying exclusively on survey data, it continuously harvests streams of publicly available tweets containing party- or candidate-related keywords via the Twitter API, creating a high-frequency record of political discussion.

The model collects basic Twitter activity metrics at regular intervals—such as daily or weekly counts—including the frequency of mentions for each party or candidate. It applies a simple lexicon-based sentiment analysis to assess whether these mentions are generally positive or negative, and examines retweet and reply counts to gauge the velocity and reach of each message.

Concurrently, the model incorporates established polling figures by collecting vote-intention percentages, either two-party or multi-party, from reputable survey organizations. By pairing these poll results with the corresponding Twitter metrics at each time point, the system creates synchronized feature vectors blending conventional survey outcomes with real-time social-media signals of public opinion.

These feature vectors feed into supervised machine learning algorithms—such as linear regression, support-vector machines, or decision trees—to predict short-term shifts in voting intentions. During model training, the poll-based time series serve as ground truth, while Twitter-derived features provide complementary, high-frequency indicators capable of detecting emergent trends that standard surveys might miss.

By continually updating the model with incoming data, this approach adapts to late-campaign developments and reduces reliance on outdated poll samples. In effect, it delivers a blended forecast that combines the statistical rigor of traditional polling with the immediacy of social-media analytics—thereby lowering the risk of “polling misses” during rapid shifts in public sentiment.

4.2.4 Leveraging Large Language Models to Mitigate Polling Misses

In recent years, the advent of large language models (LLMs) has opened transformative possibilities for social science research, particularly in the realm of election polling. The PoSSUM protocol harnesses these state-of-the-art computational tools by embedding LLMs within an entirely unsupervised analytical pipeline, ingeniously inferring detailed demographic and attitudinal profiles from the rich tapestry of social-media content (Cerina & Duch, 2023). This novel approach transcends traditional survey limitations by transforming vast, noisy, and inherently biased social-media footprints into structured and analytically tractable strata, which seamlessly feed into the robust multilevel regression and post-stratification (MRP) framework.

Crucially, PoSSUM integrates cutting-edge methodological components including structured priors and a bias-correction term originally formulated by King and Zeng (2001; as cited in Cerina & Duch, 2023), enabling it to address head-on the severe selection biases plaguing social media samples. Notably, it excels at correcting for the prevalent over-representation of highly engaged users—a notorious source of distortion—achieving predictive performance that rivals that of carefully constructed probability samples even under extreme nonrandom respondent selection.

This leap in methodological sophistication has been rendered feasible by rapid advances in natural language processing. Variables that have traditionally eluded direct measurement—such as age cohort, gender, educational background, and political orientation—can now be inferred with remarkable accuracy from linguistic signals and multimodal content. Such automatic demographic "imputation" effectively fills critical data gaps, converting sprawling, biased online populations into proxy samples whose composition closely mirrors that of the true electorate (Cerina & Duch, 2023). By harnessing diverse digital footprints—including textual posts, images, and metadata—PoSSUM creates what can be considered digital "silicon samples," affording a richer, near real-time window into public opinion than traditional polling machinery typically permits (Cerina, 2025).

PoSSUM's operational pipeline unfolds in several distinct but interconnected stages. It begins with the strategic assembly of a digital cohort drawn from social media platforms through topical, politically relevant search criteria, designed to capture a broad swath of politically engaged users. Detailed digital "moulds" are then created for individuals, compiling recent posts, reactions, and network metadata. These comprehensive profiles are analyzed using multimodal LLMs to extract survey-equivalent responses, which form the backbone of the structured strata. To counterbalance selection bias inherent in social media participation, PoSSUM implements quota sampling guided by MRP-informed stratification frames—carefully tailored to socio-demographic and behavioral heterogeneity revealed through LLM analyses. Continual model recalibration ensures that dynamic shifts in online behavior and opinion are promptly incorporated, delivering temporally sensitive and granular snapshots of the evolving electoral landscape (Cerina, 2025; Cerina & Duch, 2023).

The profound scalability and instantaneity of this approach confer striking advantages relative to conventional polling practices, which are typically episodic and resource-intensive. By enabling the processing of millions of posts in near real-time, PoSSUM facilitates continuous, high-frequency polling updates at a fraction of the typical cost and latency. This rapid responsiveness allows researchers to detect emerging polling discrepancies early and adjust weighting accordingly, preventing entrenched biases from distorting critical decision-making windows.

Empirical validation during the 2024 U.S. Presidential Election confirmed PoSSUM's capacity to generate detailed, state-level vote-share forecasts that closely approximated eventual outcomes, outperforming many traditional survey-based estimators (Cerina, 2025). Such promising results underscore the value of integrating large-language models with advanced statistical techniques like MRP, melding the speed, scale, and richness of social-media analytics with the rigor and representativeness foundational to election forecasting.

Despite the remarkable advances embodied by PoSSUM and similar LLM-empowered polling protocols, it is imperative to acknowledge inherent challenges related to the nature of social-media data. Scarano et al. (2025) provide a comprehensive examination of election polls conducted directly on social media platforms, revealing significant demographic skews—particularly an over-representation of older male users—and systematic biases notably favoring specific political actors, such as the substantial Republican tilt observed in Twitter polls during the 2016 and 2020 U.S. elections.

Moreover, their analysis highlights vulnerabilities including widespread bot activity, vote manipulation attempts, and discrepancies between publicly reported and privately accessible polling data, which collectively threaten the validity and credibility of social-media-based surveys. These findings underscore that, while social media offers unprecedented immediacy and volume of data, careful correction methods—like the bias-correction terms and structured demographic inference central to PoSSUM's design—are essential to mitigate distortions. Further efforts in bot detection, authentication protocols, and transparent methodological reporting are critical to enhance the reliability of such hybrid polling systems.

In sum, the augmentation of opinion polling through LLM-powered protocols like PoSSUM constitutes a formidable advance against the perennial problem of polling misses. By automating the extraction and annotation of demographic and attitudinal information, dynamically correcting nonrandom selection bias, and supporting near real-time monitoring of public sentiment shifts, these methods meld the best qualities of probability sampling with modern digital data abundance—equipping researchers and practitioners alike with an unprecedented toolkit to navigate today's volatile and fragmented electoral terrain. Continued interdisciplinary research and methodological innovation remain crucial to fully realize the potential of social-media-informed election forecasting without compromising representativeness or accuracy.

While methodological refinements and innovative sampling or modeling strategies are central to addressing polling inaccuracies, their practical effectiveness can only be fully appreciated when tested against real-world electoral contexts. This is why the next chapter turns to case studies drawn from a range of elections across different countries and time periods. By examining these concrete examples, it becomes possible to observe how theoretical challenges such as coverage error, nonresponse bias, or late voter shifts materialize in practice, and how alternative methodological solutions have succeeded—or failed—in mitigating them. Case studies thus serve a dual purpose: they illustrate the mechanisms through which polling errors unfold under diverse political and social conditions, and they provide critical insights into the conditions under which corrective strategies can be most effectively deployed. In this way, they bridge the gap between abstract methodological discussions and the complex realities of electoral forecasting.

