AAPOR

AMERICAN ASSOCIATION FOR
PUBLIC OPINION RESEARCH

TASK FORCE ON 2024 PRE-ELECTION POLLING

An Evaluation of the 2024 General Election Polls

AAPOR would like to thank the following Task Force members for their work on this project and service to the public opinion research community:

Josh Pasek (Chair), University of Michigan

Andrew Mercer, Pew Research Center

Ariel Edwards-Levy, CNN

Scott F. Clement, Washington Post

Ruth Igielnik, New York Times

Jonathan Ladd, Georgetown University

Emily Swanson, The Associated Press

Jeff Horwitt, Hart Research Associates

Charles Franklin, Marquette University Law School

Chris Jackson, SSRS

Kabir Khanna, CBS News

Matthew Knee, WPA Opinion Research

Angela X. Ocampo, University of Michigan

Tasha Philpot, University of Texas, Austin

Elliott Morris, Strength in Numbers

Samara M. Klar, University of Arizona

This report was commissioned by the AAPOR Executive Council as a service to the profession. The report was reviewed and accepted by the AAPOR Executive Council. The opinions expressed in this report are those of the author(s) and do not necessarily reflect the views of the AAPOR Executive Council.

Table of Contents

Executive Summary 4

1 Introduction 10

2 Accuracy of the 2024 Pre-Election Polls 12

2.1 Overall Accuracy of 2024 Polls 12

2.2 Comparison to Prior Cycles 15

2.3 Firm-Based Correlates of Accuracy 19

2.4 Methodological Correlates of Accuracy 22

2.5 Geographic Accuracy and Turnout Shifts 28

2.6 Subgroup Accuracy and Representation 34

2.7 Pollster-Level Variation in Subgroup Estimates 36

2.8 Herding and Poll Agreement 39

2.9 Shifts in Polling Volume and Geographic Focus 42

2.10 Making Sense of Past Vote 44

3 Conclusions and Future Considerations 47

3.1 Summary of Findings 47

3.2 Limitations and Interpretive Caution 48

3.3 Considerations for Future Polling 49

Appendix A. Data for the 2024 Report 51

A.1 Data sources for 2024 and historical polling margins 51

A.2 Dataset limitations 52

A.2 Replication and data access 52

Appendix B. Poll Inclusion and Cleaning Rules 53

B.1 Initial Inclusion Criteria 54

B.2 Field Date Rules and Final Window 54

B.3 Duplicate Handling and Source Precedence 55

B.4 Handling of Multiple Releases 55

B.5 Contest Coverage 56

Appendix C. Variable Definitions and Coding Crosswalk 57

C.1 Key Variables Used in Analysis 59

C.2 Coding Sources and Prioritization 60

Appendix D. Benchmark Election Results 61

Appendix E. Microdata Contributions 62

E.1 Contributing Polling Organizations 62
E.2 Variables Included in Shared Files 62
E.3 Validation, Cleaning, and Use 63
E.4 Data Access and Retention 63
Appendix F. Sub-group and geographic diagnostics 64
Appendix G. Accuracy Metrics and Statistical Tests 65
Appendix H. Alternative Accuracy Windows and Metrics 66
H.1 Alternative Field Windows 66
H.2 Alternative Error Metrics 66
Appendix I. Full Accuracy Tables by Contest and State 70
I.1 2024 Accuracy by Contest Type 70
I.2 2024 Accuracy by State (Presidential Contests) 71
I.3 Historical Accuracy by Year and Contest Type (Final Window) 71
I.4 Notes on Table Construction 72
Appendix J. Methodology Survey Instrument and Responses 73
J.1 Survey Outreach and Participation 73
J.2 Survey Instrument 73
J.3 Aggregate Results (Selected Items) 73
J.4 Use of Survey Results in the Report 76
Appendix K. Replication Files and GitHub Link 77
K.1 Repository Contents 77
K.2 Microdata and Restricted Files 77
K.3 Use and Citation 77
Appendix L. Public Narrative and Context 78
L.1 Historical Background: A Crisis of Confidence 78
L.2 Media Coverage and Forecasting Ecosystem 78
L.3 Polling in 2024: A High-Scrutiny Environment 79
L.4 Implications for the Future 79

Executive Summary

When the results of the 2024 election were tallied, they showed that the picture of the race provided by publicly released polls had been largely accurate. As the polls had indicated, the race between Kamala Harris and Donald Trump was close, in both pivotal swing states and the nation as a whole. Coming on the heels of larger-than-typical errors in 2016 and 2020, and in the face of considerable skepticism of surveys' accuracy, 2024 was a good year for public polling, even as the need for continued experimentation in future cycles remains.

The Executive Council for the American Association for Public Opinion Research established a task force with the goal of assessing the accuracy of pre-election surveys in 2024, making sense of what worked and what was less effective. To that end, the AAPOR task force gathered information about the publicly released election polls conducted in 2024 and prior years, as well as a number of different data sources with which those polls could be compared.

The main findings of the AAPOR task force are as follows:

Public polls were more accurate in 2024 than in 2020 and 2016. Across 611 polls of presidential, senatorial, gubernatorial, and congressional contests fielded in the campaign's final two weeks, the average absolute error on the two-party margin was 3.3 percentage points—down from 5.3 points in 2020 and 5.2 points in 2016. National presidential polls missed by 2.6 points on average, and state-level presidential polls by 3.0 points. State polls were their most accurate for any presidential cycle since 1944 and national errors were close to their average over the long run.

For the third straight presidential cycle, pre-election polls underestimated Republican vote shares relative to Democrats'. Polls in the last two weeks of the campaign overstated Democratic margins by 2.7 points across all offices—smaller than the 4.6-point overestimate in 2020 and 3.1 points in 2016, yet still notable. In comparison, there were relatively small average signed errors in recent midterm years, which overestimated Republicans by only 0.6 points in 2022 and Democrats by only 0.1 points in 2018. Historically, it has been rare for polls to err in the same direction for more than a couple of consecutive presidential elections: Democrats and Republicans have each been underestimated about equally often since the dawn of modern polling in the 1930s (Figure ES-1 displays average signed and absolute error for presidential general election polls since 2000).

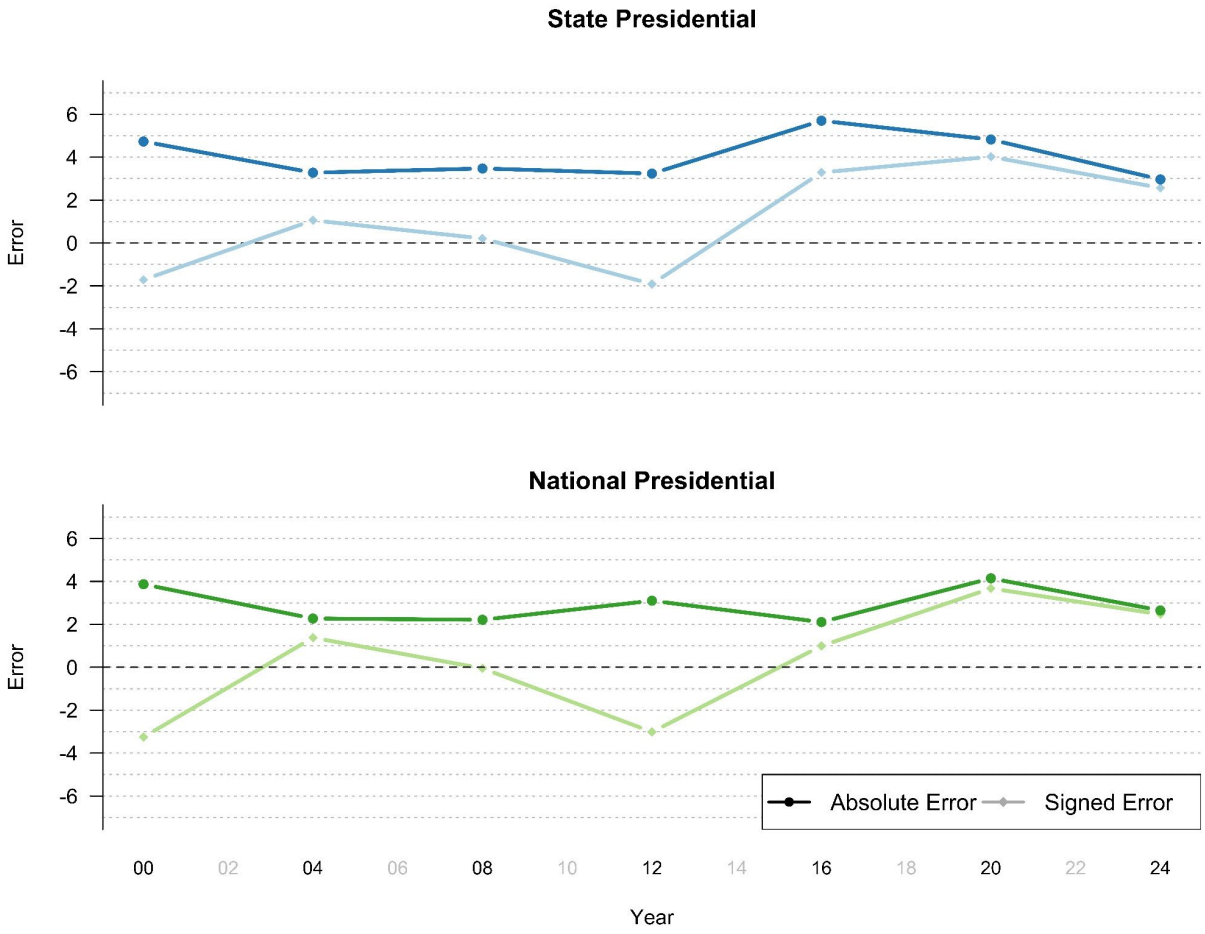

Figure ES-1 Signed and absolute polling errors for state and national presidential polls since 2000. For absolute errors (darker colors), the y-axis measures the difference in percentage points between the average poll for each type of contest and the final result. For signed errors (lighter colors), the y-axis measures the average directional difference between all polls for each type of contest and the eventual result. Positive signed errors indicate that polls overestimated Democratic performance whereas negative signed errors indicate that polls overestimated Republican performance.

No single methodological recipe guarantees higher accuracy. Polling firms used an array of sampling frames, modes, weighting variables, and likely-voter models in 2024, yet most major methodological choices showed little relationship to error size. Surveys from higher-volume firms, those weighting on partisan self-identification, those using detailed likely-voter models, and Republican-affiliated pollsters were slightly more accurate, but differences were small and may reflect other attributes of these pollsters' approaches rather than the methods themselves. For Republican pollsters in particular, this increased accuracy could possibly reflect a tendency to publicly release more Republican-friendly numbers in most elections. This tendency in itself would produce more accurate numbers in years when the average poll has a Democratic bias, but not in other years.

Accounting for past voting can improve poll performance, but how to do so is not straightforward. Many firms incorporated information about voters' 2020 voting behavior into sampling, weighting, or likely-voter modeling. The fact that most people voted the same way in 2024 as they had in 2020 means that if pollsters knew who would turn out and the past voting behaviors of those individuals, they could have almost perfectly predicted the results. Because of this predictive power, an ideal application of past vote could cut average absolute error considerably when added to a poll's weighting strategy. Yet those gains depend critically on how past voting behavior is measured and used to calibrate the sample, what firms did to account for differences between the prior and current electorates, and how well they forecast the swing from election to election. For individual surveys, the efficacy of these efforts varied.

Some key groups of voters are difficult to capture in surveys. Polls did not reliably measure the preferences of three critical blocs that fed into Democratic overestimation in 2024: (1) Republican voters in GOP-leaning areas, who were under-represented relative to Democrats in those same GOP-leaning areas; (2) Hispanic voters, whose Democratic support was overstated in pre-election polling versus post-election data; and (3) 2024 voters who had not voted in 2020—surveys signaled that they leaned Republican but still underestimated their share of the electorate.

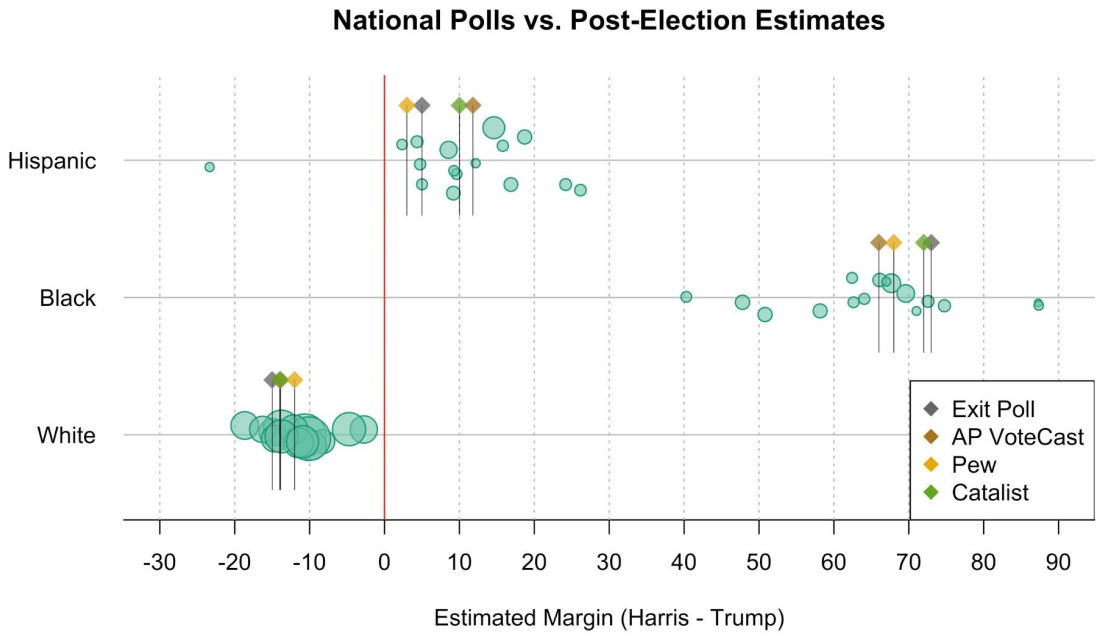

Figure ES-2: Comparing pre-election poll group estimates from microdatasets with alternate benchmarks for racial and ethnic group voting preferences using firm-provided weights. Benchmark post-election studies suggest that polls overestimated Hispanic voter preferences for Harris relative to Trump.

Although variability across survey estimates was relatively low, this does not appear to be a result of herding. Polls in 2024 displayed impressive consistency—especially in swing states, where most surveys showed the difference between Trump and Harris within the margin of sampling error. While some observers suspected that this was evidence of firms adjusting their results to match competitors ("herding"), our tests find no evidence to support this; rather, a broader reliance on political variables for sampling and weighting likely reduced spread.

Polling did not reliably anticipate within-state turnout shifts. Counties that backed Trump in 2020 saw turnout surges in 2024, whereas Biden-leaning counties in 2020 saw declines. Most surveys assumed a 2024 electorate distributed much like 2020. That turnout mis-projection explains a moderate share of the remaining directional error.

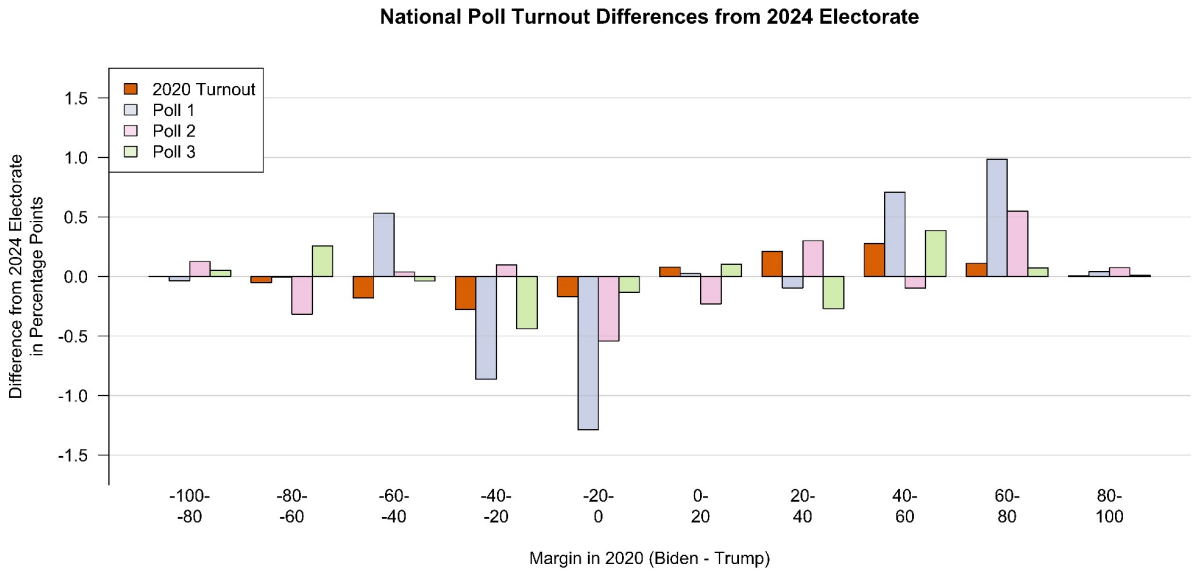

Figure ES-3: Comparing turnout in 2020 and in estimates from three county-matched national polls with actual votes cast in 2024. Plot shows differences in percentage points from 2024 turnout for each 10-percentage-point margin range of county preference from 2020. Counties that supported Trump in 2020 had higher relative turnout than projected in national polls.

Different pollsters told different stories about the electorate, while still reaching the same broad conclusions about the election's outcome. Reliance on various combinations of partisanship, past vote, and weighting variables in 2024 appears to have ensured that most surveys produced candidate-margin estimates that were very close to the final results. However, there was considerably more dispersion in estimates of how specific groups voted. Although polls generally agreed on which groups leaned toward Trump or Harris, their estimated margins varied widely, especially for Hispanic Americans, who seem to have moved sharply toward Trump in 2024.

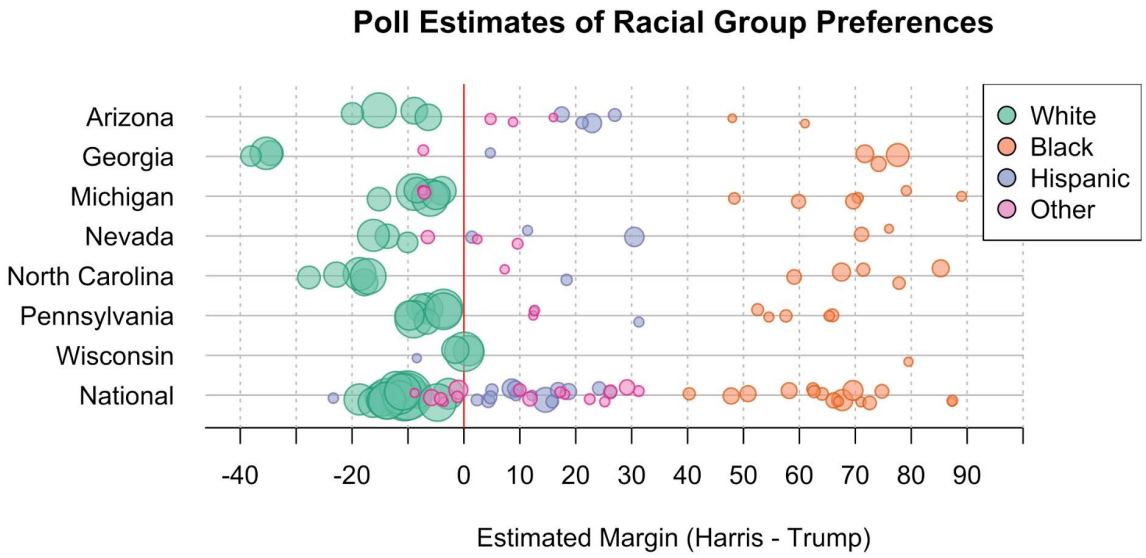

Figure ES-4: Estimated voting margins of racial groups in polls using firm-provided weights with microdata. Each dot represents a single poll's estimate for the margin within a particular racial or ethnic group. Dot size corresponds to within-group sample size. Groups with Ns<50 not shown. Although relative ordering of subgroups was consistent, high variability in estimates for some groups complicates conclusions about those groups.

Election polling is far more focused on battleground states than it used to be. In 2024, there were nearly twice as many state-level presidential polls as national ones, with most of that activity clustered in seven swing states (AZ, GA, MI, NV, NC, PA, WI). The concentration intensified in the campaign's final weeks—a continuation of a long-running shift and a notable jump even since 2020. This focus may help to better inform the public's election-night expectations (because presidential outcomes are decided state-by-state), but it does come with the tradeoff of reduced awareness of what was happening in less frequently polled states.

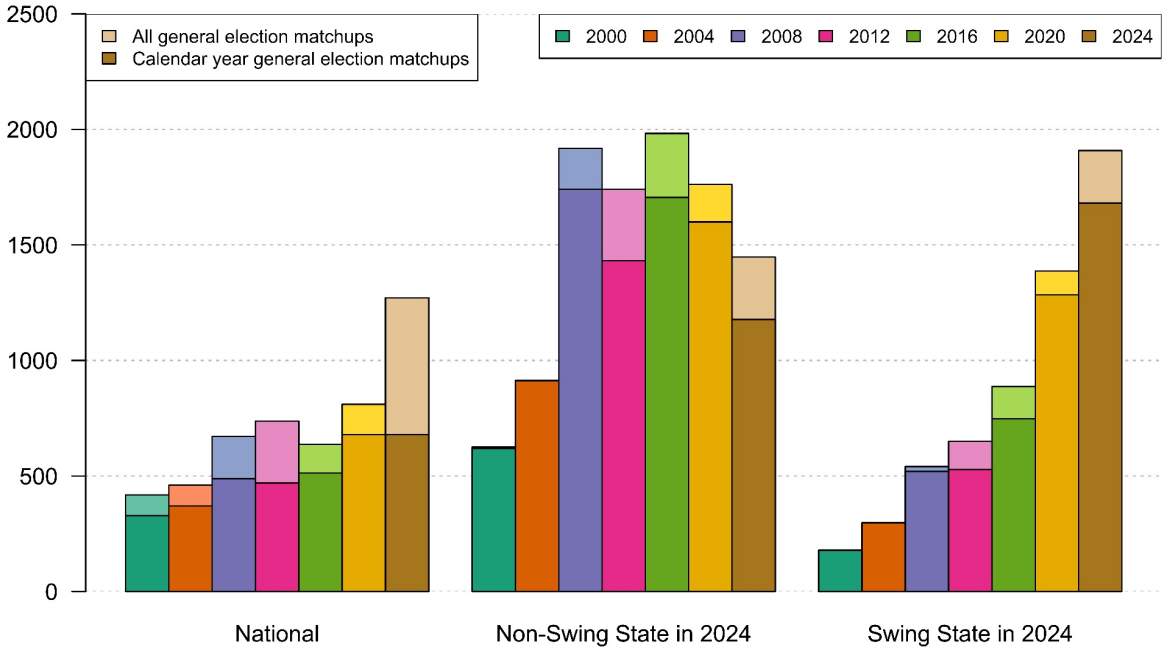

Figure ES-5: Total volume of matchups reported from all general elections by election year nationally, for states that were not considered swing states in 2024, and for states that were considered swing states in 2024 in recent presidential cycles (darker portion of each bar indicates matchups reported in the same calendar year as the election)

Taken together, these patterns show meaningful progress: public polls painted an essentially accurate picture of an extraordinarily close contest. Still, the modest Democratic overstatement, lingering errors in representing key blocs, and difficulties in modeling turnout underscore the need for continued adaptation and experimentation—especially around sampling hard-to-reach populations, weighting on political variables, and building turnout models that adapt to asymmetric changes in turnout.

Technical note: All accuracy statistics in this Executive Summary use only those polls that completed fieldwork between October 23 and November 5, 2024 and report a two-party vote share. Totals may differ in later sections that employ broader field dates, down-ballot contests, or additional methodological filters.

1 Introduction

The polling community entered the 2024 election cycle still nursing bruises from 2016 and 2020. Underestimations of Donald Trump's electoral performance in those years left many observers convinced that polling was broken, or at least not up to the task of describing upcoming elections. A healthy dose of caution about the predictive power of pre-election polls may have been overdue, but disappointment with the 2016 and 2020 misses gave way to outright skepticism in 2024. When a series of extraordinary events rocked the presidential race—a shaky debate performance by President Biden, his decision to withdraw and endorse Vice President Harris, and assassination attempts against Trump—polls moved only slightly, seemingly reinforcing critics' doubts. Yet, despite all of this turmoil, end-of-campaign polls in 2024 were quite accurate by historical standards.

How should the public interpret that apparent rebound? Did pollsters truly right the ship? Did the problems that befell 2016 and 2020 simply dissipate? Or did polling luck into a better result this cycle? These questions motivated the American Association for Public Opinion Research's (AAPOR) Task Force on 2024 Pre-Election Polling. Building on prior assessments of polling performance in 2016 and in 2020, sixteen experts in survey methodology assembled a comprehensive database of every publicly released poll, documented what pollsters did in 2024, compared those choices with earlier cycles, and evaluated how each approach related to accuracy. The findings are presented in this report.

Before turning to the evidence, it is worth underscoring a few limitations—both of this report and of polling itself. Election surveys serve multiple constituencies: campaigns, news organizations, universities, and private firms. Their methods, target populations, and goals vary widely, and not every poll's chief goal is to anticipate the final vote tally. Our analysis covers only publicly released polls and determines accuracy by comparing their published toplines to official election returns—a standard that, while conventional, may not match every pollster's intent.

Even if every survey were designed solely to predict the vote, pollsters would still confront a daunting set of choices: whom to contact, how to find them, how to persuade them to respond, and how to weight their answers. Over the past two decades, those choices have proliferated. Some innovations reflect genuine methodological progress; others are creative attempts to cope with soaring costs and falling response rates.

Another challenge is that pre-election polls also ask people about actions they will take in the future. Inaccurate responses may come from people either misreporting their current intentions or sincerely failing to predict their future actions. Late-breaking events can also upend vote intentions after a poll is fielded, and the survey results themselves can influence turnout or candidate strategy.

Finally, public understanding of "poll accuracy" now blurs into reactions to aggregated forecasts and statistical models that blend many surveys with other data. While such models dampen random error, shared biases can still skew results and inflate confidence. A flood of partisan

polls at the wrong moment can temporarily warp even the best weighting schemes when aggregated.

Against that backdrop, the task force approached 2024 with realistic expectations: polling is an informative but imperfect snapshot, shaped by human choices and subject to uncertainty.

In line with its charge, this report:

- **Assesses the accuracy of 2024 pre-election polls** (Section 2.1)
- **Compares 2024 accuracy with prior cycles** (Section 2.2)
- **Examines whether methodological choices correlate with accuracy** (Section 2.3)
- **Draws conclusions, notes limitations, and looks ahead** (Section 3)

The report relies on several sources of data in its analysis. Records of the publicly released polls conducted during the 2024 campaign were retrieved from the FiveThirtyEight and Real Clear Politics aggregates, as well as the Roper Center iPoll archives.

After de-duplication, this method identified 2,631 unique surveys covering 194 contests and fielded by 219 organizations across the calendar year; 403 of those surveys, reflecting 611 distinct matchups in specific races, were conducted in the two-week window prior to the election.

For each poll identified, the task force attempted to catalog information on four variables: the mode (how interviews were conducted), the sample frame (how potential respondents were identified and selected), weighting variables (how demographics and other factors were adjusted to make the sample look like the broader electorate) and any likely-voter modeling (whether and how pollsters attempted to determine which respondents would actually vote). Each poll's results were also matched to corresponding election results to assess accuracy.

To evaluate accuracy, this report uses two metrics to describe errors in public opinion polling: absolute error and signed error. The average absolute error reflects the mean distance—without regard to direction—between each poll's two-party margin and the certified election margin, in percentage points. It illustrates the extent to which polls missed. The average signed error also includes direction. For it, positive values signify that polls leaned Democratic whereas negative values denote a Republican lean.[1] These metrics were chosen because they have been standard in past reports and give a sense of both overall and directional inaccuracies.

In addition to this polling data, the task force also asked 86 of the most prolific 2024 polling organizations if they would be willing to share a de-identified copy of at least one data set, and to complete a questionnaire about their field practices. Thirty-nine pollsters completed the questionnaire, while 24 shared microdata. Although these questionnaire responses and data are

[1] Because polls overestimated Democratic support in 2024 and underestimated Republican support, signing the error this way makes it easier to compare the magnitudes of the signed and absolute errors. Because signed errors can be in both directions whereas absolute errors are always positive, signed errors can never be larger than absolute errors.

not representative of the polling landscape more broadly, they provide a more granular look at *where* polling errors emerged—and which voters were hardest to reach.

For more information on the report's methods and the data used, see Appendices A-G.

These appendices also provide historical background, detailed coding rules, full tables, and supplementary figures. A full discussion of how current methods and results fit into the history of public polling along with a number of additional analyses will be presented in a forthcoming book from the committee.

2 Accuracy of the 2024 Pre-Election Polls

This section presents the core findings of the report: how accurate pre-election polls were in 2024, how that accuracy compares with previous election cycles, and what factors explain the variation across surveys. It begins by summarizing overall performance in 2024 (Section 2.1), then traces accuracy trends across recent cycles (Section 2.2), and explores how polling methods, subgroup representation, geographic coverage, and pollster-specific choices shaped results (Sections 2.3–2.7). Finally, it discusses some key changes in the polling landscape and their implications (Sections 2.8 and 2.9).

2.1 Overall Accuracy of 2024 Polls

Pre-election polling in 2024 was more accurate than in the two previous presidential cycles. Among the 611 general-election polls that finished fieldwork between October 23 and November 5, the average absolute error in the two-party margin between the Democratic and Republican nominees was 3.3 percentage points—a notable drop from 5.3 points in 2020 and 5.2 points in 2016. Accuracy improved across nearly every contest type, with presidential polling leading the way: national presidential polls had an average absolute error of 2.6 points, while state-level presidential polls missed by 3.0 points. This made 2024 the most accurate cycle for presidential polling at the state level since 1944.

Table 2.1.1 - Polling Error from the Last Two Weeks Before Election in 2024

Matchup Type	# Matchups	# Firms	# Contests	Absolute Error (Margin)	Signed Error (Margin)
National presidential	60	36	1	2.6	+2.5
State (all) presidential	291	62	35	3.0	+2.6

Swing State presidential	190	44	7	2.3	+2.0
Other state presidential	101	38	28	4.3	+3.6
U.S. Senate	197	54	27	3.3	+2.4
Governor	27	17	7	5.0	+3.8
U.S. House	28	11	20	7.0	+4.9
Other presidential	8	4	5	4.7	+4.5
All Contests	611	86	95	3.3	+2.7

A majority (57.3%) of the presidential polls fielded within two weeks of the election differed from the final tallies by less than 3 percentage points and 22.9% were within a single percentage point. Figure 2.1.1 shows the distributions of the absolute errors for state and national presidential matchups. The vast majority had errors that were relatively small. Larger errors were mostly concentrated in states that were not the primary focus of the campaigns. Individual surveys in California, Iowa, New Jersey, and Wyoming erred by 10 percentage points or more, but the outcome of the presidential election in these states was never truly in doubt. Although at least some published estimates from Florida and Nevada were off by similarly large margins, these were from polls that released multiple estimates of those contests based on the same underlying survey data; other estimates produced from the same data in those states were considerably closer to what the candidates would receive on Election Day. This leaves only a single miss that was both substantively large and made a state seem out-of-play when it was actually relatively close: New Hampshire, in which one poll was off by 25.6 percentage points.

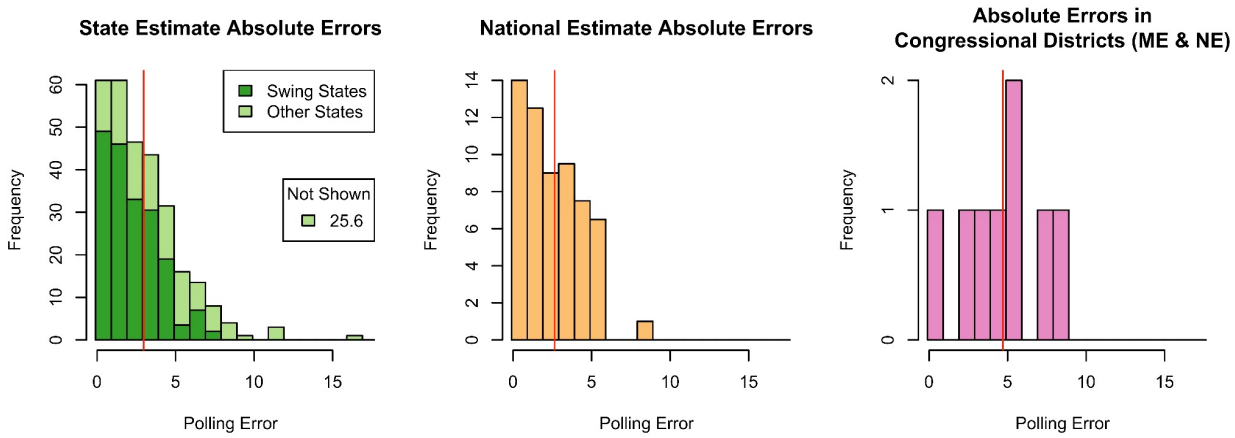

Figure 2.1.1 Distributions of absolute polling errors across 2024 Presidential contest types for the last two weeks of the election. The vertical red line indicates the average absolute error across all polls of each type.

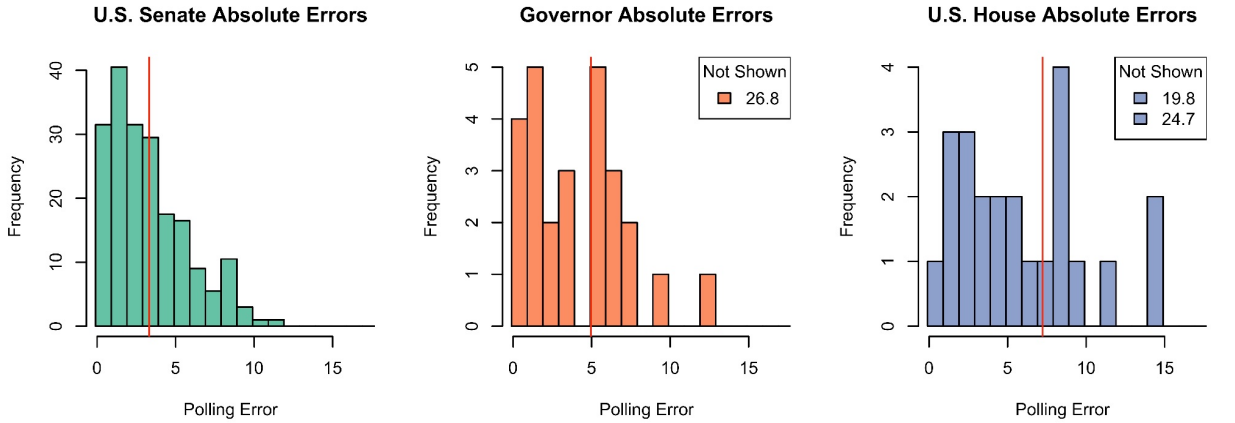

Figure 2.1.2 Distributions of absolute polling errors across other 2024 contests for the last two weeks of the election.

Senate and gubernatorial contests showed somewhat greater errors than presidential polls. Senate polls fielded in the final two weeks of the campaign had an average absolute error of 3.3 percentage points, while gubernatorial polls averaged 5.0 points. House race surveys were somewhat less accurate, with an average absolute error of 7.0 points, though still near historical norms. These results suggest that accuracy improvements were concentrated in the top-of-ticket races.

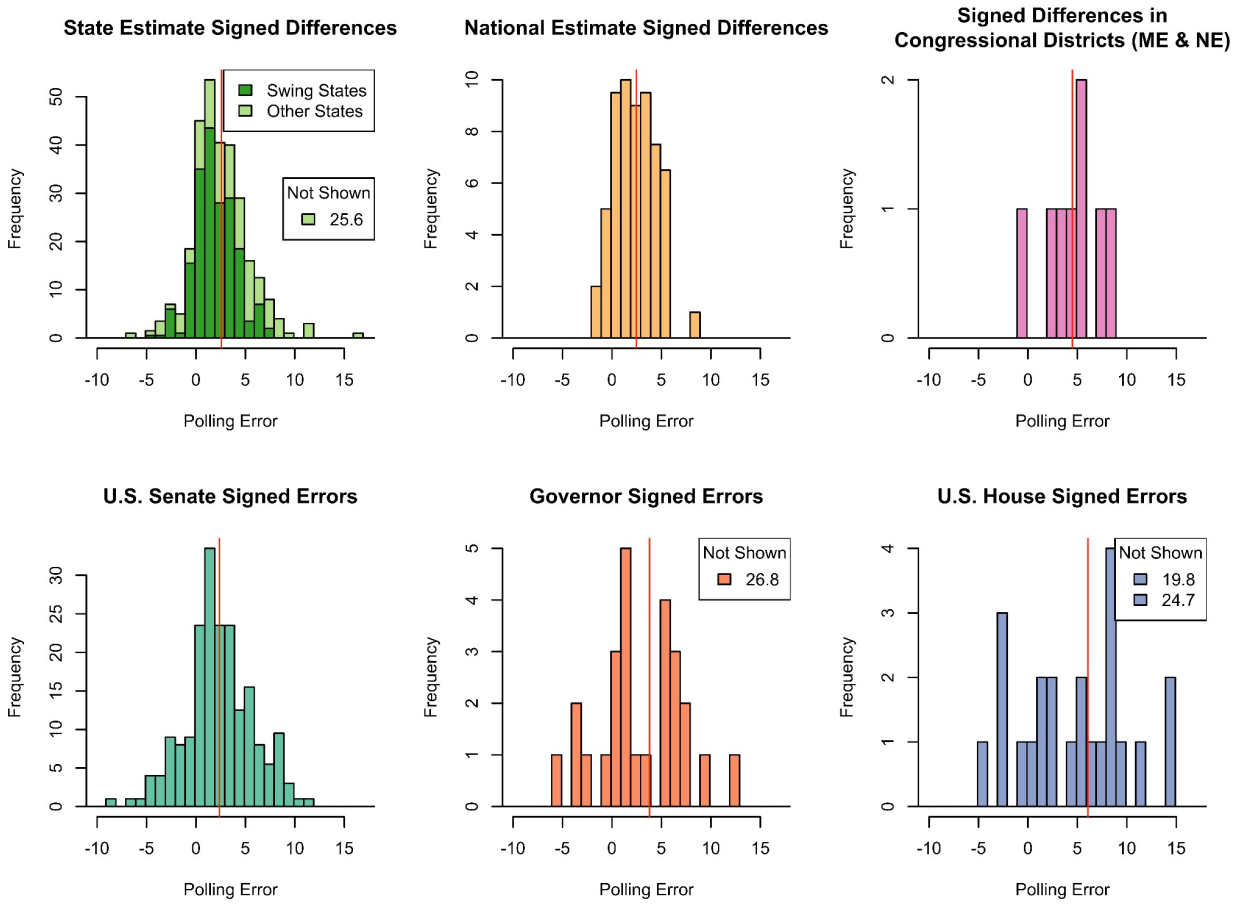

Figure 2.1.3: Distributions of signed errors across contest types for all polls in the last two weeks of the election. The vertical red line indicates the average signed error across all polls of each type.

Signed error—the direction of the polling miss—also improved. In 2024, polls overstated Democratic margins by 2.7 percentage points on average (across all contests) for polls that concluded during the final two weeks. While that figure remains non-trivial, it marks a significant improvement from 2020, when polls overstated Democrats by 4.6 points, and from 2016, when the average signed error was 3.1 points.

Together, these results show that 2024 marked a substantial recovery in overall polling accuracy, though modest Democratic overstatement persisted and remains a focus of analysis in the sections that follow.