Chapter 5: Case Studies

5.1 1936 U.S Presidential Elections

To explore the reasons behind the 1936 *Literary Digest* polling failure, Squire (1988) examined two main hypotheses that had been proposed to explain the significant discrepancy between the poll's prediction and the actual election outcome. The first hypothesis, initially suggested by pollsters such as Crossley (1937:27) and Gosnell (1937), focused on response bias, arguing that the voters who chose to return the poll's questionnaires were disproportionately inclined toward the Republican candidate, Alf Landon. The second hypothesis, proposed by Katz and Cantril (1937:167–168), highlighted sampling bias, emphasizing the fundamental flaw of the poll's sampling frame, which predominantly included wealthier voters identified from automobile registrations, telephone directories, and magazine subscription lists, systematically omitting segments of the population more supportive of Roosevelt.

To empirically assess these hypotheses, Squire (1988) undertook a secondary analysis of a 1937 Gallup survey dataset. Gallup's study was uniquely designed to capture both respondents and non-respondents of the original *Literary Digest* poll, providing a direct means to evaluate differences between these two groups. Specifically, Gallup had gathered responses regarding vote preferences, voting behavior in the 1936 presidential election, and whether respondents had returned the *Literary Digest* ballot. By comparing the voting behavior and political preferences between those who participated in the *Literary Digest* poll and those who did not, Squire was able to isolate whether nonresponse significantly influenced the poll's results or whether the poll's sampling framework itself introduced systematic bias.

The findings of Squire's analysis indicated clear support for both hypotheses. Nonresponse bias was indeed significant, as those who returned ballots to the *Literary Digest* poll were more likely to favor Landon, thus inflating Republican support in the poll. However, Squire's research also reinforced the importance of sampling bias: the pool of potential respondents was itself unrepresentative of the voting population, largely excluding lower-income and rural voters who strongly supported Roosevelt. Ultimately, Squire concluded that the 1936 polling failure resulted from the combination of these two biases—response and sampling—highlighting the crucial importance of representative sampling methods and rigorous respondent follow-up strategies in public opinion research (Squire, 1988).

The significance of the sampling bias becomes evident when examining **Table 5.1.1** from Squire's (1988) analysis. The table breaks down vote choice by ownership of a car and a telephone, the very criteria used in the *Literary Digest* sampling frame. The results show that Roosevelt received overwhelming support among individuals with neither a car nor a telephone (79 percent), while his support dropped substantially among those with both (55 percent). In contrast, Landon's support was much higher among the more affluent group with cars and telephones, reaching 45 percent. These findings illustrate how the Digest's reliance on such lists systematically favored wealthier citizens and excluded lower-income voters who were decisive in Roosevelt's landslide victory.

Presidential Vote	Car& Phone	Car, No Phone	Phone, No Car	Neither
Roosevelt	55	68	69	79
Landon	45	30	30	19
Other	1	2	0	2
Total *N*	946	447	236	657

SouRcE: American Institute of Public Opinion, 28 May 1937.

Table 5.1.1. 1936 Presidential Vote by Car and Telephone Ownership (in Percent) (as cited in Squire 1988)

Equally important is the evidence presented in **Table 5.1.2**, which highlights the effect of response bias. Those who returned the Digest ballot reported stronger support for Landon (51 percent) compared to Roosevelt (48 percent). Among those who did not return the ballot, however, Roosevelt was the overwhelming favorite with 69 percent support. This discrepancy reveals that non-respondents were disproportionately pro-Roosevelt, meaning that even within the flawed sampling frame, the act of responding was itself skewed. As Squire (1988) concludes, the Digest failure was not the result of just one flaw but of the combined effects of an unrepresentative sample and biased patterns of response.

Presidential Vote	Did Return	Did Not Return	Do Not Know
Roosevelt	48	69	56
Landon	51	30	40
Other	1	1	4
Total *N*	493	288	48

SouRcE: American Institute of Public Opinion, 28 May 1937.

Table 5.1.2. Presidential Vote by Returning or Not Returning Straw Vote Ballot (as cited in Squire 1988)

5.2 U.S National Elections 2016

This paragraph draws upon the landmark investigation led by a distinguished ad hoc committee of the American Association for Public Opinion Research (AAPOR), convened in the aftermath of the 2016 U.S. presidential election to address unprecedented public and scholarly scrutiny regarding polling inaccuracies. Guided by the influential work of Kennedy et al. (2018), the committee undertook a rigorous, multidimensional inquiry into both the anatomy and causes of polling errors—combining quantitative depth with rare breadth—and produced one of the most authoritative evaluations in the field to date.

 https://doi.org/10.26219/heal.aueb.9743

Central to the committee's methodological framework was the comprehensive aggregation of polling data, meticulously gathered from an extensive spectrum of national and state-level polls employing diverse sampling modalities. These included live telephone interviews, interactive voice response (IVR) automated surveys, and opt-in online panels, which were not merely collected but systematically scrutinized to uncover not only the "what" of polling outcomes but the "why" behind variations in accuracy across survey instruments, demographic groups, and political contexts. This expansive dataset offered unprecedented granularity, enabling the researchers to illuminate subtle yet consequential discrepancies often obscured by national aggregates.

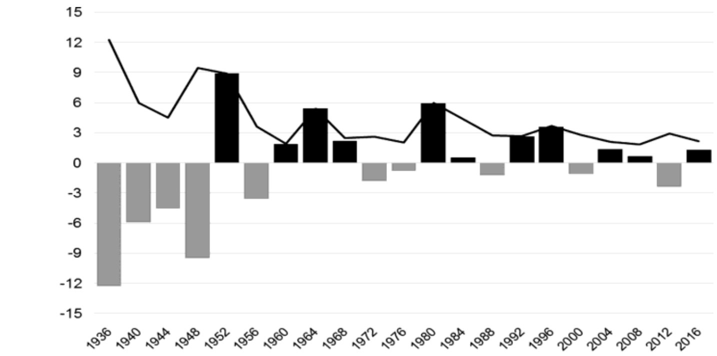
The analytical process involved an unflinching comparison between pre-election poll forecasts and certified election results, probing questions not only of "how far off" projections were, but crucially "where," "why," and "under what methodological or contextual circumstances" these errors became most pronounced. By disaggregating results at the level of key battleground states, the committee captured with forensic precision the dynamics that ultimately tipped the electoral scales—a task essential for meaningful post-election reflection and learning.

Beyond surface-level measurement, the committee employed advanced quantitative and qualitative techniques to investigate drivers of inaccuracy. Special emphasis was placed on the persistent challenge of obtaining representative samples—most notably, the chronic underweighting of voters without college degrees, whose decisive impact in 2016 was substantially underestimated due to incomplete demographic stratification. Further areas of inquiry included the predictive shortcomings of likely voter models, distortions introduced by differing survey modes, and the frequently hypothesized (but empirically elusive) effect of social desirability or "shy voter" bias on candidate preference reporting.

Crucially, the committee situated its findings within a historical framework, placing the surprises of 2016 in the context of nearly a century of American electoral polling. In so doing, it offered both a sobering reminder of the recurring complexities inherent in measuring mass opinion and a hopeful roadmap toward scientific progress—characterized by enhanced methodological rigor, reflexive adaptation, and increased institutional transparency.

Overall, this research advanced the understanding of polling error in an election of historic consequence and elevated the standards for critical self-assessment and methodological innovation throughout the discipline. By marrying statistical precision with analytical depth, the committee's work stands as an exemplar of scholarly inquiry committed to intellectual honesty and civic responsibility—embodying the kind of science capable of transforming challenges into enduring improvements.

Figure 5.2.1 presents a historical overview of polling error in U.S. presidential elections from 1936 to 2016, charting both the average absolute error (ABS error) and the signed error of national polls. The black bars represent signed error (indicating directionality—positive values reflect Republican underestimation, negative values Democratic), while gray bars show the unsigned magnitude of error. The line across the top indicates the number of polls included per year. The figure reveals that although recent elections have exhibited modest polling error, the 2016 election showed a relatively small average absolute error (2.2 percentage points) compared to some earlier contests, such as 1948 or 1952. However, the signed error in 2016 (+1.3) reveals a consistent underestimation of Republican support. The data suggest that while polling inaccuracies are not new, 2016 stands out for the directional bias favoring Democrats, rather than for extreme overall error magnitude.



	1936	1940	1944	1948	1952	1956	1960	1964	1968	1972	1976	1980	1984	1988	1992	1996	2000	2004	2008	2012	2016
Polls	1	1	1	1	1	1	1	2	2	3	3	4	6	5	6	9	10	16	19	25	22
ABS error	12.3	6.0	4.5	9.5	8.9	3.6	1.9	5.4	2.5	2.6	2.0	6.0	4.4	2.7	2.7	3.7	2.8	2.1	1.8	2.9	2.2
Signed error	-12.3	-6.0	-4.5	-9.5	8.9	-3.6	1.9	5.4	2.2	-1.8	-0.8	6.0	0.5	-1.2	2.7	3.6	-1.1	1.4	0.7	-2.4	1.3

Figure 5.2.1. Average error in vote margin in national presidential polls, 1936–2016 (Kennedy et al, 2018).

Figure 5.2.2 compares the average absolute error for various polling methodologies used during the 2016 U.S. general election. The bar graph distinguishes between national-level and battleground-state polls, using different shading or coloring, and categorizes polls by their mode and sample design—for instance, Interactive Voice Response (IVR) only, live telephone (random-digit-dial or registration-based), internet (opt-in), and combinations thereof.

The data reveal subtle yet critical differences in the precision of each survey approach. IVR-only polls, which utilize automated calls primarily to landlines, achieved the lowest average absolute error among battleground-state polls (2.8 points), despite potential concerns about their limited coverage of younger and more mobile populations. By contrast, opt-in internet polls and live telephone polls using random-digit dialing showed somewhat higher errors, around 3.8–3.9 points for battleground states. National-level polls using blended methods—such as IVR combined with online sampling—posted even lower average errors in some cases, reflecting the potential benefits of methodological hybridization.

This figure illuminates a key insight: polling accuracy is not solely a function of technological modernity or traditionalism, but depends intricately on the match between survey design, population behavior, and electoral context. The 2016 results suggest certain modes perform surprisingly well under specific conditions, challenging assumptions about the universal superiority (or inferiority) of any one approach.

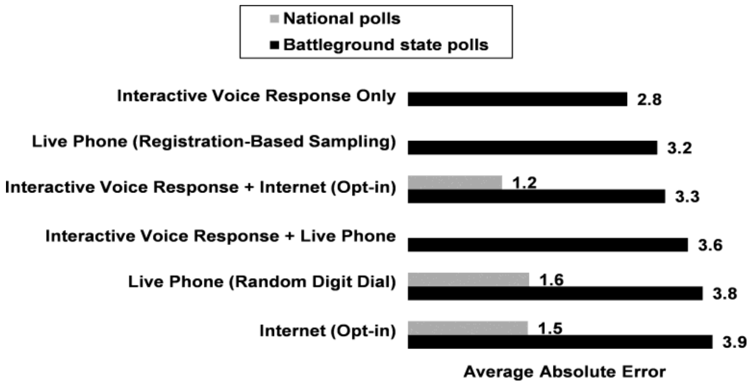


Figure 5.2.2. Average absolute error for 2016 general election polls, by design (Kennedy et al, 2018).

As a direct consequence of these findings, the polling industry undertook significant adaptations in the training and operationalization of polling methodologies for subsequent election cycles. Survey organizations decisively revised their sampling and weighting procedures—most notably by incorporating education as a core stratification variable and systematically adjusting weights to correct for known demographic imbalances. This shift is vividly illustrated in Figure 5.2.3, which depicts the evolution of education weighting practices across major polling organizations before and after the 2016 election. The figure highlights a notable increase in the proportion of polls implementing explicit education-based weighting following 2016, reflecting an industry-wide acknowledgment of the methodological shortcomings identified by the AAPOR committee.

Furthermore, **Figure 5.2.3** provides tangible evidence that such methodological recalibrations produced measurable reductions in systematic polling error, particularly in states with historically high proportions of non-college-educated voters. The diagram not only charts the rise in the adoption of best practices but also stands as a visual testament to the polling community's swift and evidence-based response to empirical deficiencies. These enhancements were complemented by increased emphasis on methodological transparency and advanced training for polling professionals, further strengthening the resilience of the field. Ultimately, the adjustments captured in Figure 5.2.3 encapsulate how methodological learning—driven by critical post-election analysis—can be translated into concrete improvements in both the credibility and accuracy of contemporary electoral polling.

Notably, the horizontal axis of **Figure 5.2.3** in Kennedy (2018) provides a clear comparative framework, labeling three distinct data points for each state: the poll estimate before adjusting for education, the estimate after educational weighting, and the actual certified vote margin. By presenting these values sequentially along the horizontal dimension, the figure enables a direct visual assessment of how educational weighting narrows the gap between pre-election survey predictions and electoral realities. This structure reinforces the critical role that post-stratification adjustments played in enhancing polling accuracy and underscores the methodological evolution underpinned by empirical analysis.

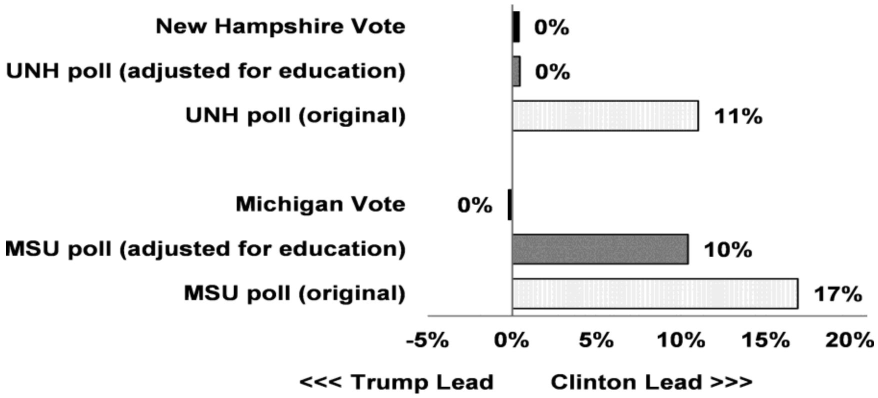


Figure 5.2.3. Poll estimates with and without weighting adjustment for education, relative to 2016 presidential vote outcome (Kennedy et al, 2018).