2.2 Comparison to Prior Cycles

The size of the polling errors in 2024 were not historically unusual. Where 2024 was somewhat notable, however, was that it reflected the third straight presidential cycle in which Republican vote shares were underestimated relative to Democrats', though the size of that imbalance declined from the previous two cycles.

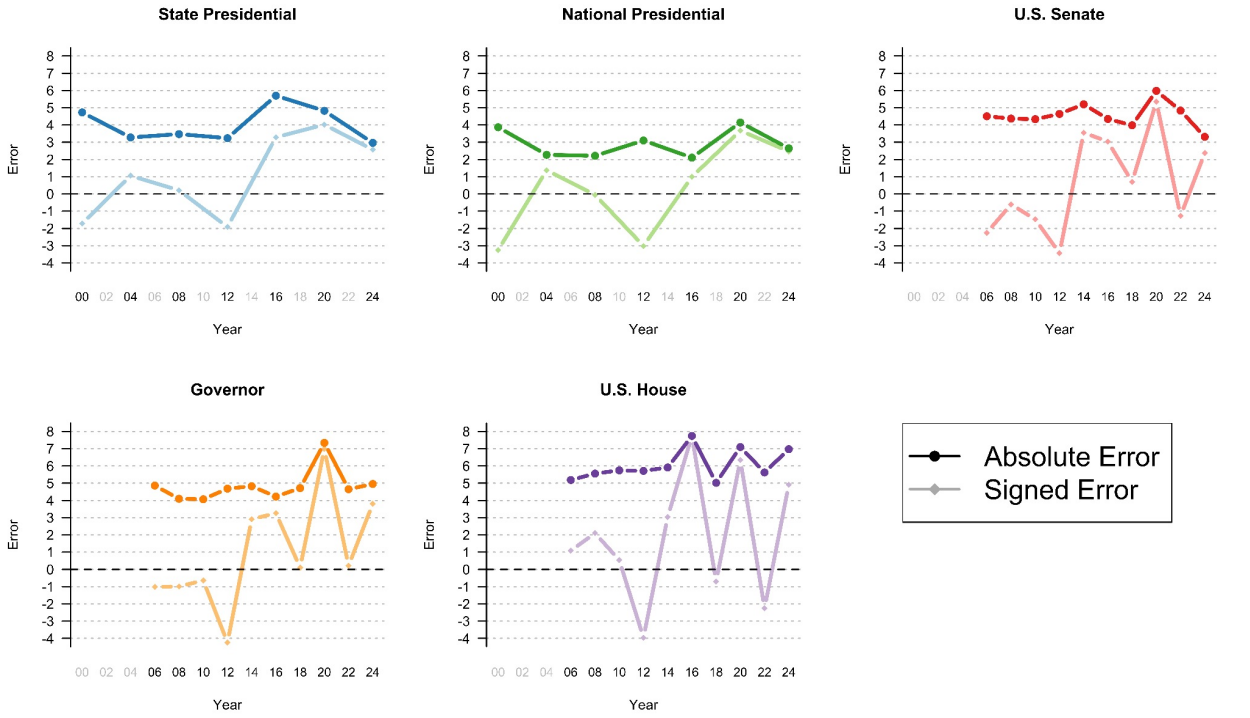

Figure 2.2.1: Signed and absolute polling errors for various offices since 2000. For absolute error, y-axis measures the difference in percentage points between the average poll for each type of contest and the final result. For signed errors, the y-axis measures the average directional difference between all polls for each type of contest and the eventual result. Positive signed errors indicate that polls overestimated Democratic performance whereas negative signed errors indicate that polls overestimated Republican performance.[2]

Recent midterm cycles, by contrast, showed smaller directional error. Polls in 2022 slightly overestimated Republican performance (average signed error: –0.6 points), while polls in 2018 slightly overestimated Democrats (+0.1 points). These fluctuations are consistent with the long-run tendency for polling error direction to vary from cycle to cycle. Historically, it has been rare for the same party to be underestimated across three consecutive presidential elections, underscoring the lingering challenges of reaching and modeling parts of the Republican electorate as it has evolved in years when Donald Trump was on the presidential ballot.

Table 2.2.1 Signed and Absolute Polling Errors in Recent Years for the Presidential contest

	Average National Presidential Popular Vote Polling Error	Average State-Level Presidential Polling Error

[2] Historical data on poll accuracy in this report differs slightly from previous AAPOR reports, as the 2024 committee sought to assemble a more comprehensive dataset of polls and adopted new methods to account for polls that produced multiple vote choice estimates and for tracking polls. The updated accuracy estimates for previous years differ by less than one percentage point with the exception of 2000 national polls, where the new dataset found a signed error of -3.3 compared with -1.1 in previous reports, largely due to considering all polls in the final two weeks rather than the just the final poll from a limited set of national polls.

Year	Signed Error	Absolute Error	Signed Error	Absolute Error
2000	-3.3	3.9	-1.7	4.7
2004	1.4	2.3	1.1	3.3
2008	-0.1	2.2	0.2	3.5
2012	-3.0	3.1	-1.9	3.2
2016	1.0	2.1	3.3	5.7
2020	3.7	4.1	4.0	4.8
2024	2.5	2.6	2.6	3.0
2024 (Last week)	2.3	2.3	2.5	2.9
2024 (Last 3 days)	2.5	2.5	2.4	2.7

Broadly, then, polling accuracy in 2024 represented a notable return to form. As shown in the long-run series of national presidential polling errors, the average absolute error across all final two-week surveys fell sharply from 4.1 points in 2020 to 2.6 points in 2024, bringing accuracy back in line with historical norms. For state-level presidential polling, the improvement was similarly pronounced, with the average absolute error dropping from 4.8 points in 2020 and 5.7 points in 2016 to 3.0 points in 2024. This places 2024 among the most accurate cycles in the last half-century and makes it the smallest absolute error for state-level presidential polling since 1944.

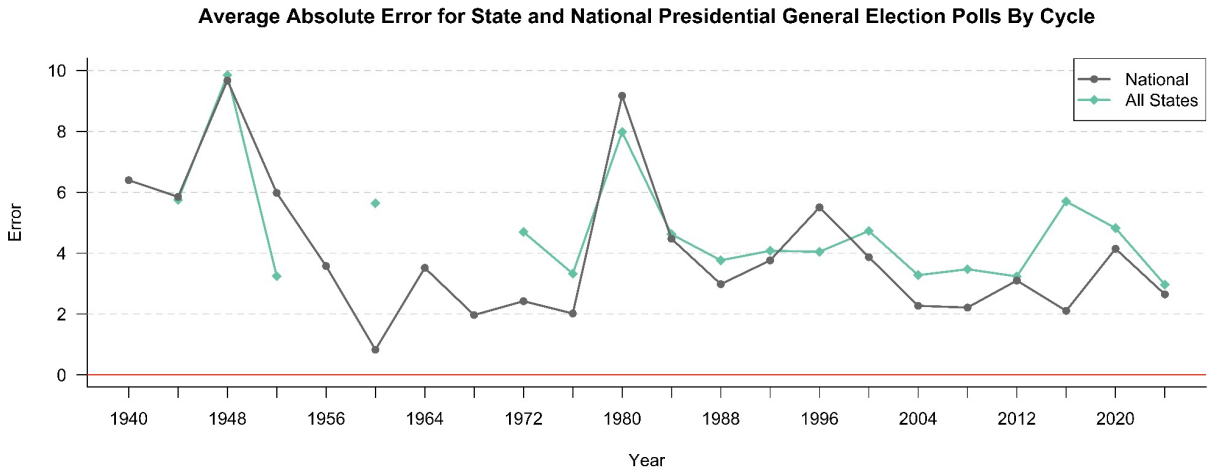

Figure 2.2.2: Average national and state-level absolute errors in Presidential general election margins since 1940 using all aggregated polls from the last two weeks of each election.

The modest overestimation of Democratic shares in 2024 comes amidst a long-term pattern of variability in polling bias. As shown in Figure 2.2.3, presidential election polls have historically oscillated in direction—sometimes overestimating Republican support (e.g., 1948, 2000, 2012), sometimes Democratic (e.g., 1964, 1980, 1996). Over most of the modern polling era, these signed errors have tended to balance out across cycles. What distinguishes the recent period is its persistence: 2024 marks the third consecutive presidential election in which Democratic support was systematically overestimated. This kind of directional consistency in presidential race errors has few precedents in presidential cycles.

However, midterm election polling has not shown the same pattern—signed errors in 2018 and 2022 were small and varied in direction. Midterm and presidential electorates differ in key respects, with far more marginal voters voting in presidential years than in midterms. This makes it a more predictable electorate for turnout modeling, and one that contains fewer voters who are difficult to reach to discover their voting preferences.

Whether this recent string of presidential-year Democratic overstatements signals a structural shift in polling error—or simply reflects temporary features shared across the 2016, 2020, and 2024 campaigns—remains unclear. All three of these presidential elections featured Donald Trump on the ballot as the Republican nominee. It is unclear if the type of challenges pollsters have faced in the past three elections will persist when Trump is not on the ballot.

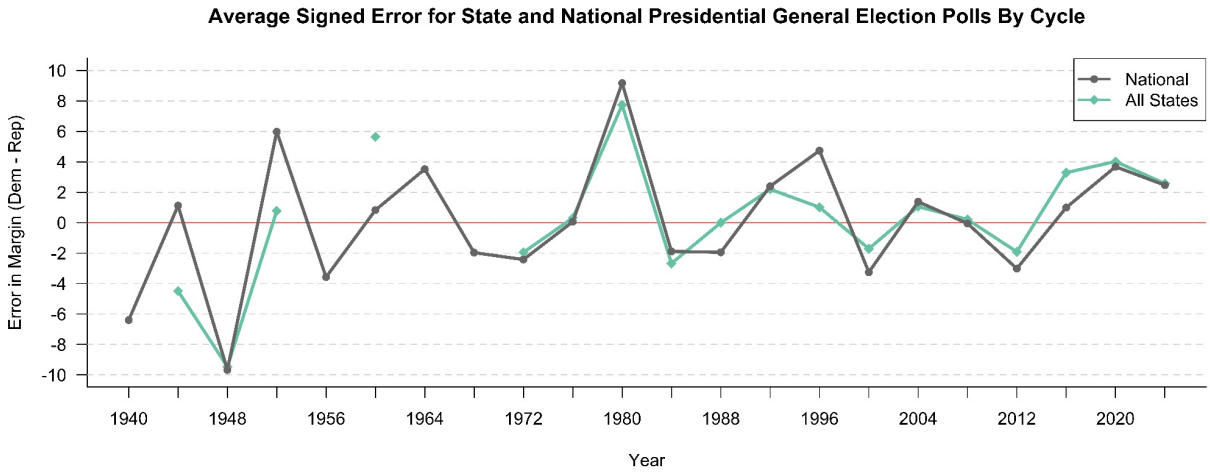

Figure 2.2.3: Average national and state-level signed errors in presidential general election margins since 1940 using all aggregated polls from the last two weeks of each election.

Performance also improved across contest types. In 2020, gubernatorial polls missed by 7.3 points on average and Senate polls by 6.0 points. In 2024, those figures dropped to 5.0 and 3.3 points, respectively. House polls were effectively unchanged, however, with the error edging down from 7.1 to 7.0 percentage points.

All results reported here use a consistent 14-day final field window, the same poll-cleaning rules, and the same benchmark vote files described in Appendix B. This allows for direct comparisons across years, rather than inference from shifting samples or metrics.

2.3 Firm-Based Correlates of Accuracy

While much attention in polling accuracy assessments focuses on methodological choices, differences across polling organizations themselves—independent of reported design choices—also shaped poll results in 2024. This section examines how firm-level characteristics such as partisanship, polling volume, and political specialization were associated with performance.

Among the most visible patterns was that Republican-affiliated firms produced slightly more accurate results on average than Democratic-affiliated or nonpartisan firms. This reverses the pattern seen in some prior cycles, where Democratic-aligned or nonpartisan firms had fared better. In 2024, the modest overstatement of Democratic performance in the broader polling environment meant that Republican firms, which historically project stronger Republican vote shares, often produced estimates that more closely matched certified outcomes.

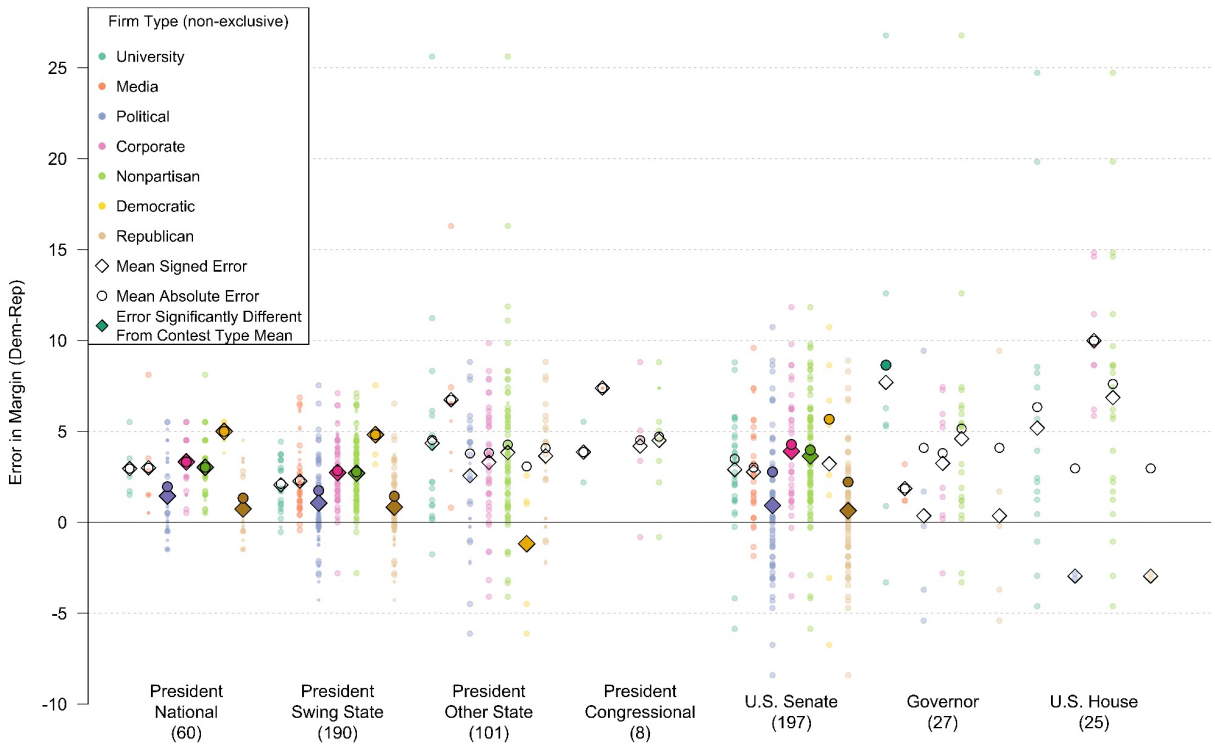

Figure 2.3.1: Absolute and signed errors for polls based on what type of firm was collecting the data (hand coded). Errors tended to be smallest for Republican and political firms, which likely reflects the tendency of these firms to provide estimates that were a little more favorable toward Republicans in a year with a Democratic overestimate. A similar analysis in 2022, a year without a Democratic overestimate, did not replicate this finding.

It was not just partisan affiliation that mattered. Firms that have a track record of conducting campaign polling—whether partisan or nonpartisan—also tended to outperform firms lacking this history. This advantage may stem from greater familiarity with likely-voter modeling, sample frame construction, or weighting strategies tailored to election contexts. These firms also fielded more polls overall, allowing for greater calibration over time.

Notably, the firms with the longest track records did not notch the best performance. Instead, the lowest errors came from firms that were newer, but were not tracking their first cycle. It's possible this effect is related to which firms survived other recent cycles or some success with emerging methods.

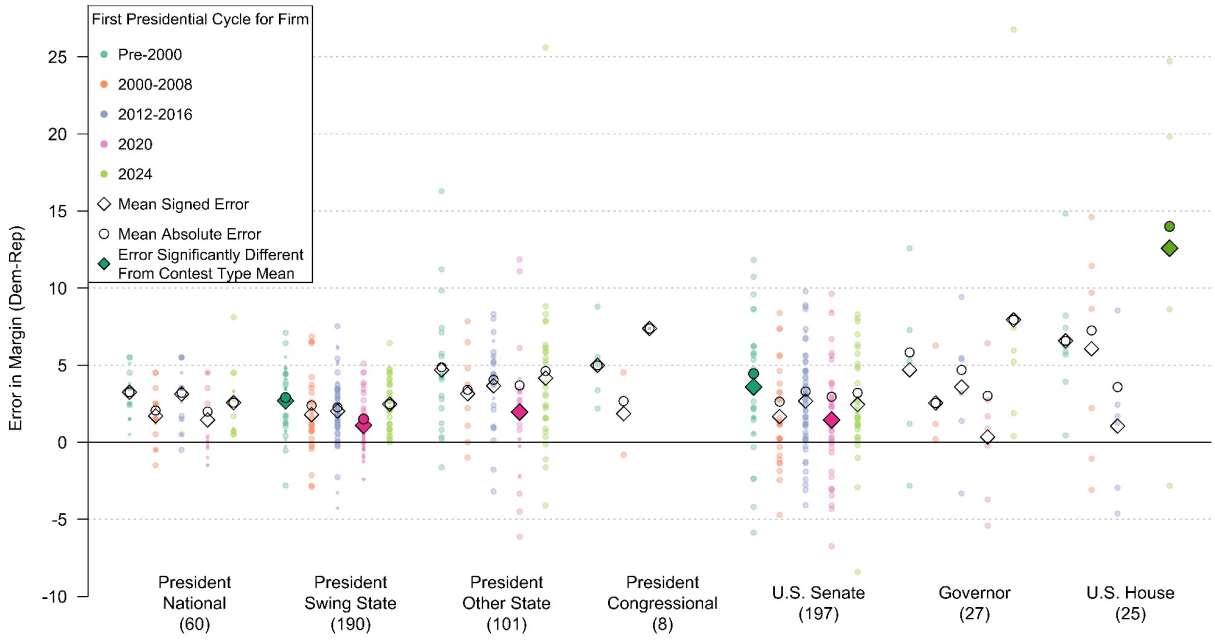

Figure 2.3.2: Absolute and signed errors for polls based on how long firms have been polling elections (firms identified and matched using aggregate data sources and keywords). Errors tended to be largest for firms that were present in 2024, but had not been present in 2020. Firms with the longest histories also had slightly larger errors than firms that have been around for a cycle or two.

Pollster volume also showed a weak but consistent relationship with performance. High-volume organizations—those that released many polls across contests and states—exhibited lower average error rates than firms that issued only one or two polls. These organizations often had greater resources, more robust approaches, and experience adapting their procedures to different races and electorates. That said, several low-volume firms also achieved very strong results. Volume alone was not determinative.

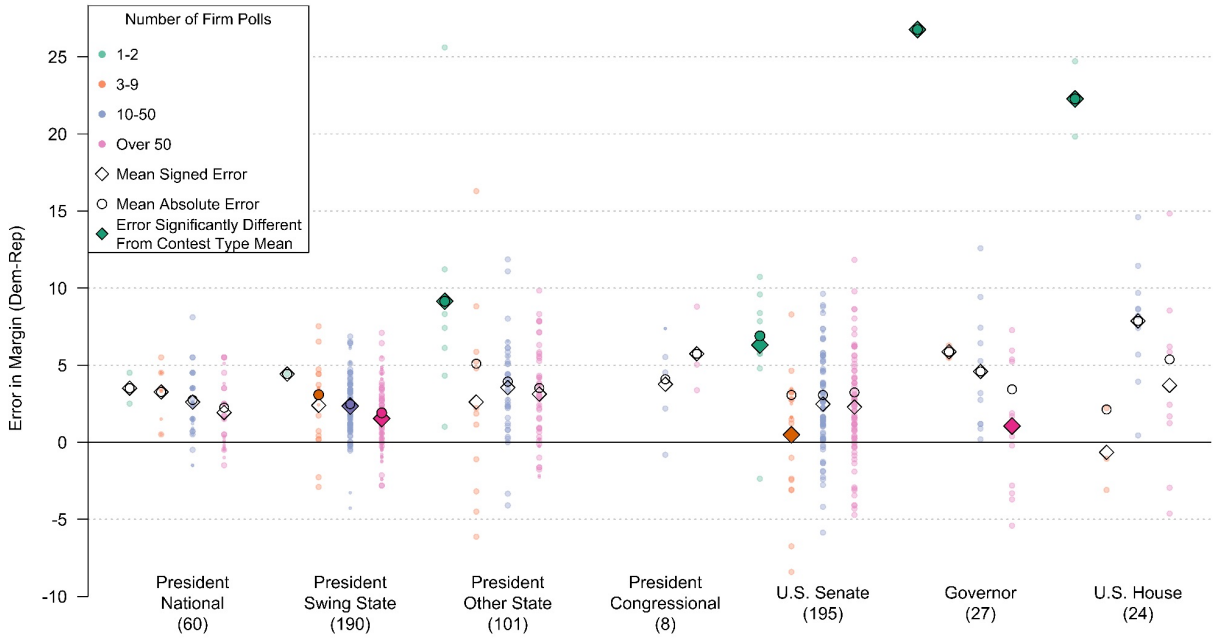

Figure 2.3.3: Absolute and signed errors for polls based on how many polls each firm released in the 2024 cycle. Firms that released fewer polls tended to have larger errors than those that released a larger number of polls, with the smallest errors among the most prolific pollsters.

Caution is warranted in interpreting these relationships. Many of the traits described above—partisan affiliation, pollster experience, and volume—correlate with methodological features such as sample source or weighting approach, which are addressed separately in Section 2.4. Moreover, not all firms disclosed enough metadata to categorize every design element, limiting our ability to fully disentangle firm effects from survey procedures.

Nevertheless, the 2024 data provide clear evidence that organizational characteristics—especially political alignment and experience—were modestly associated with polling accuracy. This underscores the importance of treating firm-level attributes as part of the broader accuracy landscape and points to continued value in tracking institutional patterns alongside methodological ones.

2.4 Methodological Correlates of Accuracy

The task force also examined whether specific methodological decisions were associated with better pre-election performance in 2024. These decisions included sampling frame, interview mode, weighting variables, and the type of likely-voter model used. Although some combinations were associated with slightly lower average error, no single methodological recipe guaranteed high accuracy.

Sampling strategy was not consistently related to errors. As shown in Figure 2.4.1, polls that used voter-file-based sampling frames occasionally performed slightly worse than those using

opt-in panels, or address-based samples. It is not clear why this would be the case given that voter-file based samples start out somewhat closer to the eventual electorate. It might be that firms using these samples rely too heavily on information in the voter file (e.g. partisan registration) or that they have more difficulty identifying which new or irregular voters will participate. Polls using random digit dialing (RDD) sometimes performed quite well and sometimes produced relatively large errors.

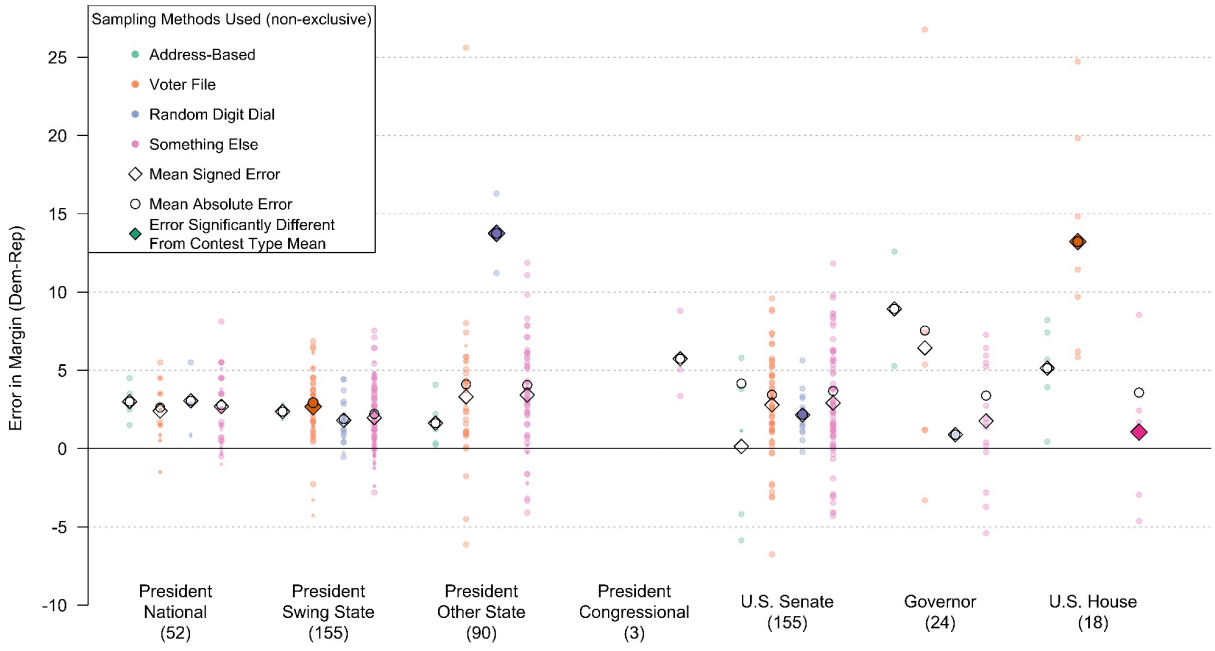

Figure 2.4.1: Absolute and signed errors for polls based on which sampling strategy each firm used in the 2024 cycle (committee coded). Slight evidence that voter file samples performed worse for some estimates.

Mode of data collection also showed weak associations with error. On average, surveys that included at least some interactive voice response interviews (IVR) were the most accurate for national and swing state presidential polls, but errors when using this method were not statistically significant from when it was not employed, and no surveys solely relied on IVR; all combined its use with at least one other method.

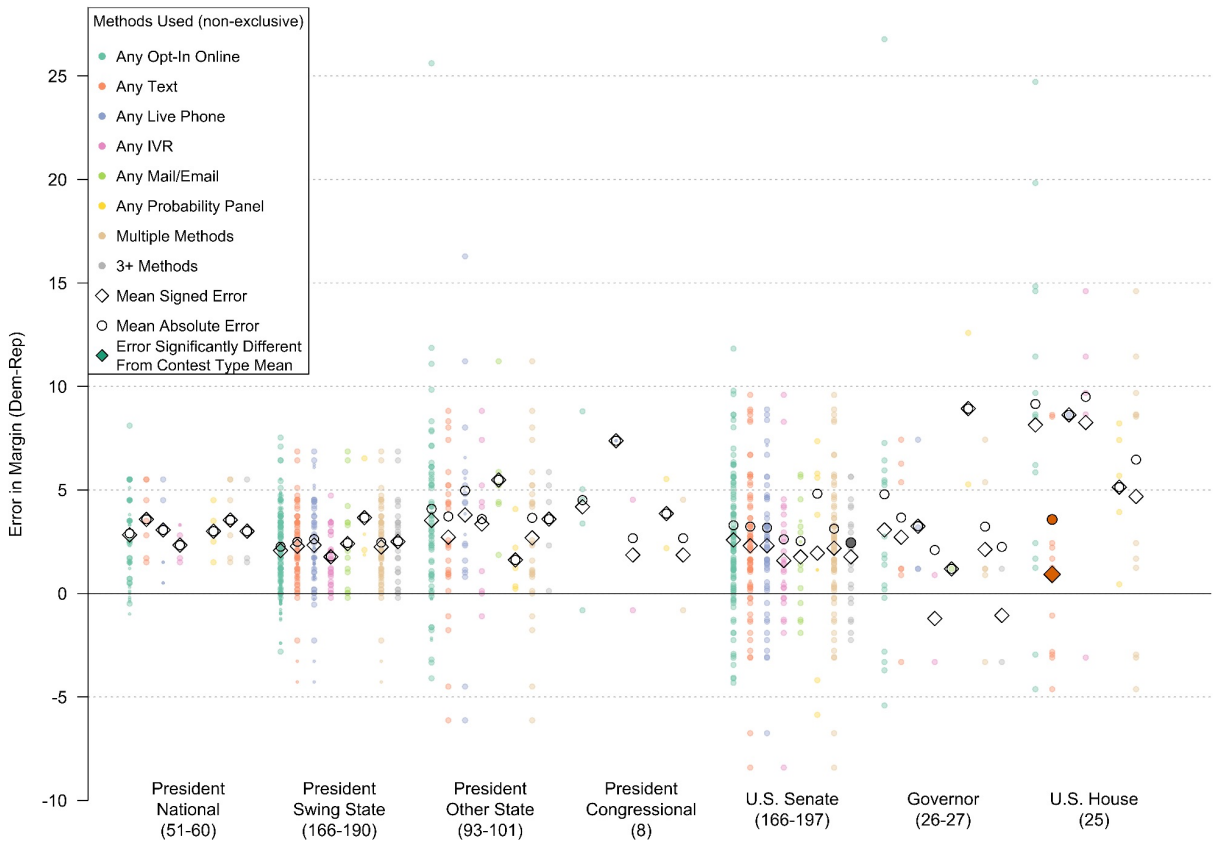

Figure 2.4.2: Absolute and signed errors for polls based on which methods each firm used in the 2024 cycle (FiveThirtyEight Coded). No consistent patterns in which methods are related to better polls.

Weighting on partisanship was associated with modestly improved accuracy. Polls that included party identification—alongside demographic factors such as age, education, and race—in their post-stratification routines tended to show slightly smaller average absolute errors. As in previous cycles, it remains unclear whether these gains reflect the effectiveness of party weighting itself or the broader set of additional practices adopted by firms who use it.

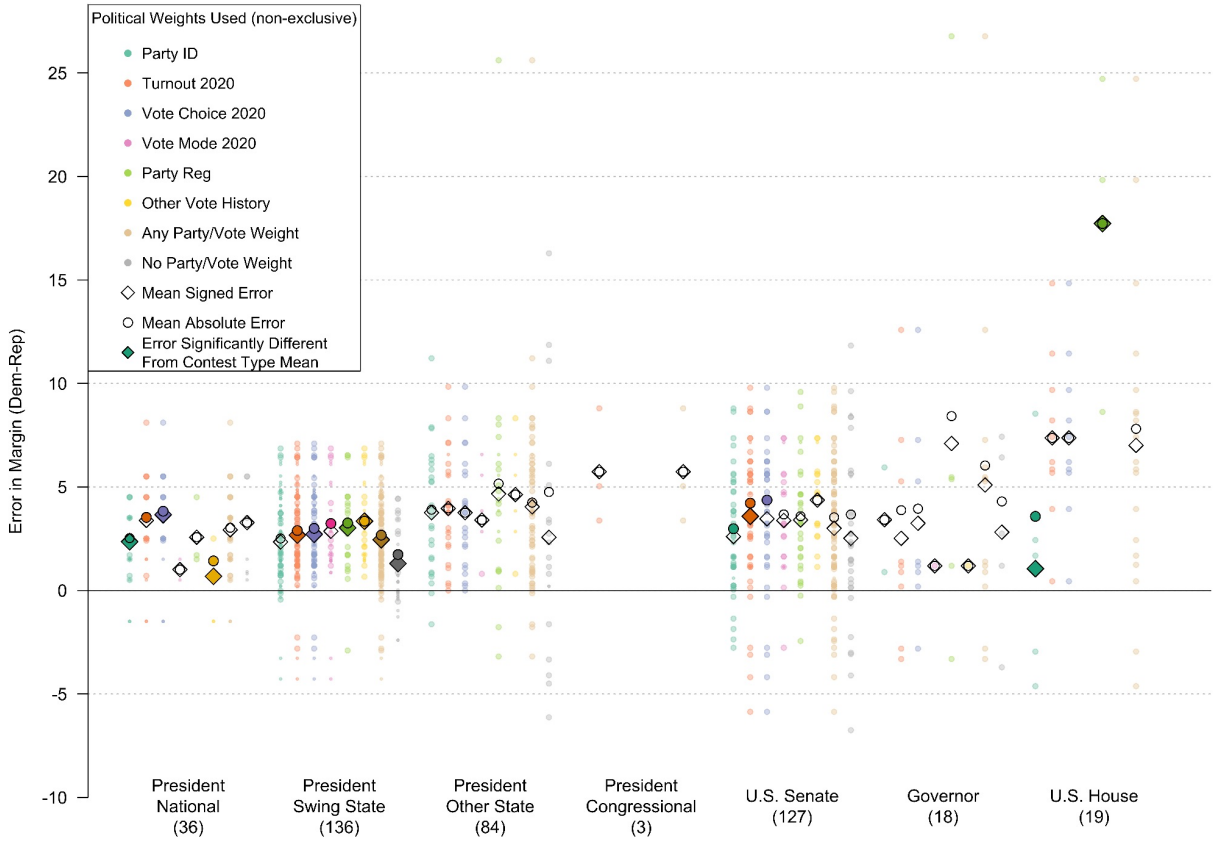

Figure 2.4.3: Absolute and signed errors for polls based on use of different political weighting variables (committee coded). Weighting on self-reported partisanship sometimes outperformed other methods, weighting on party registration from the voter file sometimes performed worse.

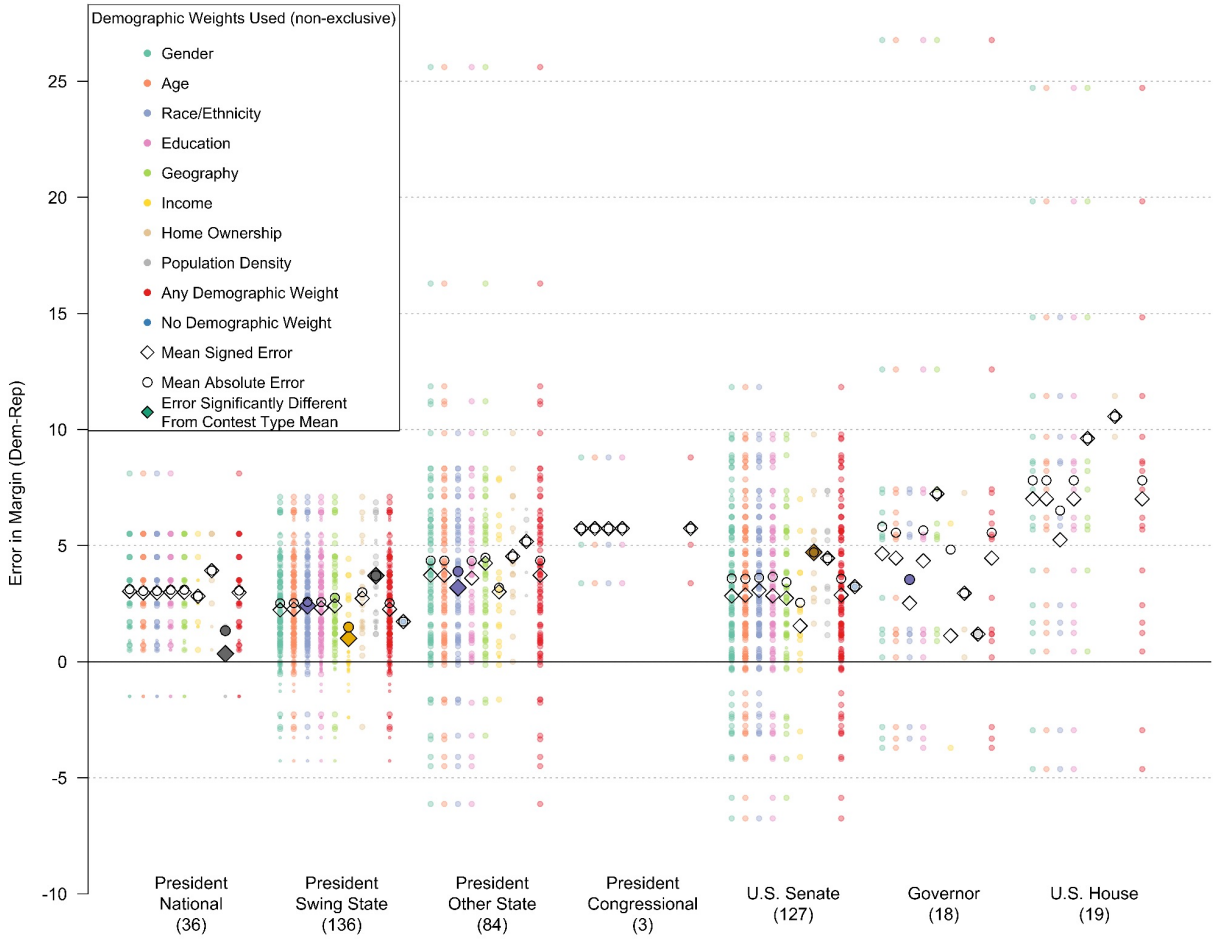

Figure 2.4.4: Absolute and signed errors for polls based on use of different demographic weighting variables (committee coded). Weighting on respondent race/ethnicity and income was sometimes associated with higher accuracy polling. Note that many polls used a common set of variables, resulting in similar estimates across weighting strategies.

Likely-voter modeling showed more pronounced differences. Polls that used multivariate turnout models or registration-based screens had somewhat lower average errors than those relying on simple self-reported likelihood-of-voting questions or those that reported results for all adults or registered voters, at least among those firms that completed the task force survey. The advantage of these more complex models likely stems from their ability to better approximate actual turnout—particularly in a year with shifting participation patterns and a significant share of voters who had not turned out four years earlier. Firms that categorized respondents as having a probability of voting were also more accurate than those that treated respondents

dichotomously as either likely voters or as likely non-voters.

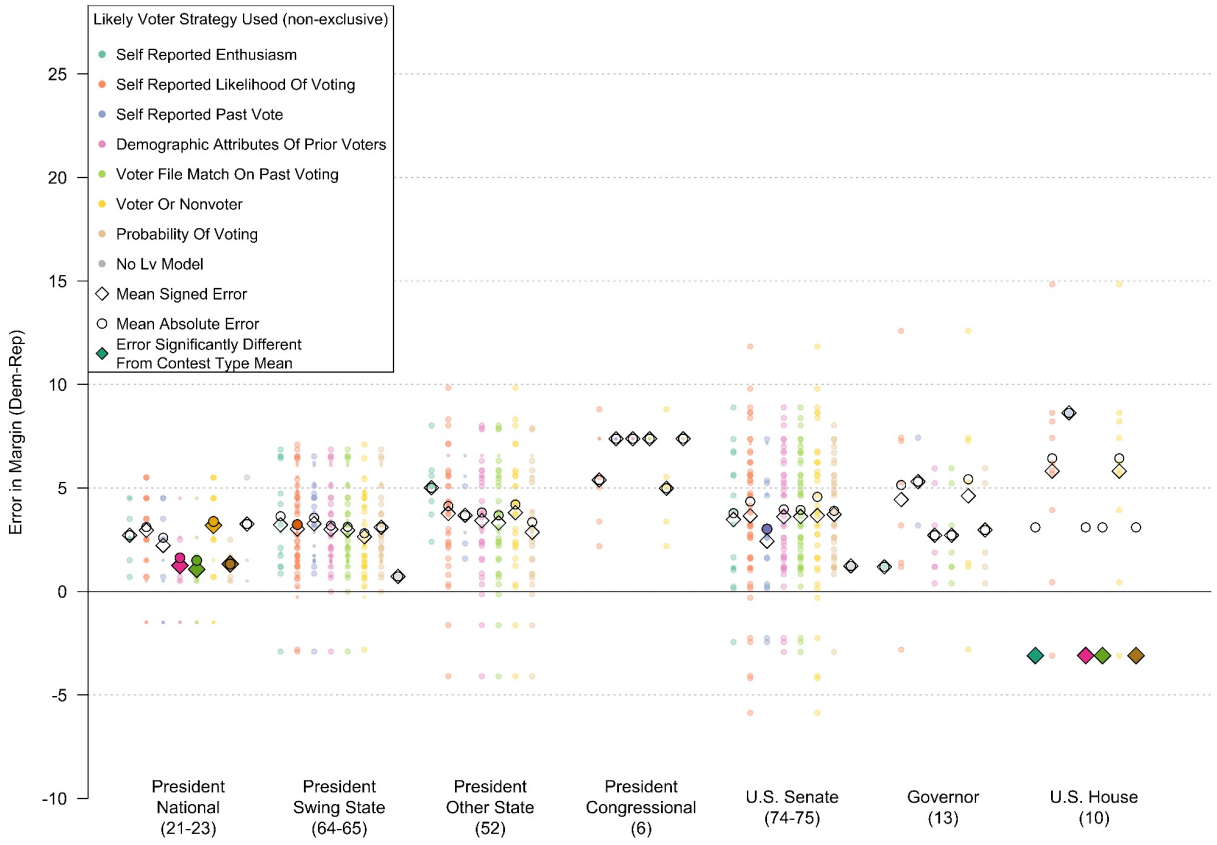

Figure 2.4.5: Absolute and signed errors for polls based on use of different likely-voter indicators (from firm survey). Firms that modeled attributes of voters or that used voter file records for identifying likely voters were more accurate than those that used self-report metrics. Firms assigning a continuous probability of voting were also more accurate than those using dichotomous classifications.