<u>Source</u>: University of New Hampshire poll conducted November 3–6, 2016, with 707 likely voters. Michigan State University poll conducted September 1–October 30, 2016, with 743 likely voters.

5.3 U.K Presidential Elections 2015

The investigation conducted by Sturgis et al. (2017) was initiated following the pronounced polling errors in the lead-up to the 2015 UK general election, where a near-universal underestimation of the Conservative Party's vote share occurred. This systematic miss, averaging approximately seven percentage points, ignited widespread concern about the integrity and validity of opinion polling in British electoral politics.

Sturgis and colleagues assembled and analyzed an extensive collection of polling microdata derived from multiple national-level polls. Their approach entailed a detailed decomposition of the total forecast error, aiming to isolate and quantify the respective contributions of underlying methodological and contextual factors. Central among these were unrepresentative sampling frames, weighting and turnout modeling issues, late voter swing, questionnaire design effects, mode of interview, and the possibility of respondent misreporting.

The research compellingly demonstrated that the paramount cause of the polling inaccuracy was unrepresentative sampling. Specifically, the recruited samples disproportionately overrepresented Labour supporters while underrepresenting Conservative voters, a bias that standard weighting and post-stratification adjustments failed to fully correct (Sturgis et al., 2017). Contrary to some media narratives, the study found limited evidence that last-minute changes in voter preferences or turnout weighting errors substantially contributed to the discrepancy. At most, late swing accounted for around one percentage point of error, while effects related to survey mode or question wording proved negligible.

The phenomenon of "herding" also emerged as a contributing factor, with pollsters tending towards similar sampling and weighting strategies, which possibly amplified the collective error through convergence on shared flawed assumptions (Sturgis et al., 2017). Furthermore, ancillary elements such as the treatment of postal or overseas voters, and misclassification of voter intention, were found to exert minimal impact on the overall accuracy.

This seminal study culminated in a series of twelve methodological recommendations aimed at overhauling polling practice. These recommendations emphasized the imperative of improving sample representativeness, notably through refinement of sampling frames and recruitment methods, and the enhancement of weighting procedures to better align samples demographically with the electorate. By doing so, Sturgis et al. (2017) not only clarified the multifaceted origins of the 2015 polling failure but also charted a path for methodological reforms designed to restore confidence and accuracy in future electoral polling.

To support the analysis, several types of visualizations are particularly informative. A stacked bar chart decomposing the total forecast error into components such as sample composition error, weighting adjustment error, and likely voter model error vividly illustrates that sample composition represented the majority share of polling bias. Similarly, comparative graphs contrasting poll predictions with actual election outcomes highlight the systematic underestimation of Conservative support. Visualizations depicting the disparity between the demographic composition of poll samples versus actual electorates further expose representativeness gaps, underscoring the central sampling issue identified by the research. Campaign period timelines demonstrate the convergence of poll estimates across organizations ("herding"), while error analysis by polling mode provides insight into how survey administration methods influence accuracy.

Collectively, these findings underscore the critical need for continued methodological vigilance and innovation in electoral polling, with sampling design at the forefront of priorities. The work by Sturgis et al. (2017) remains a foundational reference for scholars and practitioners seeking to understand and rectify polling errors in complex democratic environments.

5.4 Quebec Elections 2018

Bodet, Laflamme, Brie, and Ouellet (2024) conducted an evaluation of polling accuracy during the 2018 Quebec elections through a streamlined exit poll carried out in twelve ridings. At each site, interviewers approached every fourth voter leaving the polls with a short, anonymous questionnaire, yielding 540 fully vetted interviews after eligibility and quality checks. Importantly, the ridings and polling stations were deliberately stratified to mirror key cleavages in Quebec's electorate, including linguistic divisions between francophones and anglophones, urban-rural contrasts, and partisan balance. This stratification was vital, as exit polls rely on relatively small samples collected under tight time constraints, and without careful design a systematic "nth-voter" rule would risk over- or under-sampling decisive subgroups. After fieldwork, demographic and regional weights were applied to align the sample with census benchmarks and turnout figures, and weighted results were then compared against official outcomes. This approach demonstrated that even a modest, well-stratified exit poll could produce a considerably more accurate picture of party support than pre-election surveys.

To further probe the sources of inaccuracy, the authors estimated a series of ordinary least squares models using their microdata. In the first two models, the outcome was the signed vote-share error for each party, with a dummy variable for right-of-centre parties and controls for official vote share and riding fixed effects. Both specifications revealed a systematic underestimation of around ten percentage points for right-leaning parties, statistically significant at the 95 percent level and nearly matching the Root Mean Square Error (RMSE) of the sample (**Table 5.4.3**). This confirmed the hypothesis that ideology can influence willingness to disclose vote choice even in onsite surveys. The analysis then shifted from bias to precision. Model 3 regressed absolute error on the raw number of respondents captured for each party, showing that every nine additional interviews reduced the error by about one percentage point. By contrast, Model 4 substituted proportional representation in the sample for raw counts, but the relationship disappeared, reinforcing the conclusion that sheer interview volume, rather than proportionality, drives exit poll precision.

Accuracy was evaluated with two complementary metrics: the Mean Absolute Error (MAE), capturing average deviations between estimates and certified vote shares, and the RMSE, which weighted larger misses more heavily to

https://doi.org/10.26219/heal.aueb.9743

reflect "typical" errors. Both measures showed improvements compared to pre-election polling, but also revealed persistent biases and pockets of imprecision. Taken together, the findings highlight two practical lessons for mitigating polling misses: first, oversampling right-of-centre voters to counter chronic underestimation, and second, prioritising larger absolute sample sizes in low-turnout ridings to enhance precision.

This pair of scatterplots (**Figure 5.4.1**) visualizes the relationship between polling error (E) and the representation of each party within exit poll samples from the 2018 Quebec election. The left panel relates error to the absolute number of party supporters sampled, while the right focuses on the percentage these voters represent within the total sample. Black dots represent right-leaning parties, which, as the plots illustrate, are almost never heavily overestimated. Most significant overestimations are seen among other parties, especially at low sample numbers or proportions. Undercounting, particularly for right-leaning parties, tends to cluster at the lower end of sample size, demonstrating a persistent bias that cannot be explained by proportional representation alone. This pattern underscores the importance of oversampling certain groups, like right-leaning voters, in order to reduce systematic error and avoid misrepresenting electoral support in small or demographically skewed samples.