Methodological choices rarely operate in isolation. Firms that used voter files were also more likely to weight on partisanship and employ detailed likely-voter models, making it difficult to isolate the effect of any single decision. Moreover, firms differ in implementation—two surveys may both weight on partisanship, for example, but do so using different target distributions or variable definitions.

These patterns align with findings from prior AAPOR reports: design decisions can contribute to improved performance, but their effects are often conditional on execution, context, and interactions with other features. In 2024—as diversity of methods increased but more pollsters incorporated political variables—the effects of variation across methods was narrower than in past cycles.

2.5 Geographic Accuracy and Turnout Shifts

While headline polling accuracy improved in 2024, performance was uneven across states—and within them. Figures 2.5.1 through 2.5.4 present a comprehensive view of state-level polling error, its geographic distribution, historical comparison, and relationship to reported margins of error for the presidential contest.

Final two-week presidential polls varied substantially in their signed error across states. As shown in Figure 2.5.1, errors tended to lean modestly Democratic, with most state-level polls overstating Harris's support relative to certified returns. This directional lean was consistent with the national pattern, but some states—including several pivotal battlegrounds—exhibited notably smaller or even pro-Republican signed errors.

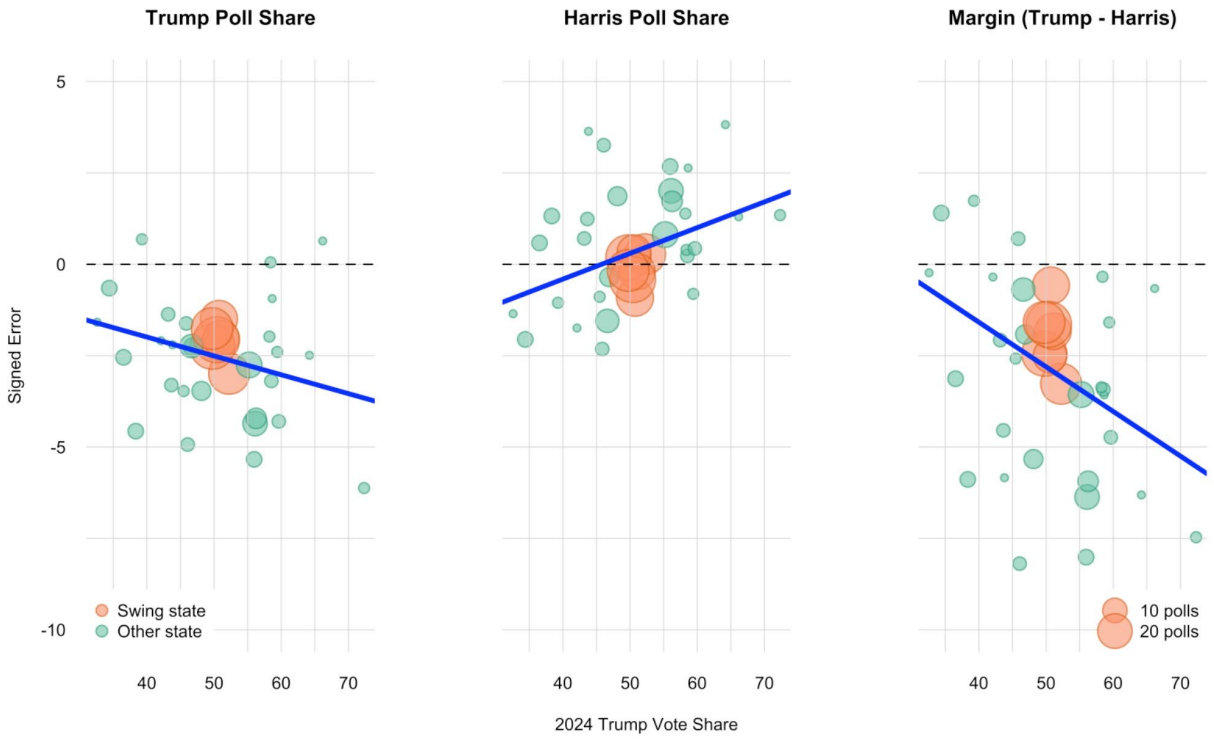

Figure 2.5.1: Distributions of signed errors for candidate estimates across states, regression line weighted by the number of polls in each state. Polls in states that supported Trump were more likely to underestimate his support compared to Harris support.

Figure 2.5.2 maps signed errors across the 50 states, illustrating regional clustering of polling performance. While battleground states generally showed tighter error margins, less-polled states often exhibited wider variability. This geographic imbalance reflects both methodological challenges and the allocation of polling resources: heavily polled states benefit from larger sample sizes, more frequent releases, and likely additional scrutiny given the importance of those contests. In contrast, low-frequency states are often covered by fewer smaller-sample surveys whose designs may be less transparent or harder to validate.

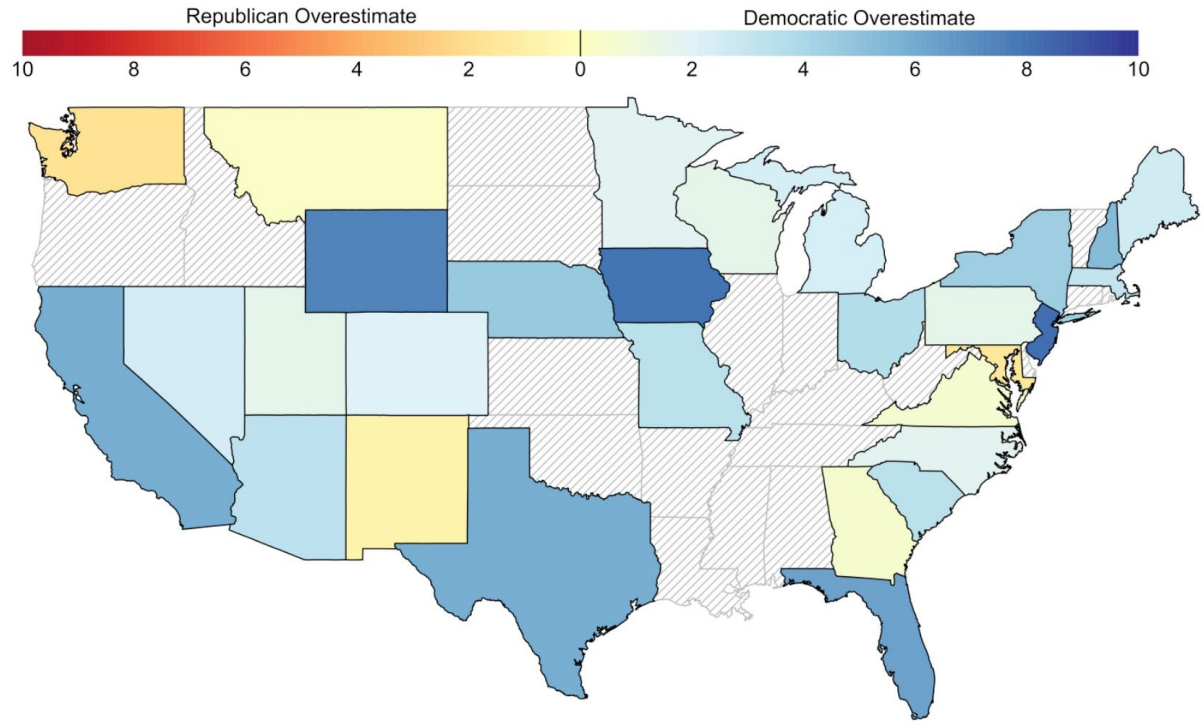

Figure 2.5.2: Distributions of signed errors by state.

To put these patterns in context, Figure 2.5.3 compares 2024 error by state to 2020. Polls in 19 states and the nation as a whole were significantly more accurate in 2024 than they had been in 2020, versus only four states where the results were significantly worse. No states saw polling that erred in 2024 by more than 10 percentage points on average, even though six had done so in 2020.

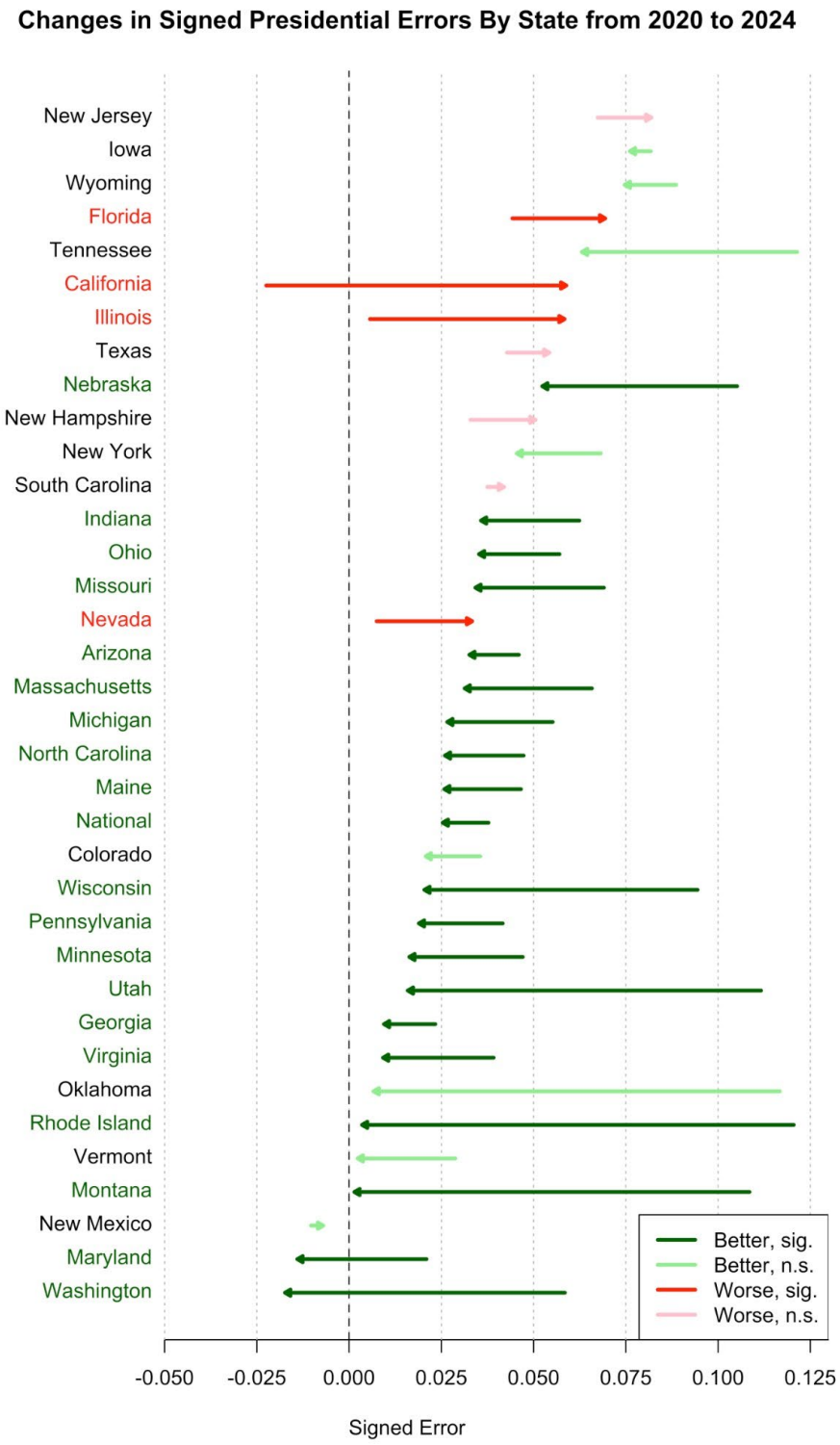

Figure 2.5.3: Changes in signed error from 2020 to 2024 by state. Positive signed errors indicate overestimations of Democratic vote shares in polls, where negative signed errors indicate overestimations of Republican vote shares.

Despite these errors, the vast majority of polls in 2024 correctly forecasted how states would perform in the presidential election. Figure 2.5.4 shows how estimates from each individual poll as well as the overall polling average differed from the eventual results by location. For all but three locations, a majority of polls forecast the correct election winner; the overwhelming majority also accurately indicated whether the election was likely to be close or not.

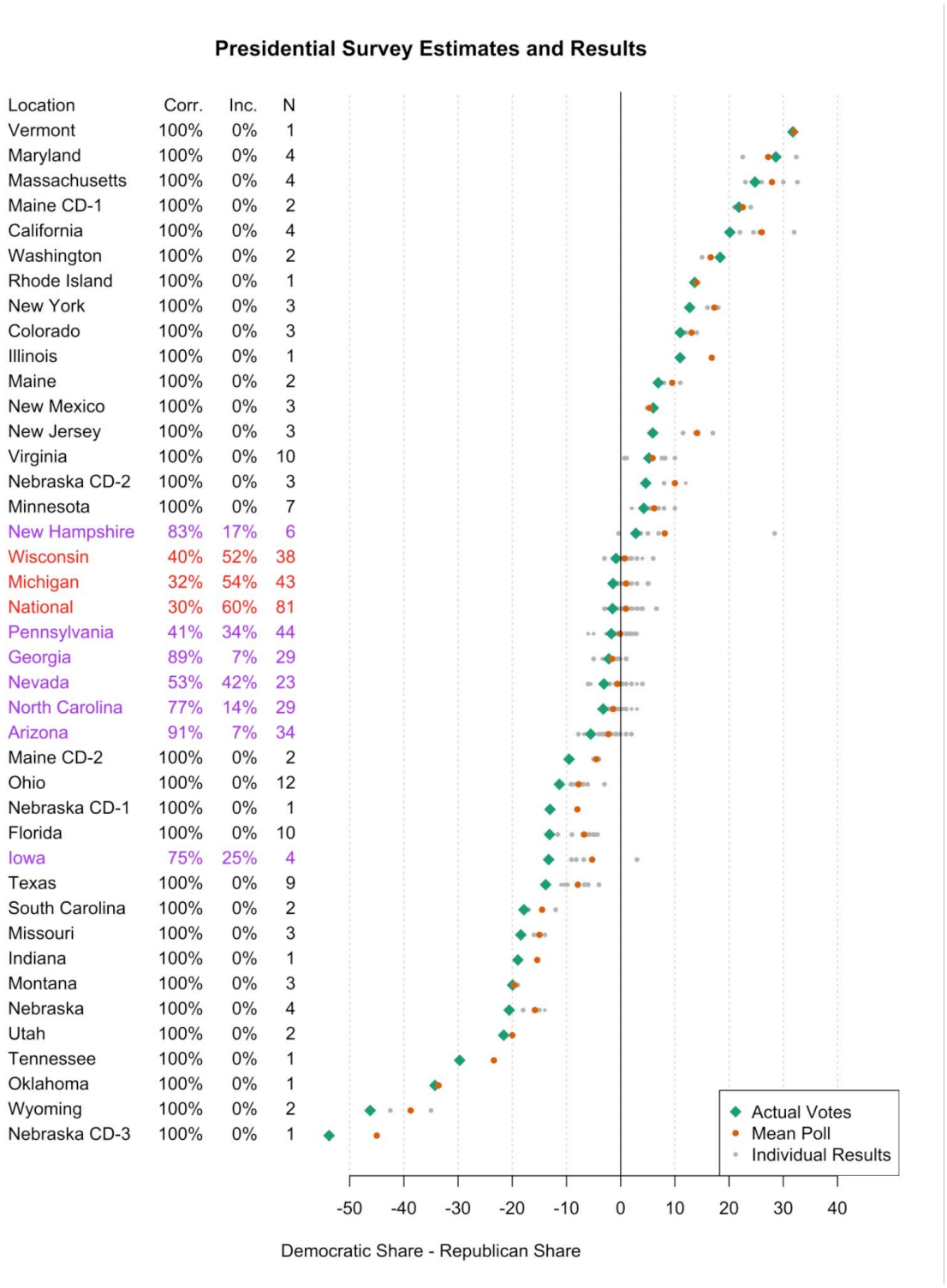

Figure 2.5.4: Signed errors in 2024 by state. States shown in purple had at least a few reported results which showed an incorrect election winner. States shown in red indicate cases where the plurality of pre-election polls said that the wrong candidate was leading.

State-level differences, however, only tell part of the story. The most granular errors often stemmed from misrepresenting local turnout patterns. As described in earlier sections, Republican turnout surged in rural and exurban counties, while Democratic turnout fell in some urban centers. Most polls assumed that the 2024 electorate would resemble that of 2020, and only a few incorporated local geographic weighting or regionally stratified turnout projections. As a result, surveys often missed where the electorate was shifting—especially in high-stakes battlegrounds like Georgia, Pennsylvania, and Nevada.

To better understand these dynamics, the microdata shared by participating pollsters enabled comparisons between the geographic distribution of survey respondents and certified county-level turnout. Figure 2.5.5 compares weighted respondent distributions from national and state polls with actual 2024 turnout, grouped by the 2020 partisan lean of each county. It shows that counties that voted heavily for Biden in 2020 were overrepresented in the polls relative to their actual turnout in 2024, while Trump-leaning counties were underrepresented. These disparities help explain why modest Democratic overstatements persisted in many states, even when the marginal errors were low overall.

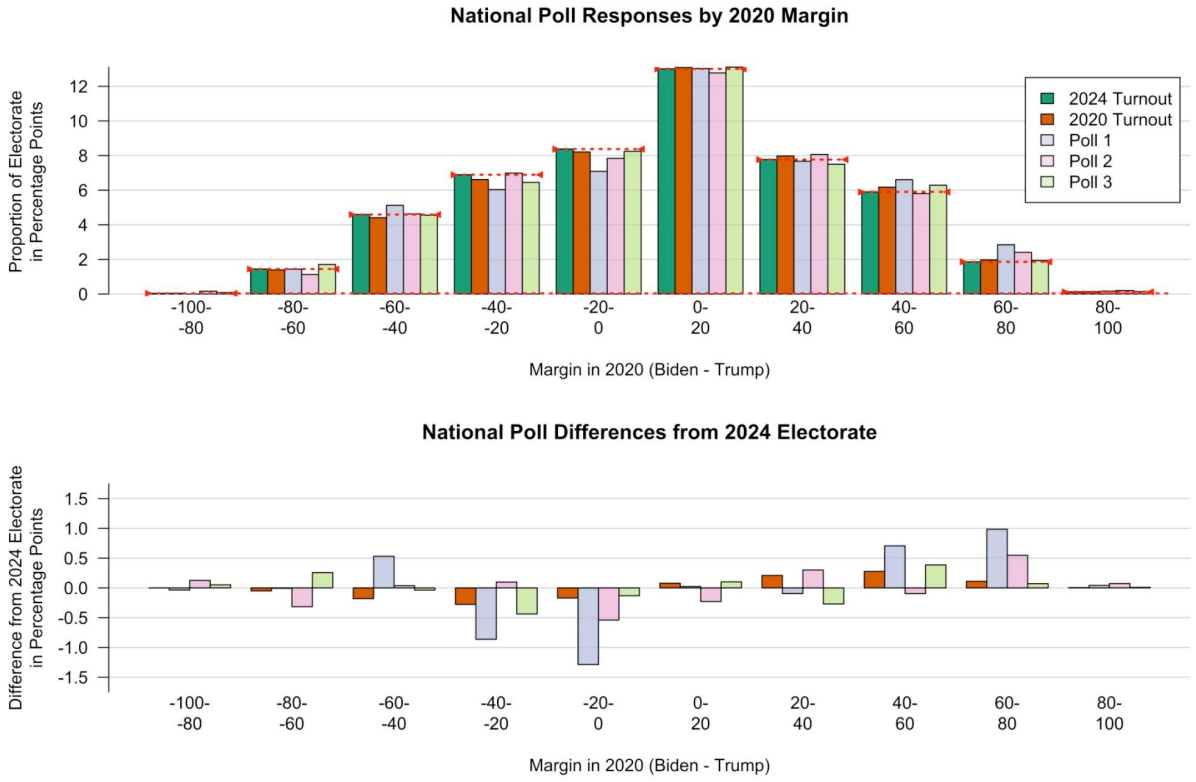

Figure 2.5.5 - Comparing turnout in 2024 with turnout from 2020 for three county-matched national microdata polls for each 10-percentage-point margin range of county preference. Top plot shows weighted representation across each sample. Dashed lines allow for comparisons with 2024 turnout levels. Bottom plot shows differences in percentage points from 2024 turnout. Polls indicate overrepresentations of voters from counties that leaned toward Biden in 2020 compared to eventual 2024 turnout by county.