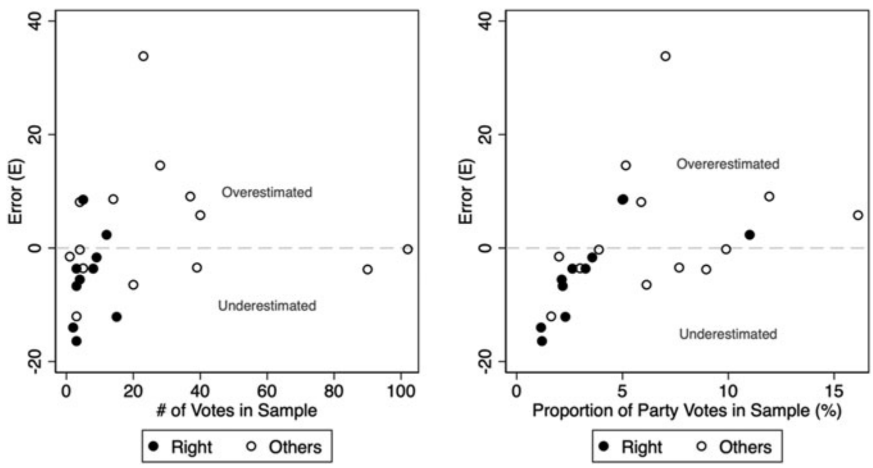


Figure 5.4.1. Over- and under-estimation (E) of parties' vote share (Canadian Journal of Political Science, Marc André Bodet et al 2024).

This boxplot (**Figure 5.4.2**) compares the magnitude and dispersion of absolute polling errors by party/district, revealing notable disparities in exit poll reliability. The figure shows that parties such as TSC and LD experienced higher median errors and greater dispersion—frequently exceeding 10 points—compared to consistently lower errors for groups like LH and QC. The visual summary highlights not only the skewness and outliers present for certain parties but also the substantial variation in estimate quality that can arise even within the context of a carefully planned, stratified exit poll. Importantly, the differences captured here reflect both local turnout dynamics and the challenges of capturing accurate estimates for smaller or politically marginal parties, reinforcing the need for larger samples or targeted oversampling strategies in these cases.

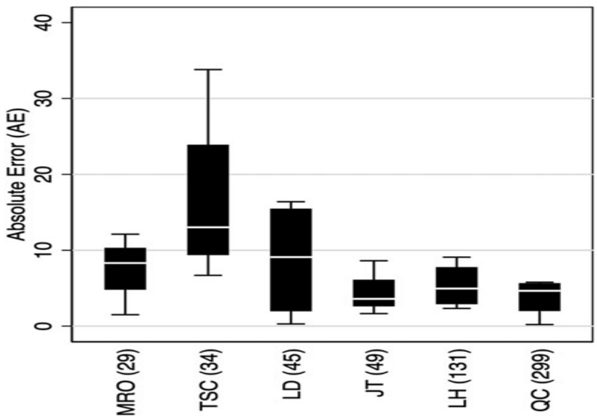


Figure 5.4.2. Distribution of Absolute Error (AE) per district (Canadian Journal of Political Science, Marc André Bodet et al 2024).

The table below (**Table 5.4.3**) is a regression results table that summarizes model estimates for both error (E) and absolute error (AE) as dependent variables, measured across four different models. The table presents the estimated coefficients for several covariates: the number and percentage of votes for a party in the sample, whether the party is classified as "Right," and the district election results. Standard errors are reported in parentheses below each coefficient, and asterisks denote statistical significance. The models show that "Right Parties" are associated with statistically significant negative errors in the first two models, indicating these parties tend to be underestimated in exit polling. The number of votes is marginally significant in Model 2, suggesting a relationship between larger sample sizes and reduced error. The intercepts, sample size (N), R-squared values, and root mean squared error (RMSE) indicators provide additional detail about overall model fit and explanatory power. This table provides quantitative evidence of systematic and party-based sources of error in Canadian exit polling, reinforcing findings that sample size and party type matter in predicting polling accuracy.

Covariates	Model 1 (E)	Model 2 (E)	Model 3 (AE)	Model 4 (AE)
# of votes	−0.03		−0.11*	
% of votes	(0.06)	−0.25	(0.05)	0.05
Right Parties	−9.64*	(0.34)	−3-39	(0.19)
		−9.82*		−0.44
	(4.47)	(4.30)	(3.00)	(2.66)
District Results	0.11	0.37	0.30	0.15
	(0.30)	(0.59)	(0.16)	(0.33)
Intercept	1.80	1.17	4.33*	2.95
	(5.05)	(4.75)	(1.99)	(1.93)
N	24	24	24	24
R2	18.65%	21.29%	27.38%	14.82 %
RMSE	10.34	10.17	6.58	7.13

Table 5.4.3. Regression analysis (OLS) to identify the sources of errors (Canadian Journal of Political Science, Marc André Bodet et al 2024).

Key Findings

The weighted exit poll proved markedly more accurate than the campaign-period surveys that had pointed to a hung parliament. Province-wide, the poll overstated Liberal support by just 0.8 percentage points and understated the Coalition Avenir Québec (CAQ) by

1.3 points, producing an MAE of 1.2—well within sampling tolerance and roughly one-third the average error recorded by pre-election telephone and online polls. Errors at the riding level were slightly larger, reflecting local turnout differentials, but still below the five-point threshold often deemed acceptable for seat-allocation modelling. Demographically, the exit poll captured francophone and non-college voters at rates almost identical to census benchmarks, though it modestly over-sampled university graduates. The authors conclude that carefully designed, probability-based exit polls remain a viable audit tool for Canadian elections, capable of validating or correcting pre-election expectations—provided they employ rigorous post-stratification and adequate geographical coverage.

5.5 U.S National Elections 2020

In his comprehensive study of the 2020 U.S. general election, Costas Panagopoulos undertook a detailed assessment of polling accuracy and bias across both the national and state levels. His analysis drew on 14 national surveys conducted in the final week before Election Day, along with more than 200 statewide polls covering presidential, Senate, gubernatorial, and congressional contests. To evaluate the precision of these polls, he applied a set of widely used statistical accuracy measures. Chief among them was the “A” index developed by Martin, Traugott, and Kennedy, which summarizes the average deviation between poll estimates and actual election outcomes, providing a general benchmark of overall accuracy. Complementing this, he employed two of Mosteller’s classic measures: M3, which captures the error in the estimated margin between the top two candidates—particularly crucial in closely contested races—and M5, which records the single largest error observed in a poll, highlighting the most consequential miscalculations. By combining these measures, Panagopoulos was able to distinguish between polls that were broadly accurate but missed on margins, and those that suffered from outlier errors with disproportionate effects. In addition, his approach incorporated multivariate regression analysis to examine how factors such as survey mode, sponsor affiliation, sample size, and fieldwork timing shaped patterns of accuracy and bias (Panagopoulos, 2021).