A second analysis, shown in Figure 2.5.6, further illustrates how polling estimates diverged from actual outcomes depending on the partisan makeup of counties. This figure compares average candidate support in the 2024 polls with both certified 2024 vote margins and 2020 benchmarks, grouped by deciles of county-level Democratic vote share in 2020. Within each decile, the figure displays how weighted respondent preferences differed from the true average preference in the counties where those respondents lived. The results show that polls consistently overstated Democratic support in the most Republican-leaning counties and slightly overstated Republican support in the most Democratic-leaning ones—patterns that held across national, state, and district-level surveys. These imbalances suggest that even when polls were geographically representative in aggregate, differential representation or weighting within partisan geographies remained a source of error.

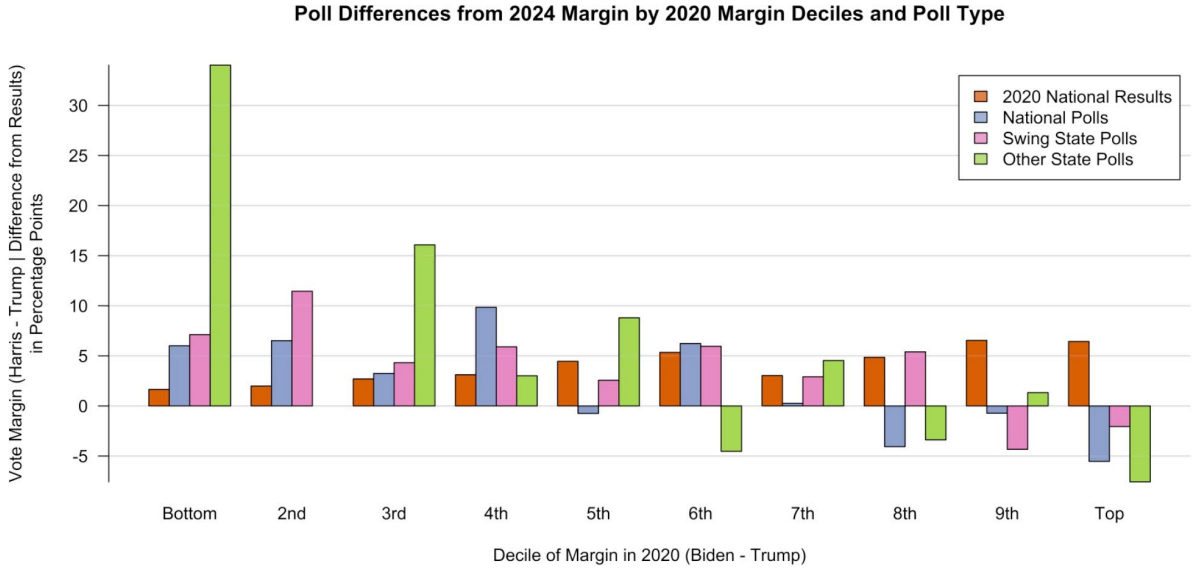

Figure 2.5.6 - How election margins in 2024 compared with margins in 2020 and poll results by 2020 county-level preference margin for different types of polls. Deciles are defined nationally. Within each decile, respondents' weighted average preferences are compared to the average preference that would be expected based on vote distributions in the counties where those same respondents lived. Higher decile numbers indicate greater Democratic support.

Future survey designs may benefit from more fine-grained geographic modeling, particularly in swing states where turnout asymmetries can tip statewide outcomes.

2.6 Subgroup Accuracy and Representation

Although the electorate overall was modeled with reasonable accuracy, polls consistently underrepresented or mischaracterized three critical blocs—all of which contributed to the modest Democratic overstatement:

1. **Republican voters in Republican-leaning counties**, who were underrepresented relative to Democrats in those areas.
2. **Hispanic voters**, whose levels of Democratic support were overstated in most surveys relative to post-election surveys and voter file data.
3. **2020 nonvoters** who cast a ballot in 2024, many of whom leaned Republican and were undercounted or modeled incorrectly.

The most consequential of these issues was partisan imbalance within geography. As discussed in Section 2.5, Republican voters in rural and exurban areas were harder to reach than their Democratic neighbors, even though they lived in the same counties. That imbalance created subtle but persistent distortions—polls that achieved regionally representative samples still might not accurately capture *who* within those regions would turn out and vote.

Hispanic voters, meanwhile, posed a different challenge. Surveys correctly identified the direction of change—Trump made gains among Hispanic Americans in 2024—but they underestimated the size of the shift. Most polls showed Harris winning Hispanic voters by a comfortable margin, but validated voter studies and post-election surveys suggest that the margin was narrower than pre-election polls reported. This mirrors patterns observed in 2020 and raises questions about language, frame coverage, and trust in polling among segments of the Hispanic electorate.

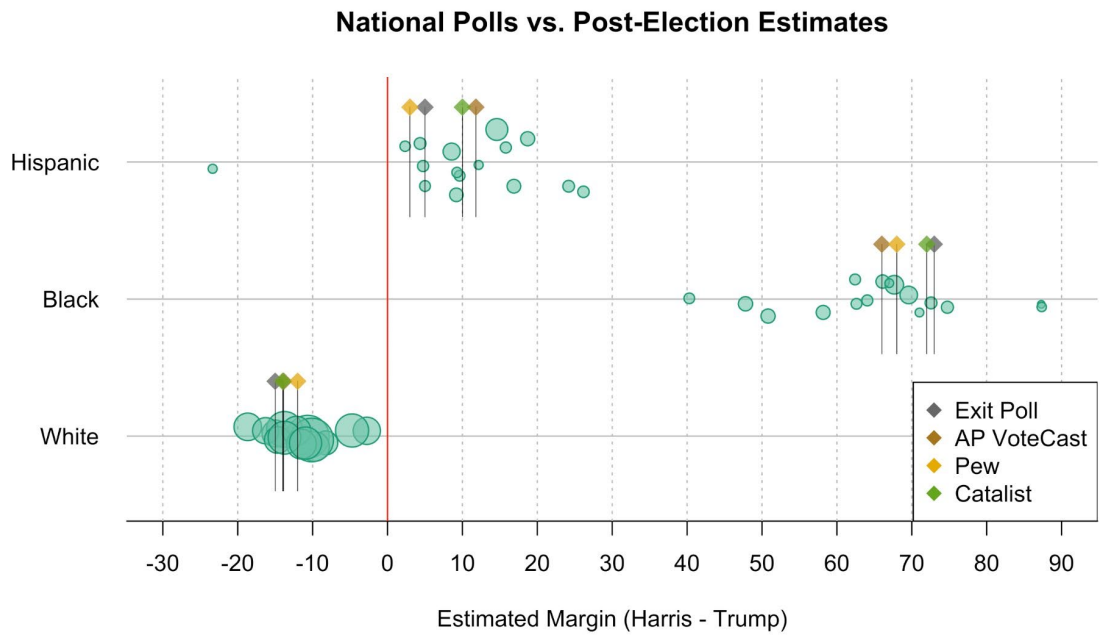

Figure 2.6.1 - Comparing pre-election poll group estimates from microdatasets with alternate benchmarks for racial and ethnic group voting preferences using firm-provided weights. Benchmark post-election studies suggest that polls overestimated Hispanic voter preferences for Harris relative to Trump.

Finally, polls struggled to fully capture 2024 voters who had not voted in 2020. These voters were especially common in battleground states. While some polls detected their partisan lean, many underestimated their share of the electorate. Because 2020 nonvoters tend to differ from habitual voters on engagement, demographics, and survey responsiveness, missing them can distort overall estimates even when models are well-calibrated for known populations.

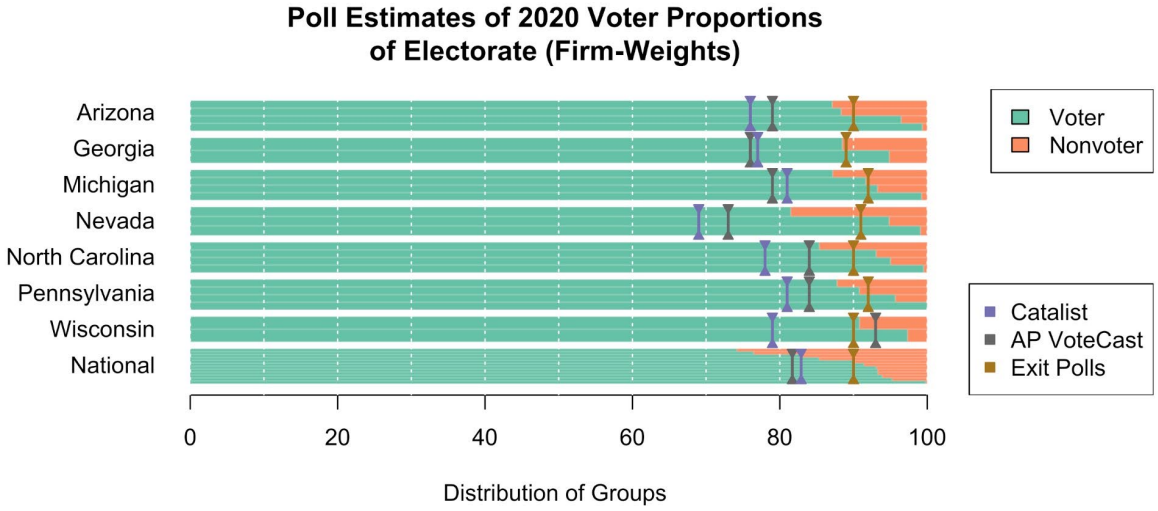

Figure 2.6.2 - Estimated proportions of the 2024 electorate that had voted in 2020 using firm-provided weights for microdata. Within each state, each row shows the distribution for a separate poll. Vertical bars provide benchmark comparison estimates from Catalist voter file data, AP VoteCast, and exit polls.

Across all three groups, the problems were less about identifying *preferences* than about coverage and composition. Most polls correctly detected that rural voters and 2020 nonvoters leaned Republican and that Hispanic support for Democrats had declined—but they misestimated either the extent of these shifts or how many voters in those categories would turn out.

Future work may benefit from better integrating administrative turnout indicators into sampling and weighting routines. Improved frame coverage, oversampling of hard-to-reach groups, and more nuanced likely-voter modeling could also help correct these persistent gaps.

2.7 Pollster-Level Variation in Subgroup Estimates

Even when pollsters reached similar conclusions about the outcome of the race overall, they sometimes told different stories about the electorate's composition and subgroup voting behavior. In 2024, most surveys showed Kamala Harris and Donald Trump locked in a close contest, with little disagreement about the national or swing-state margins. Yet when those same surveys reported breakdowns by age, race, education, and partisanship, the differences between pollsters were striking.

The most visible example was Hispanic voters. Some surveys showed Harris winning this group nationally by 25 points; others showed her behind. While polls agreed that Trump made gains among Hispanic Americans compared to 2020, the size of those gains varied substantially. The dispersion was not limited to Hispanic voters: estimates of the vote margin among young adults, college-educated Whites, and independents also varied by 15–20 points across firms in the microdata provided to the task force.

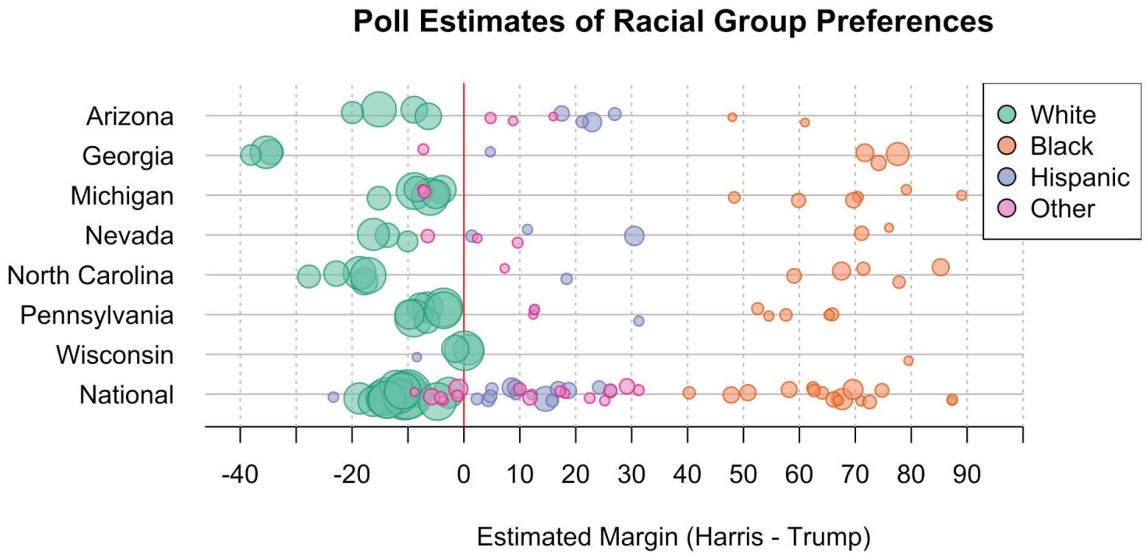

Figure 2.7.1: Estimated voting margins of racial groups in polls using firm-provided weights with microdata. Each dot represents a single poll's estimate for the margin within a particular racial or ethnic group. Dot size corresponds to within-group sample size. Groups with Ns<50 not shown.

We might expect greater variations in results for these groups simply because the sample sizes are often smaller than the samples used for estimates of the entire electorate. But these larger gaps between polls were not driven by statistical noise alone. Some reflect real differences in how respondents were asked about their preferences, weighting strategies, sampling frames, or how respondents were grouped and categorized. In other cases, firms simply had smaller sample sizes for certain subgroups, making their estimates more volatile.

To better understand this variability in subgroup vote margin estimates and how weighting affected the range of estimates, Figure 2.7.2 presents a set of boxplots comparing results across the microdata collected from national surveys, grouped by key demographics—race, education, age, and gender. Each row shows the range of estimates across polls at three stages: unweighted raw responses, firm-level weighted preferences, and post-hoc reweighted results designed to align with election results. While firm-weighted estimates typically narrowed group differences and brought them closer to expectations, substantial variation remained across pollsters—especially for subgroups like Hispanic voters, those with less than a college degree, and younger voters.

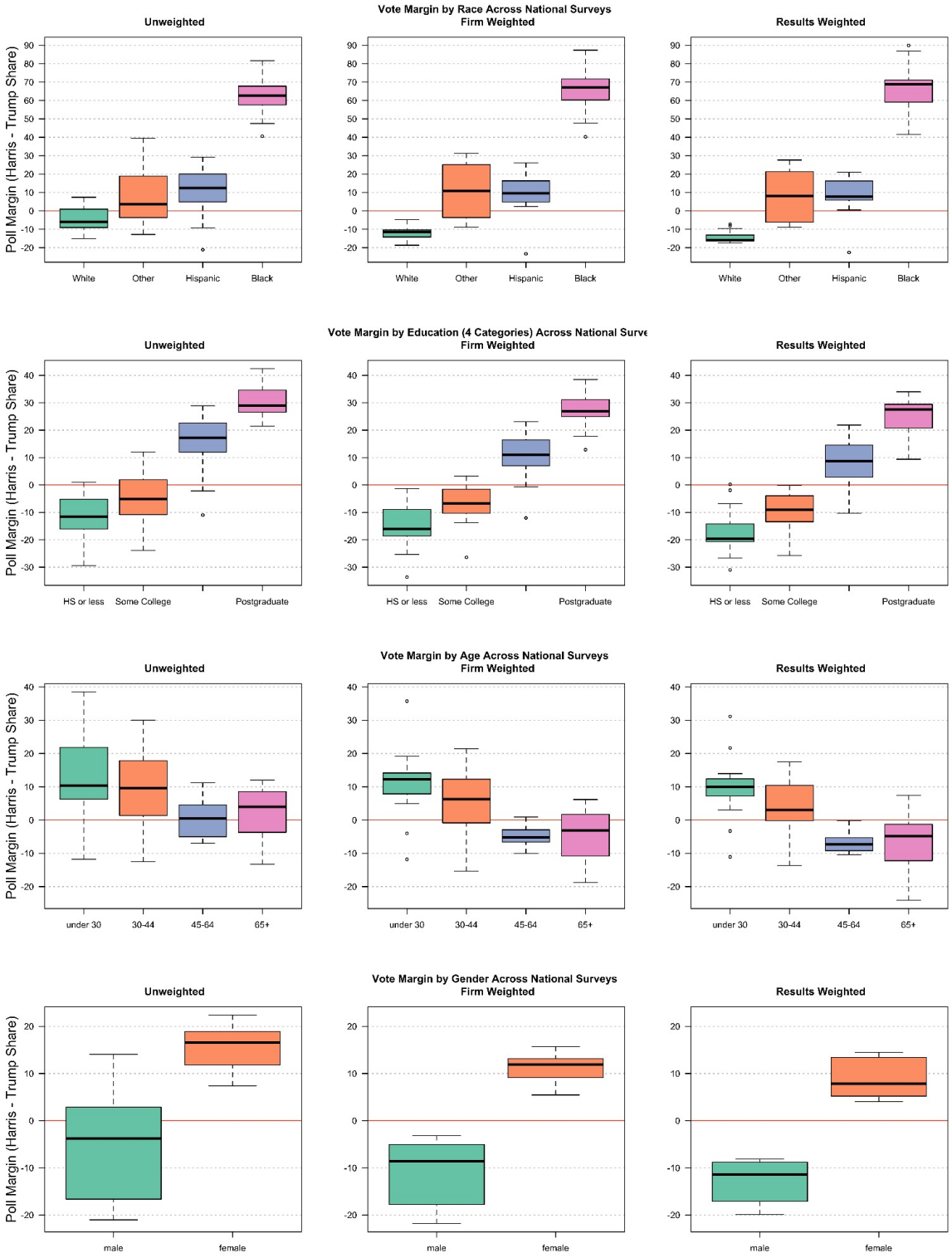

Figure 2.7.2 - Estimated voting margins of demographic groups in polls using raw data, firm-provided weights, and weighting to results with microdata. Distributions of group preferences across polls depending on weighting strategies. Categories with N<50 omitted from estimates.

Importantly, most pollsters agreed on whether a group leaned Democratic or Republican. There was some disagreement on the *magnitude* of those differences. This distinction matters: while it means polling was still informative for identifying shifts and trends, it also suggests that subgroup-level estimates remain far less reliable than overall margins.

At least for now, analysts and the public should interpret subgroup crosstabs cautiously, especially when comparing across pollsters or drawing fine-grained conclusions.

2.8 Herding and Poll Agreement

One of the most noticeable features of the 2024 polling cycle was the tight clustering of poll margins in competitive races. In many battleground states, most surveys fell within a narrow band, raising concerns that firms might have adjusted their results to align with others. This kind of convergence, often referred to as herding, can artificially understate uncertainty and mask important methodological differences.

To explore this possibility, poll-to-poll dispersion was calculated within each contest: the standard deviation of poll margins after subtracting the contest average. Among presidential polls in battleground states with at least five releases in the final two weeks, the average within-state dispersion was 2.3 percentage points, compared to 3.3 points in 2020 and 3.5 points in 2016. This narrowing was particularly evident in states like Pennsylvania, Michigan, and Wisconsin.

The overall dispersion of poll results declined in 2024, particularly in presidential swing states. As shown in Figure 2.8.1, the average standard deviation of poll margins within contests dropped by nearly a full point from 2020 to 2024. While some variation remains—especially in down-ballot races—the tighter clustering in top-tier contests signals a broader shift toward design convergence. The reduction in variability is presented in Figure 2.8.1, and reductions are similar when individual polls are compared to other temporally similar polls across the entire 2024 election cycle as well.

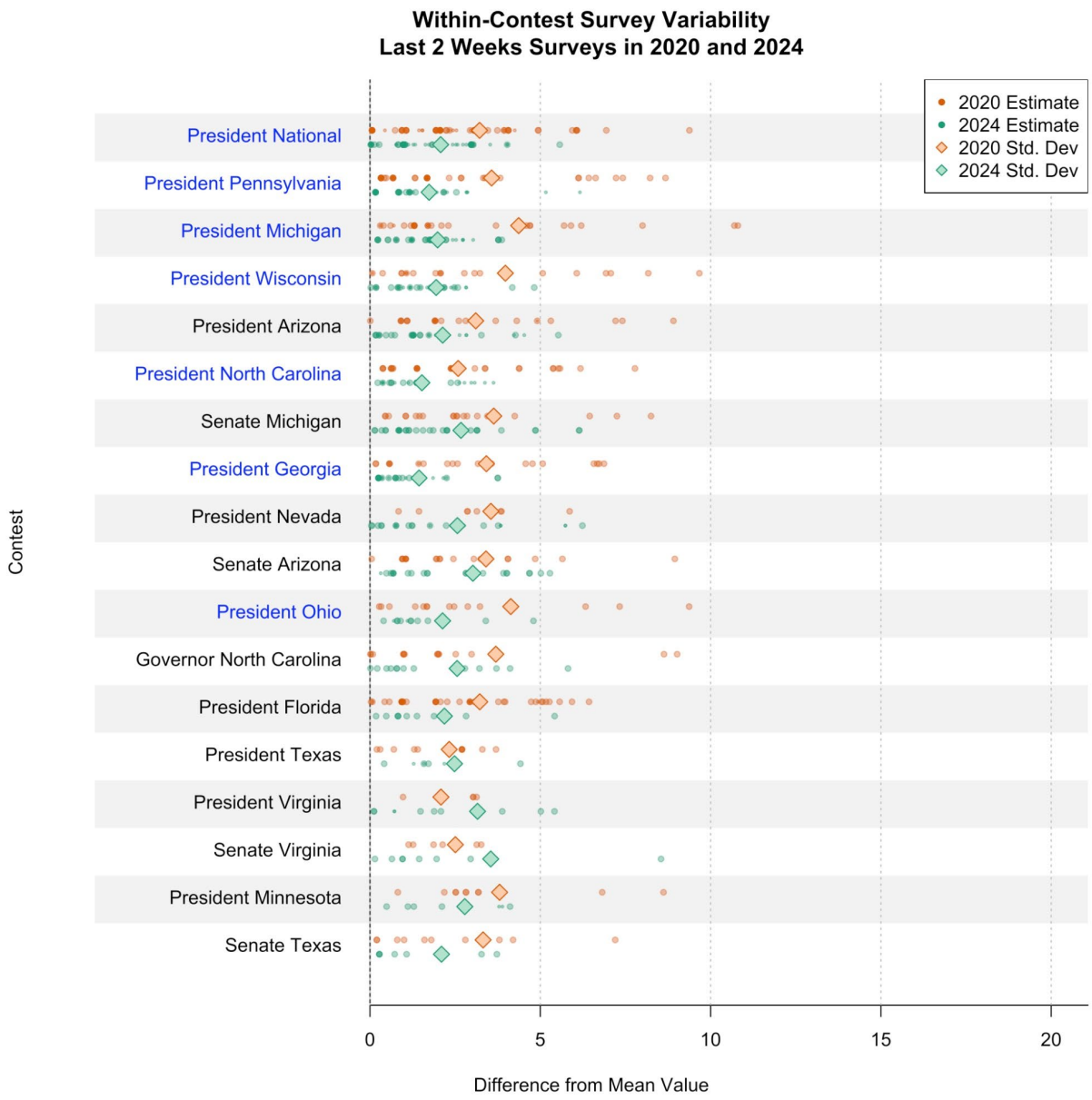

Figure 2.8.1 Variability in estimates from surveys around the average survey estimate from the last two weeks in each contest in 2020 and 2024. Contests sorted based on the number of matchups reported from largest to smallest, significant differences (F-test comparisons of variance) indicated with blue text.

To assess whether this reduced variation stemmed from strategic convergence, a herding test was applied. For each poll, the deviation from the 7-day moving average of other polls prior to its release was computed and compared to the deviation from the 7-day average of polls released afterward. If pollsters were adjusting their estimates to match existing results, the distance from the prior average would be significantly smaller than the distance from future releases for these firms. In 2024, no such pattern emerged. Differences between pre- and post-release deviation were statistically indistinguishable, suggesting that poll results were not artificially nudged toward the consensus.

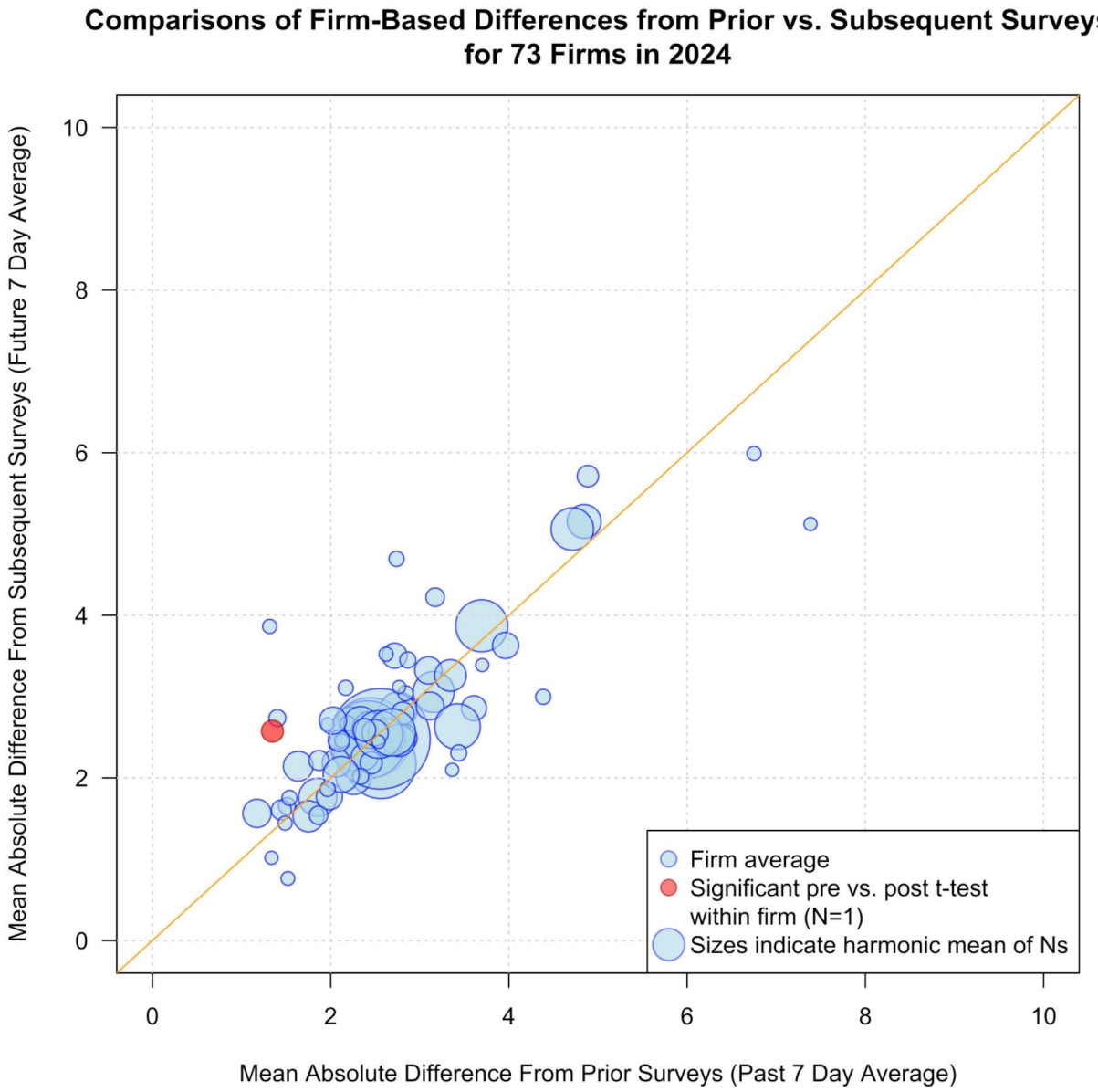

Figure 2.8.2 How surveys from each firm compare to prior versus subsequent results from other firms conducted for the same contest. Only one firm produced results that were significantly more similar to prior survey releases than to subsequent ones. Results are replicated when controlling for house effects as well.

The most plausible explanation for this reduced spread is increased methodological convergence. More pollsters used voter-file-based samples, weighted on partisanship or past vote, and implemented complex likely-voter models. These approaches tend to produce more stable and consistent toplines—even across firms—without any need for coordination. Indeed, when the task force conducted an experiment independently reweighting the raw microdata samples to match common demographics and 2020 margins, variability across samples notably

decreased and generally matched the variability in the data when using firm-provided weights.

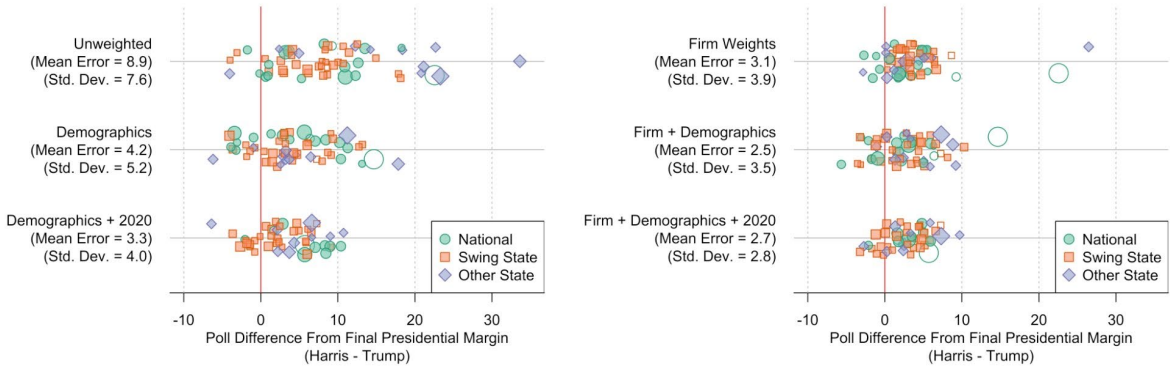

Figure 2.8.3 - Influence of weighting choices on signed errors in and variability of election estimates across contest types (dot size corresponds to sample size). N = 59 for unweighted, firm-weighted, and demographic estimates, 11 samples did not include 2020 preferences for respondents, so N=48 when weighting to 2020 vote share. 3 samples did not provide firm weights, so all cases for these were given a weight of 1 (shown as empty dots and included in calculations for left panel, but not right panel when basing on firm weights).[3]

Still, caution is warranted. Methodological similarity can amplify shared blind spots if most firms rely on similar turnout assumptions, weighting targets, or demographic frames. But the available evidence points to real convergence in design and estimation, not herding behavior, as the primary reason for the tight agreement seen in 2024.

2.9 Shifts in Polling Volume and Geographic Focus

The geographic footprint of pre-election polling has changed substantially over time. In 2024, more than twice as many state-level presidential polls were fielded as national ones—a continuation of a trend that began in the mid-2000s but accelerated notably over the past three cycles. Most of that growth came from intense concentration in seven swing states: Arizona, Georgia, Michigan, Nevada, North Carolina, Pennsylvania, and Wisconsin.

[3] Firm weights were not provided for 3 samples. Excluding these samples from the analyses that did not begin with the firm weights, mean signed errors and standard deviations (in parentheses) were: Unweighted: 8.6 (7.6), Demographic: 5.5 (4.0), Demographic + 2020: 3.4 (4.1). 2020 vote was not available for 11 samples. Excluding these from other analysis, mean signed errors and standard deviations (in parentheses) were: Unweighted: 9.9 (7.5), Demographic: 5.1 (5.2), Firm-Weighted: 4.3 (4.8), Firm + Demographic: 3.5 (4.0).

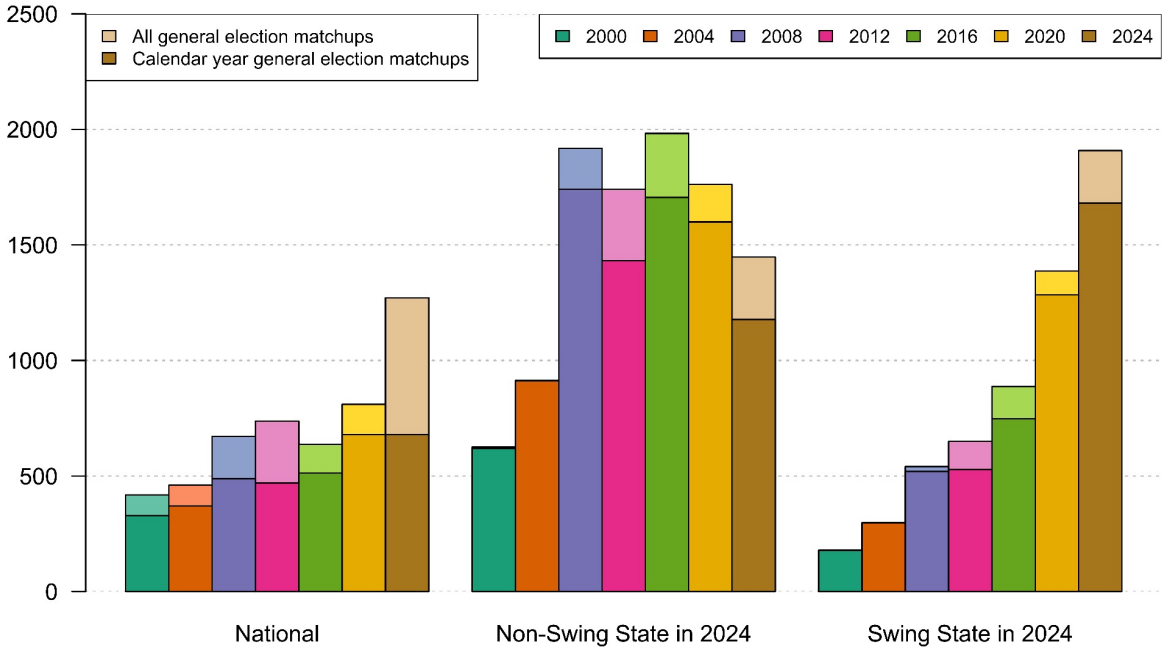

Figure 2.9.1: Total volume of matchups reported from all general elections by election year nationally, for states that were not considered swing states in 2024, and for states that were considered swing states in 2024 in recent presidential cycles (darker portion of each bar indicates matchups reported in the same calendar year as the election)

Between October 23 and November 5, more than 60% of all state-level presidential polls were conducted in those seven states. By contrast, 15 states and the District of Columbia saw no publicly released presidential polling during the same period and an additional 11 states had only one or two publicly released surveys in the last two weeks (meaning that a majority of states had two or fewer polls). This imbalance reflects the structure of the Electoral College and the growing dominance of battleground forecasting in media and campaign strategy. It also reflects a large shift from prior years. In both 2016 and 2020, there had been at least one public poll in every state within the final two-week window.

The shift toward battleground polling has a practical benefit in funneling more polling resources toward the most closely watched and consequential races. But it also narrows the informational window available to the public about broader trends. With fewer national polls and even fewer surveys in noncompetitive states, the picture of public opinion across the country becomes more fragmented.

This pattern is not new, but 2024 marked a high point in its concentration. It also parallels changes in the polling sponsor landscape: partisan pollsters, media organizations, and

corporate pollsters increasingly focus on battlegrounds, while academic and nonprofit pollsters tend to cover broader ground but conduct polls less frequently.

The implications for future cycles are mixed. State-level accuracy has improved alongside this shift, but at the partial expense of broader coverage.

2.10 Making Sense of Past Vote

Few variables hold as much power in election polling—or present as many challenges—as past vote behavior. In 2024, many pollsters incorporated self-reported or voter-file-verified 2020 turnout and vote choice into their survey designs. Some used this information to target samples, others to adjust weights, and still others to build likely-voter models. If the relationship between 2020 behavior and 2024 behavior could be known in advance, modeling using 2020 behaviors could have accounted for the vast majority of the differences between who responded to polls and who actually cast a vote.[4] In practice, though, the effects of modeling using 2020 behaviors were inconsistent across applications, and some firms achieved similar accuracy without

[4] To assess this, the committee re-weighted 50 October and November surveys to account for differences between the raw data they collected and election results. This was done either using all political and demographic variables available to the committee or using just each individual's voting. Weighting using 2020 election results alone accounted for from 50% to more than 80% of the relations between survey data and 2024 results. Adding in additional political and demographic variables brought the variance explained in local results to between 65% and 90%.

incorporating past vote.

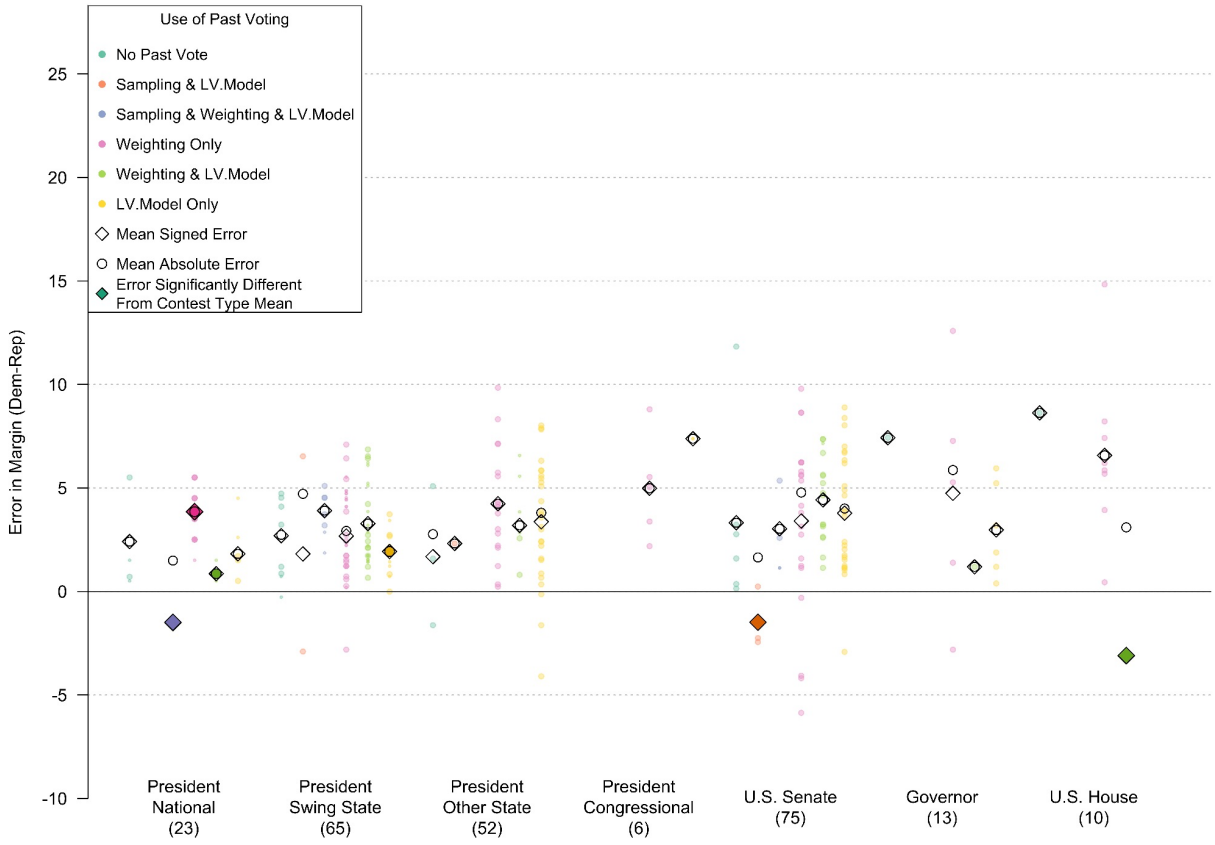

Figure 2.10.1: Correspondence between past vote usage strategies and observed error (estimates of each approach aggregated across hand-coded, keyword searched, and firm survey results depending on availability).