The **figure 5.5.1** presents a comparison between various national pre-election polls conducted by prominent polling firms and the actual results of the 2020 U.S. General Election. Specifically, it includes both the presidential race (between Trump and Biden) and the election for the U.S. House. Each poll's predicted vote share for Trump (Republican) and Biden (Democrat), sample size, and a calculated accuracy metric known as the Mosteller Measure are reported. The Mosteller Measure evaluates the deviation of poll predictions from actual election outcomes, with lower values indicating greater accuracy.

In the presidential race, the actual election result shows Trump obtaining 46.8% and Biden 51.3%. The polls varied in their accuracy, with IBD/TIPP and Reuters/Ipsos being notably accurate (Mosteller Measure of 1.05 and 1.25 respectively), closely matching actual results. Conversely, Quinnipiac exhibited the largest deviation, with a Mosteller Measure of 4.55. On average, presidential election polls had a Mosteller Measure of 2.19, indicating moderate deviations overall.

Regarding the U.S. House election, the actual vote shares were 47.7% Republican and 50.8% Democratic. Poll predictions for the House showed higher deviations, averaging a Mosteller Measure of 3.73. The USC Dornsife poll performed relatively better (Mosteller Measure of 2.45), while the polls conducted by Politico and NBC News/Wall Street Journal significantly diverged from actual outcomes (Mosteller Measures of 5.25 and 4.75 respectively). Overall, the table underscores a range of polling accuracy in the 2020 election, with House election polls demonstrating slightly lower accuracy than presidential election polls.

Rank/Firm	*Trump/Rep*	*Biden/Dem*	*Sample Size (N)*	*Mosteller Measure 3*
Election Result (President)	*46.8*	*51.3*		
1. *The Hill*/HarrisX	45	49	2,359	2.05
2. IBD/TIPP	46	50	1,212	1.05
3. *Reuters*/Ipsos	45	52	914	1.25
4. JTN/RMG Research	44	51	1,200	1.55
5. Harvard-Harris	46	54	2,093	1.75
6. Rasmussen Reports	47	48	1,500	1.75
7. FOX News	44	52	1,246	1.75
7. SurveyUSA	44	52	1,265	1.75
7. *USA Today*/Suffolk	44	52	1,000	1.75
10. *Economist*/YouGov	43	53	1,363	2.75
11. CNBC/Change Research	42	52	1,880	2.75
11. NBC News/*Wall Street Journal*	42	52	1,000	2.75
13. USC Dornsife	43	54	5,423	3.25
14. Quinnipiac	39	50	1,516	4.55
Average				*2.19*
Election Result (U.S. House)	*47.7*	*50.8*		
1. NBC News/*Wall Street Journal*	43	48	1,000	3.75
2. *The Hill*	42	47	2,359	4.75
3. *Politico*	43	50	14,663	2.75
4. USC Dornsife	45	53	5,413	2.45
5. YouGov	42	52	1,363	3.45
6. *USA Today*	39	49	1,000	5.25
Average				*3.73*

Figure 5.5.1. Final, National Presidential and U.S. House Preelection Poll Accuracy, 2020 (Panagopoulos,2021).

At the national level, Panagopoulos (2021) identified a notable and systematic pro-Democratic bias in pre-election polling. Specifically, he reported an average "A" measure score of −0.069, equating to an approximate overestimation of Joe Biden's actual popular vote margin by about 1.7 percentage points. This represented the most significant national polling bias detected since 1996. Of the 14 national surveys evaluated, only two—The Hill/HarrisX and Investor's Business Daily/TIPP—managed to project the actual election outcome within an acceptable range of ±0.5 percentage points from Biden's eventual 4.5-point popular vote victory. The remaining twelve polls consistently overstated Biden's advantage, suggesting a pervasive methodological issue related to sampling or weighting strategies that tended to disproportionately include or highlight voters inclined towards the Democratic Party.

In addition to national polls, the state-level analyses presented even more pronounced evidence of polling inaccuracies. Panagopoulos (2021) documented that roughly 86 percent (117 out of 136) of statewide polls conducted for presidential, Senate, and gubernatorial contests showed significant pro-Democratic bias, with an average "A" measure of −0.090, translating to about a 2.25-point overestimation in Democratic support. Notably, the bias was slightly greater in Senate races than in presidential or gubernatorial elections. The regression analyses conducted by Panagopoulos revealed limited association between poll accuracy and survey characteristics such as polling mode (live telephone versus online), sponsor affiliation, or sample size, implying that the issue was fundamentally embedded within the polling methodologies themselves rather than specific procedural decisions made by individual pollsters.

Placing these findings within a broader historical context, Panagopoulos (2021) highlighted that while previous U.S. election cycles, including those in 1996, 2004, 2008, and notably 2016, had also demonstrated some level of pro-Democratic polling bias, none were as statistically significant or widespread as those observed in 2020. He attributed this persistent issue in part to nonresponse bias, particularly among voter demographics traditionally associated with Republican support, including rural voters and those without college degrees. Additionally, he suggested the unprecedented circumstances surrounding the COVID-19 pandemic substantially altered voting behaviors, voter turnout dynamics, and the accessibility of voters to pollsters, thereby exacerbating the difficulties inherent in accurate voter modeling and sampling. Despite methodological refinements introduced by pollsters following the 2016 election, such as enhanced voter screens and improved weighting strategies, Panagopoulos concluded that the polling industry had yet to overcome fundamental structural challenges, highlighting the necessity for continued innovation and adaptability in future electoral polling efforts.

 https://doi.org/10.26219/heal.aueb.9743

Chapter 6 Measuring Polling Performance: An Over-Time, Cross-National Analysis of Electoral Accuracy.

Election polling has long stood as both a mirror and a barometer of democratic societies, offering a tangible window into the intentions and sentiments that animate electorates across the globe. From newsrooms to campaign headquarters, poll results shape narratives, expectations, and strategic choices, assuming a central place not only in the pre-election discourse but also in the very architecture of decision-making. Yet, the journey of polling from its genesis to its contemporary form has been punctuated by moments of profound doubt—occasions when the ostensibly rigorous machinery of survey science has produced misses that reverberate across political systems and erode public confidence.

Against this backdrop, it is essential to engage with empirical investigations that systematically evaluate the performance of election polling, moving beyond anecdotal accounts to comprehend the conditions under which polling errors occur. This chapter builds on the extensive synthesis of cross-national and longitudinal evidence assembled by Jennings and Wlezien (2018), among other foundational studies. Their comprehensive analysis, spanning numerous decades and diverse political contexts, reveals the persistent yet context-dependent and occasionally paradoxical nature of polling inaccuracies.

The aim of this body of work is not simply to enumerate polling failures but to extract enduring insights that can inform methodological refinements and improve the reliability of electoral forecasting. By integrating a substantial volume of polls and elections worldwide, Jennings and Wlezien provide a crucial empirical foundation for understanding when, where, and why polling errors arise.

Core Analysis: Historical, Cross-National, and Methodological Perspectives on Polling Errors

Polling inaccuracies are not a novel phenomenon confined to isolated elections or specific regions; rather, they have recurred consistently across democratic nations and electoral cycles. Jennings and Wlezien (2018) provide a compelling longitudinal and comparative narrative, charting quantitative and qualitative variations in polling performance across diverse political arenas—including the British General Elections of 1992 and 2015, the 2016 U.S. Presidential Election, and the Brexit referendum.

Their comprehensive dataset encompasses election polls from multiple countries such as the United States, United Kingdom, France, Germany, Canada, and Australia, covering several decades to capture temporal dynamics. The findings demonstrate that while some geographies exhibit lower average polling errors, no nation is immune to high-profile misses. The errors emerge from complex interactions involving sampling designs, voter turnout assumptions, and late shifts in public sentiment close to election day.

From a methodological standpoint, Jennings and Wlezien categorize polling errors mainly into sampling-related and response-related issues. Sampling-related problems arise from non-representative sample frames, declining participation rates, and inadequate coverage of key demographic subpopulations, such as young voters, minorities, or socioeconomically disadvantaged groups. Response-related errors stem from phenomena like social desirability bias and the "shy voter" effect, where respondents may conceal their true preferences due to social pressure or fear of ostracism. Turnout modeling remains especially challenging, as fluctuating participation rates and behavioral uncertainties compound predictive difficulties.

Significantly, the evolution of polling methodologies—transitioning from landline telephone surveys to mobile phones and online panels—has introduced new complexities without fully resolving these entrenched issues. Changes in communication habits and media consumption engender challenges in sampling frames, while internet-based approaches introduce concerns regarding self-selection and non-probability sampling biases.

The cross-national comparison underscores further nuances: electoral systems with proportional representation and multiparty competition impose distinct obstacles compared to two-party majoritarian contexts. Moreover, elections marked by ideological polarization or volatile, last-minute shifts in voter allegiances exhibit amplified polling errors. Even within a single country, polling accuracy varies substantially across electoral cycles, driven by shifting political landscapes and voter dynamics.

Taken together, this evidence strongly encourages a comprehensive and nuanced approach to polling. Instead of seeking a silver bullet, pollsters and researchers must integrate improved sampling frameworks, sophisticated turnout and behavioral models, transparent disclosure practices, and embrace innovative methodologies—such as model-based inference, real-time data integration, and adaptive weighting—to effectively contend with the multifaceted challenges of contemporary polling.

Conclusion

The narrative arc of election polling, as traced through comparative and longitudinal scrutiny, is one defined less by linear progress than by cycles of optimism, shock, and methodological reckoning. Despite the arsenal of modern tools at the disposal of pollsters—sophisticated weighting schemes, predictive turnout models, digital sampling methods—persistent discrepancies between poll predictions and electoral outcomes remain. These are not isolated blips but rather symptoms of the dynamic, multifactorial environment within which all polling operates: society's shifting currents, the enigmatic motivations of voters, and the fundamental challenges inherent to sampling a heterogeneous polity in flux.

As the evidence synthesized here compellingly demonstrates, there is no panacea for polling error. Rather, it is through an honest confrontation with complexity—acknowledging both the strengths and the intrinsic limitations of existing approaches—that the field can hope to make meaningful strides. The future of polling demands not only technical refinement and transparency, but also humility: a willingness to adapt, to innovate, and to rigorously test new methodologies against the rich tapestry of real-world electoral contests.

In conclusion, this chapter has revisited key sources of polling error while underlining the need for continuous reflection on how changing electoral contexts affect accuracy. By maintaining methodological rigor and adapting to new challenges, polling can remain a valuable tool for understanding public opinion and supporting democratic processes.

Chapter 7: Conclusions and Future Directions

This concluding chapter brings together the main findings presented across the previous sections, aiming to provide a coherent synthesis of the evidence and arguments discussed. It reflects on the key themes and methodological challenges that have been identified and considers their broader implications for the study and practice of election polling. At the same time, it points to areas where further research and methodological development are needed, offering a forward-looking perspective on how polling practices can continue to evolve in response to persistent and emerging challenges.

Recapitulating the Core Insights

As extensively unpacked in Chapter 3's discussion on *Sampling Frame & Coverage* (Section 3.1), the cornerstone of polling accuracy firmly rests on the representativeness of the sample. The transformation from traditional landline-based random-digit dialing techniques towards more contemporary approaches—such as address-based sampling and composite sampling frames—takes place against a backdrop of sweeping sociotechnical shifts. These shifts introduce both challenges and opportunities in ensuring all relevant voter subpopulations are adequately captured.

Delving deeper, Chapter 5's focused case study on the 2016 U.S. National Elections illuminated how underrepresentation of critical demographic segments—especially non-college-educated voters—coupled with inherent limitations in *likely voter* models and survey modes, significantly skewed polling outcomes. These findings, anchored in empirical evidence, reinforce the notion that sampling quality and methodological nuance are paramount.

Further enriching this discourse, the seminal work by Sturgis et al. (2017), as discussed in Chapter 6, cogently identified unrepresentative sampling as the principal cause behind the notorious polling shortfall in the 2015 UK General Election. Their analysis demonstrated that while turnout weighting adjustments were applied, their influence on the accuracy of final vote-share estimates was relatively minor. Instead, the bulk of the discrepancy was traced to deficiencies in the underlying samples themselves, which failed to adequately capture key segments of the electorate. This finding firmly positioned sampling error at the core of the polling miss, highlighting the critical importance of representativeness in survey design.

A Critical Appraisal of Findings and Their Significance

These converging lines of evidence spotlight the imperative for perpetual methodological evolution in response to dynamic technological and societal contexts. As articulated in Chapter 4, the erosion of stable telecommunications infrastructures and evolving communication habits present acute challenges to established sampling paradigms, threatening the very fabric of representativeness upon which polling validity depends.

Moreover, the exploration of social media as both a potent data reservoir and a methodological wildcard (elaborated in Chapters 5 and 6) reveals a delicate balancing act: while such platforms afford immediacy and rich, high-frequency signals, they are fraught with demographic skews and vulnerabilities to manipulation (Cerina, 2025; Scarano et al., 2024). This duality necessitates scrupulous correction techniques and critical scrutiny to elevate social-media-derived data from noisy chatter to actionable insight.

Hence, this work reinforces the contention that enhancing polling accuracy transcends incremental fixes, demanding an integrated, multilayered methodological framework that harnesses established surveying rigor alongside innovative data sources and analytical techniques.

Charting the Path Forward: Recommendations for Research and Practice

In light of the cumulative evidence and reflective synthesis presented, this chapter proposes emphatically the adoption of hybrid, multilevel predictive models—such as multilevel regression with post-stratification (MRP) and advanced machine learning algorithms (see Chapters 5.5 and 6)—to synergistically blend traditional survey data with auxiliary digital trace data.

It promotes the proactive and systematic approach of pilot testing, ongoing evaluation, and continuous adjustment of the sampling frames, which is an

Asia Chapter 3 theme, to prevent and address coverage gaps for underrepresented voters.

The study also calls for enhanced efforts to identify and address distortions arising from social media platforms, utilizing advanced techniques for bot detection and bias correction, in line with the approaches advocated by Scarano et al. (2024).

Final Reflections: Integrating Insight into Impact

In summation, this chapter collates and elevates the central themes traversed from Chapters 3 through 6, transforming them into a cohesive narrative of methodological vigilance and innovation. By foregrounding the intricacies of sample representativeness, weighting sophistication, and data source integration, the work advances a nuanced framework for understanding and overcoming the endemic challenges of electoral polling.

The insights presented go beyond academic theory. They offer a practical and visionary framework designed to enhance the reliability and influence of future public opinion research. Ultimately, this contributes to strengthening the foundations of democratic participation by promoting more accurate and trustworthy polling methodologies.

References

AAPOR Ad Hoc Committee (2018). *Final Report on the 2016 U.S. Presidential Election Polling Errors*. Public Opinion Quarterly.

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., et al. (2010). *AAPOR Report on Online Panels*. Public Opinion Quarterly.

Battaglia, M. P., Frankel, M. R., & Link, M. W. (1995). *Improving Standard Post-Stratification Techniques for Random-Digit-Dialing Telephone Surveys*. Survey Methodology.

Bodet, Marc André, Lydia Laflamme, Evelyne Brie and Catherine Ouellet. 2024. “Exit Polls in Canada: A Methodological Note.” Canadian Journal of Political Science 57 (4): 928–938.

Broniecki, P., Leemann, L., & Wüest, R. (2021). *Improved Multilevel Regression with Poststratification through Machine Learning (autoMrP)*. Political Analysis.

Callegaro, M., & Yang, Y. (2020). *The Role of Surveys in the Era of Big Data*. Journal of the Royal Statistical Society: Series A.

Cerina, R., & Duch, R. (2023). [Relevant works on the use of LLMs and the application of PoSSUM within the MRP framework]. (If this is a manuscript or unpublished communication, format accordingly.)

Crossley, P. (1937). “Response Bias in Opinion Polls.” *American Journal of Sociology*.

Couper, M. P., & Miller, P. V. (2008). *Web Survey Methods: Introduction*. Public Opinion Quarterly.

Curtin, R., Presser, S., & Singer, E. (2005). *Changes in Telephone Survey Nonresponse Over the Past Quarter Century*. Public Opinion Quarterly.

Durand, C. P., Heath, A., Jeffery, B., Curtice, J., & Johns, R. (2001). *Shy Tory or Low Turnout?*. British Election Study Working Paper.

https://doi.org/10.26219/heal.aueb.9743

Erikson, R. S., & Wlezien, C. (2012). *The Timeline of Presidential Elections*. University of Michigan Press.

Gallup, G. H., & Rae, S. B. (1940). *The Pulse of Democracy: The Public-Opinion Poll and How It Works*. Simon & Schuster.

Gelman, A., Park, D. K., & Bafumi, J. (2004). *Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls*. Political Analysis.

Goldberg, A., et al. (n.d.). *Random Domain Intercept Technology as a Real-Time Data Collection Method*. [Working paper].

Groves, R. M. (1989). *Survey Errors and Survey Costs*. Wiley.

Jennings, W., & Wlezien, C. (2018). *Election Polling Errors across Time and Space*. Nature Human Behaviour.

Katz, D., & Cantril, H. (1937). *Public Opinion Polls: Their Use in a Democracy*. Annals of the American Academy of Political and Social Science.

Keeter, S., Kennedy, C., Clark, A., Javitz, H., Best, J., & Peskin, J. (2006). *Assessing the Cell Phone Challenge to Survey Research*. Public Opinion Quarterly.

Kennedy, C., Blumenthal, M., Clement, S., Clinton, J. D., Durand, C., Franklin, C., McGeeney, K., Miringoff, L., Olson, K., & others. (2018). *An evaluation of the 2016 election polls in the United States*. **Public Opinion Quarterly, 82**(1), 1-33.

Kennedy, C., Popky, D., & Keeter, S. (2023). *How Public Polling Has Changed in the 21st Century*. Pew Research Center Report.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, 9(2), 137–163.

Kish, L. (1965). *Survey Sampling*. Wiley.

Krosnick, J. A., & Chang, L. (2009). *Attitudinal Consistency Over Time*. In *Handbook of Survey Research*. Emerald.

Marsh, C., & McElroy, G. (2003). *Herding in Opinion Polls: A Study of the 2003 UK General Election*. Journal of Elections, Public Opinion and Parties.

Mellon, J., & Prosser, C. (2017). *Correcting Polling Error in Over-engaged Samples Using Representative Turnout Weighting*. British Election Study.

Meng, X.-L. (2018). *Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox*. **Annals of Applied Statistics, 12**(2), 685-726.

Murray-Watters, A., Zins, S., Silber, H., Gummer, T., & Lechner, C. (2023). *River Sampling – A Fishing Expedition: A Non-Probability Case Study*. Survey Practice.

Panagopoulos, C. (2020). *Accuracy and Bias in the 2020 U.S. General Election Polls*. Journal of Elections.

Park, D. K., Gelman, A., & Bafumi, J. (2004). *Bayesian Multilevel Estimation with Post-stratification: State-Level Estimates from National Polls*. Political Analysis.

Prosser, C., & Mellon, J. (2016). *Modelling Turnout in the 2015 UK Election*. British Journal of Political Science.

Silver, N. (2012). *The Signal and the Noise: Why So Many Predictions Fail—but Some Don't*. Penguin.

Simon, J. (1954). "The Influence of Polls on Voter Behavior." *Public Opinion Quarterly*.

Squire, P. (1988). *Why the 1936 Literary Digest Poll Failed*. Public Opinion Quarterly.

Sturgis, P., Kuha, J., Baker, N., Callegaro, M., Fisher, S., Green, J., Jennings, W., Lauderdale, B. E., & Smith, P. (2017). *An Assessment of the Causes of the Errors in the 2015 UK General Election Opinion Polls*. Journal of Elections.

Sturgis, P., Baker, N., Callegaro, M., Fisher, S., Green, J., Jennings, W., & Prosser, C. (2016). *Report of the Inquiry into the 2015 British General Election Opinion Polls*. British Polling Council.

Thomas, M., Sopata, C., Rogers, B., & Marusco, S. (2021). *Forecasting the 2020 Presidential Election: A Comparison of Methods*. University of Virginia Working Paper.

Traugott, M. W. (1992). *The Role and Impact of Polls in U.S. Presidential Campaigns*. Public Opinion Quarterly.

Wilks, S. S. (1940). "Representative Sampling for Public-Opinion Polls." Public Opinion Quarterly.

Weisberg, H. F. (2005). *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. University of Chicago Press.

Wang, S., Rothschild, D., Goel, S., & Gelman, A. (2015). *Forecasting elections with non-representative polls*. **International Journal of Forecasting, 31**(3), 980-991.

West, B. T., & Andridge, R. (2023). *A diagnostic measure for non-ignorable selection bias in pre-election polling estimates*. **Public Opinion Quarterly, 87**(S1), 475-501.

Wu, S. (2022). *Mobile interviewing and coverage gaps: Challenges for modern probability samples*. **Survey Methodology Journal, 48**(2), 201-222.