The logic is straightforward: how someone voted—or whether they voted—in the previous cycle is almost always a strong predictor of whether and how they will vote again. But the practice is far from simple. Past vote can be used at multiple stages of the survey process, each with different implications:

- **Sampling**: Voter files can be filtered to include only likely or recent voters; they can also be subset on party registration, past party primary voting (in states that record it), or modeled vote choice data, based on past turnout and registration records. For polls using online panels, longtime panelists can also be categorized based on who they said they voted for in 2020.
- **Weighting**: Respondents' self-reported (or modeled) 2020 vote can be used to post-stratify the final sample to a desired distribution.

- **Turnout modeling**: Data on voters' past turnout (and sometimes partisan data as well) is a core predictor in most multivariate likely-voter models, whether via rule-based screens or regression/machine learning scores.

- **Vote choice modeling**: Expected voting choices can be estimated by asking people how they voted in prior elections and can be modeled in complex ways by looking at demographics, locations, campaign contacts, and other variables that can be linked to a voter file to generate a voter profile. These estimates can be used for sampling as well as to adjust data to account for groups of individuals that may be underrepresented in a particular poll.

Each of these uses introduces potential sources of error or distortion. Self-reported vote is subject to recall error, social desirability bias, and partisan misreporting. Voter files can contain incorrect matches, outdated registrations, or unverifiable entries. Efforts to model the 2024 electorate based on turnout patterns from 2020 may fail to anticipate the behavior of the 2020 nonvoters who participated in the subsequent election. And models used to impute vote choices used by voter file vendors are also sometimes wrong.

These problems are amplified when past vote is used as a benchmark—that is, when weights are forced to match the past electorate rather than the current one. Doing so can effectively hard-code past electoral composition into a poll, making it less responsive to genuine shifts in preferences over time. In 2024, where the composition of the electorate changed meaningfully—especially due to 2020 nonvoters, turnout shifts in rural counties, and demographic change—polls that over-relied on 2020 targets may have captured precision at the cost of adaptability.

Below the state level, calculating the baseline presidential vote in 2020 is often not straightforward. Some jurisdictions, such as congressional districts, have to be painstakingly reconstructed using precinct results and precinct lines, which often themselves changed in the interim, requiring additional imputation and subjective decisions. And in at least one case between 2020 and 2024, a state changed its county level equivalent units, similarly complicating reverse compatibility.

Pollsters varied widely in how they implemented past vote adjustments.[5] Some firms used only turnout history from the voter file; others combined self-report and file data. Many weighted directly on 2020 vote choice; others included it only as a covariate. Still others avoided using it altogether, due to concerns about transparency, data availability, or questions around the appropriateness of partisan adjustment.

But one of the most persistent election polling challenges—and one that modeling using past vote often fails to address—is how to account for irregular and newly eligible voters. In 2024, individuals who did not vote in 2020 but turned out in 2024 made up a substantial share of the electorate, especially in closely contested states. These voters were more Republican-leaning

[5] A trend that was apparent both from the methodological information reported with public polls and from the Task Force survey of selected firms.

than many expected and were underrepresented in samples that relied heavily on prior turnout to construct their frames or weights. Even when surveys included such voters, their turnout likelihood was often underestimated by models trained or selected on habitual election participants.

Relying on past vote aligns samples more with validated history, but it can also miss emerging constituencies. Pollsters need to balance the benefits of using past vote against the possibility that the electorate has changed—a balancing challenge that 2024 made especially clear.

Ultimately, the 2024 results show that judicious use of past vote to inform likely-voter models or supplement weighting can improve accuracy but also carries risk.

3 Conclusions and Future Considerations

The findings in Section 2 offer a broadly encouraging picture: polling in 2024 was more accurate than in the previous two presidential elections, with narrower average signed and absolute errors, reduced variability, and fewer glaring misses. Many long-standing challenges—coverage gaps, turnout modeling, partisan nonresponse—remained, but pollsters appeared to make tangible gains in addressing them.

Still, accuracy alone does not tell the full story. Understanding *why* polls improved and how the lessons of 2024 might carry forward and lead to further improvements requires a closer look. This section summarizes the task force's key takeaways, reflects on what polling can and cannot do, and outlines practical considerations for future pre-election work.

3.1 Summary of Findings

Publicly released election polls in 2024 provided an accurate picture of the race between Kamala Harris and Donald Trump—both nationally and in the battleground states that ultimately decided the outcome. The average absolute error in final two-week surveys across all contests was 3.3 percentage points, a substantial improvement from the 5.3- and 5.2-point average errors observed in 2020 and 2016, respectively. National presidential polls missed by just 2.6 points on average, and state-level presidential polls by 3.0 points. This was the lowest state-level error in a presidential year since 1944.

The tendency to overstate Democratic margins persisted, but was smaller: the average signed error in 2024 was +2.7 points for Democrats, down from +4.6 in 2020. That directional lean appeared in all contest types. 2024 represents the third consecutive presidential election in which the public polls underestimated Republican support, even though midterm election cycles continue to show smaller and more variable signed errors.

No single methodological choice guaranteed more accurate results. Firms used diverse sampling frames, interview modes, weighting variables, and likely-voter models in 2024. While some combinations, such as partisanship weighting and multivariate turnout modeling, were

associated with modestly lower error, these effects were relatively small and often intertwined with firm-level practices.

Some groups remained difficult to measure accurately. Republican voters in GOP-leaning counties were underrepresented relative to Democrats in the same areas; Hispanic voters' Democratic support was overestimated; and 2020 nonvoters—many of whom leaned Republican—were underrepresented or mischaracterized in terms of turnout likelihood.

Despite the tight clustering of results in competitive states, statistical tests found no evidence of herding. The relatively low poll-to-poll dispersion appears to reflect real convergence in field practices and modeling strategies, not artificial alignment across firms.

Most pollsters reached similar conclusions about the outcome of the election—but differed more substantially in their estimates of *who* was voting and the extent of their support for each candidate. Subgroup estimates, particularly among Hispanic voters and 2020 nonvoters varied widely, underscoring the importance of caution when interpreting demographic crosstabs.

Finally, the focus of pre-election polling continues to shift toward battleground states, which is a mixed blessing. In 2024, state-level presidential polls outnumbered national ones by a factor of nearly two-to-one, with seven core swing states receiving the bulk of attention. This targeted focus may help improve election forecasting, but it narrows the public's view into broader national trends and underexamined contests.

3.2 Limitations and Interpretive Caution

Several limitations constrain how the results presented here should be interpreted.

First, this report includes only publicly released polls. These surveys are not a random sample of all the polls conducted during an election cycle. Many internal or proprietary polls—especially those commissioned by campaigns—never appeared in public. Others may have been selectively released when results aligned with a desired narrative. As a result, this report reflects the performance of visible polling, which may differ systematically from the full body of work conducted during the campaign.

Second, not every poll aims to predict the vote margin. Some surveys are designed to track issue salience or test messaging. Others serve internal strategy purposes, where perfect election-day alignment is not the goal. Even among those intended to provide an estimate of the election outcome, methodological variation reflects different priorities: some pollsters emphasize transparency, others speed or cost-efficiency. Benchmarking all polls against the final certified vote helps create comparability—but not all polls should be judged solely on that basis.

Third, late-breaking events can shift results in ways that are not captured in survey data, especially for polls that close several days before Election Day. Although there did not appear to be major events that changed preferences in the days leading up to the election, get-out-the-

vote efforts and last-minute messages could change behaviors in ways that are invisible to pre-election polls.

Fourth, some methodological features are inconsistently reported. Published toplines often omit details about weighting variables, likely-voter models, sample frame, or how weighting and sampling targets are benchmarked. When available, this information was coded, but it was not always complete. Comparisons across design types should therefore be interpreted with care.

Finally, the precision of subgroup and geographic diagnostics depends heavily on the microdata files that were shared with the task force. Those files represent a rich but nonrandom subset of 2024 polls. Findings drawn from them, particularly in Sections 2.5 and 2.6, highlight important patterns but should not be generalized beyond the surveys that contributed data.

3.3 Considerations for Future Polling

As researchers and pollsters prepare for future cycles, several areas are priorities for continued innovation and attention.

First, coverage of hard-to-reach voters remains uneven. Republican-leaning populations in rural and exurban counties continue to be underrepresented in many surveys, especially those using opt-in panels or RDD-based frames. Even when these areas are sampled proportionally, differences in response rates and weighting efficacy can skew results. Expanding frame quality, increasing oversampling in low-response strata, and improving contact strategies for historically underrepresented groups will be essential.

Second, incorporating past voting behavior into survey design—whether through self-report, voter files, or modeling—offers real accuracy gains but must be done thoughtfully. The 2024 data show that past turnout and vote choice data are useful in selecting respondents from panels, building weights, or creating likely-voter models, but they can introduce new biases if applied rigidly or without calibration, especially if the composition of the electorate is changing rapidly. Pollsters will need to continue refining how and when to use this information: as part of sampling, poststratification, screening, or all three.

Third, subgroup estimates require special care. While overall margins were tightly estimated in 2024, demographic crosstabs varied widely across firms. Hispanic vote margins, in particular, differed by as much as 20 points depending on the pollster. Improved subgroup precision may require more consistent reporting practices, use of external benchmarks (e.g., ACS, validated turnout data), and greater adoption of multilevel or model-based poststratification approaches, ideally across data collections, especially when subgroup sizes are small.

Fourth, the shift toward battleground states is likely to persist, given its value for forecasting and media coverage. But this narrowing limits insight into broader electoral dynamics and leaves many races and states underexamined. Striking a balance between national and state-level coverage, as well as between competitive and overlooked races remains key.

Finally, transparency and documentation remain uneven. The field made real progress since 2016, especially with the rise of open aggregators and stronger industry norms. Yet the task force found that full methodological detail is still often missing from public poll releases. Continued adoption of transparency initiatives, such as AAPOR's Disclosure Standards and third-party verification tools, would support stronger accountability and replication.

Similarly, the fact that polling is currently aggregated by independent groups also raises the potential that key archives will not be maintained. Our collective evaluation of polling suffered a significant loss when the *Pollster.com* archive was no longer maintained following the 2016 election. And *ABC*'s closure of the data-reporting site FiveThirtyEight—which tracked polls and provided data used by this Committee—risks a similar potential loss, although other outlets are continuing some of their work. The field should work to ensure that there are ways to keep track of poll releases, rather than leave that task to organizations lacking a preservation strategy.

The 2024 cycle showed that public polling can still offer valuable, reliable insights about the electorate. Yet sustaining those gains will require vigilance, improvement of existing methods, and the development of new methods, all to continue adapting to the shifting political and technological landscape.

Appendix A. Data for the 2024 Report

The analyses in this report rest on a deliberately broad evidence base: every publicly released 2024 pre-election poll located by the task force, multiple official vote files, large-scale demographic surveys, commercial voter-file records, and poll-method metadata scraped from online aggregators. This section documents how those pieces were found, cleaned, coded, and linked—and lays out the error metrics and statistical tests used in later chapters. Readers who want only the headline findings can skim quickly; those who wish to replicate or extend the work will find the precise inclusion rules, variable definitions, and benchmark sources spelled out in the rest of this appendix and the following ones.

A.1 Data sources for 2024 and historical polling margins

The task force archive begins with 2,631 publicly released 2024 general-election polls collected between January 1, and November 5, 2024. These include vote-intention questions for the presidential race, 32 of the 35 U.S. Senate contests, all 11 gubernatorial contests, and district polling for 94 U.S. House contests. From that full universe, the task force isolated the 611 matchups reported across 403 surveys whose fieldwork ended between October 23, and November 5, 2024; this “final-two-weeks” subset serves as the basis for headline accuracy statistics in Section 2.

Because the task force was also charged with assessing change over time, the same collection and cleaning protocols were replicated for earlier cycles. Parallel archives cover 2016, 2018, 2020, and 2022, yielding 2,505, 1,219, 2,827, and 1,031 survey releases, respectively. Where long-run error trends are shown (e.g., Figure ES-1), the task force additionally draws on historical election result files, which extend presidential-poll accuracy estimates back to 1940.

All polls—2024 and historical—are linked to a common set of benchmarks and auxiliary files:

- Certified national and state vote returns (Associated Press, Federal Election Commission and historical data from CQ)
- State-level canvas and turnout files for geographic-coverage diagnostics
- American Community Survey (ACS) microdata for demographic baselines
- National voter-file extract to identify voters who had not participated in 2020
- FiveThirtyEight poll-method metadata
- Hand coding of mode, sampling frame, weighting variables, and likely-voter models
- Confidential respondent-level microdata supplied by individual pollsters (aggregated results only are reported)
- Responses to a questionnaire about methods used that was sent to the firms that had produced the most polling
- Figures from post-election analyses of the electorate, including the Exit Polls, AP VoteCast, and reports by Catalist and Pew.

In this report, a survey is a single data collection by a pollster and may include questions on multiple contests. A matchup (or contest-poll) is a candidate×geography contest estimate

produced by a survey and is the unit used in most counts, tables, and models. The term ‘poll’ is used somewhat informally in prose; unless noted otherwise, ‘poll’ here refers to a contest-poll (matchup).

The harmonised poll-level dataset, variable codebook, and R scripts that reproduce every figure and table will be posted to the Task-Force GitHub repository upon publication.

A.2 Dataset limitations

Several cautions accompany the archive. Publicly released polls are a convenience sample of all the interviewing that takes place; some campaign surveys never see daylight, while others appear only if the sponsor thinks they look good. Late-breaking events sometimes occurred after many “final” polls were finished collecting data. A subset of releases offer only registered-voter results, and a few omit methodological detail altogether. Finally, special elections and run-offs fall outside our frame. None of these gaps upends the topline findings, but readers should keep them in mind when drawing fine-grained conclusions about subgroups or timing.

It is also important to note that conclusions about election surveys may be limited in what they tell us about polling more generally. One reason for this is that election polling needs to address questions that differ from the goals of most public polls. In contrast to other polls, which largely report on the attitudes or past behaviors of the entire public, election polling is designed to figure out what a subset of the population will do in the future. Because the set of people who will vote and their choices have not yet been determined, this adds additional uncertainty to election polls. It is inherently more challenging to accurately measure behaviors that haven't happened yet than ones that have.

On the other hand, election polling can benefit from the fact that we have exceptionally good data about some factors that are closely related to turnout and candidate preferences on election day—e.g., past voting behaviors and partisan identification—that can be used to carefully calibrate these survey estimates. Attitudes and behaviors measured in other polling cannot be as finely tuned.

A.2 Replication and data access

A de-identified poll-level file, the full R workflow used to clean and analyze the data, and scripts that generate every figure and table will be posted on the task-force GitHub page upon publication. Respondent-level files supplied by individual pollsters will remain accessible only to the committee. Aggregate statistics derived from those files appear are fully reproducible from the public code.

Appendix B. Poll Inclusion and Cleaning Rules

This appendix outlines the precise steps used to identify, screen, deduplicate, and weight publicly released pre-election polls for inclusion in the 2024 polling archive. These procedures ensure that only relevant, comparable surveys were included in accuracy assessments and that overrepresented polls did not skew averages.

Poll collection began with automated grabs from the FiveThirtyEight polling API, followed by matching scrapes of RealClearPolitics and, finally, a sweep of the Roper Center iPoll archive and sponsor-hosted PDFs for surveys that never reached the big aggregators. Once all rows were stacked, a precedence rule kept only one record per firm-contest-year combination, favoring FiveThirtyEight over RCP and RCP over Roper links.[6] That single step removed most cross-platform duplicates while preserving genuinely distinct questionnaires.

A survey remained in the master file only if its topline results were publicly available before Election Day, it covered the presidential, Senate, House, or gubernatorial general election, and it reported a two-party vote share (or margin) for the Democratic and Republican nominees who were actually on the ballot at the time of interviewing.[7] Each poll was time-stamped by its final field date; for rolling studies, the task force retained just one observation per reporting window. When a release reported several ballot versions—say, with and without minor-party candidates—ballots that included both major party candidates or that best matched the certified slate were kept.

Even after cross-platform deduplication, the same questionnaire could still surface more than once within a matchup. This could occur when aggregators included multiple versions of the same poll with different questions or weighting strategies or when tracking polls only partially overlapped. Inside each surviving survey release, those repostings were collapsed to identify the number of unique reports and surveys, and the result's analytic weight was set to 1 / nReports, with *nReports* equal to the number of public repostings that cleared all filters. This down-weighting prevents multiply-reported matchups and tracking polls from dominating averages and harmonises the differing reposting conventions of FiveThirtyEight (one record per survey estimate) and RCP (one record per matchup). Where tracking waves cover only part of a unique field period, contributions were prorated to their share of unique interview days (and still divided by nReports).

[6] FiveThirtyEight was preferred because it had the best coverage of recent elections, included the most results per matchup, and included the most complete metadata. Real Clear Politics generally included a single report per matchup, but had better coverage than Roper.

[7] There were also a few matchups where one major party candidate was competing against an independent candidate. These occurred in 2024 for Nebraska's Senate seat between Fischer (R) and Osborn (NP) and the Vermont Senate seat contested by Sanders (I) and Malloy (R). For these contests, the independent candidate was treated as having the opposite party from the declared major party candidate. That is, Osborn and Sanders were both regarded as Democrats when comparing partisan vote shares.

After screening and cleaning, the 2024 file contains 2,631 unique surveys covering 194 contests and fielded by 219 organizations across the calendar year; 403 of those surveys, reflecting 611 matchups, fall inside the final-two-weeks window and form the basis for the accuracy statistics in the next section. A complete audit trail—including raw aggregator dumps, R scripts, and the cleaned poll-level dataset—will be posted to the task force GitHub repository upon publication.

Terminology: A **survey** is a single data collection by a pollster and may ask about multiple contests. A **contest** is a candidate×geography race (e.g., U.S. Senate in State S). A **matchup** (a contest-poll) is a survey×contest estimate and is the unit used in counts, tables, and accuracy. For any given survey×contest, there is not more than one matchup; if multiple ballot versions exist, we adjust for these per B.1. A **result** is a reported estimate for a specific matchup (e.g., alternative weights, likely-voter screens, or split releases); multiple results can exist within a matchup and, after cross-platform de-duplication, are treated as repostings in B.4 (weighted 1/nReports; partial overlaps handled per B.2–B.4). We use **poll** informally in prose; unless noted otherwise, "poll" refers to a contest-poll (matchup).

B.1 Initial Inclusion Criteria

A matchup was retained in the full dataset if it met all of the following conditions:

- The survey was publicly released prior to Election Day (November 5, 2024).
- It covered a general-election contest: presidential (national or state), Senate, governor, or at-large U.S. House.
- It reported a two-party vote share for the named Democratic and Republican nominees on the ballot at the time of fieldwork.
- It included a clearly stated field period end date.
- If multiple ballot versions were reported (e.g., with and without minor-party candidates), the ballots most consistent with the certified candidate list were used.[8]

Surveys that reported only registered-voter results, or that included respondents who could not be clearly assigned to a party matchup, were excluded from accuracy statistics but retained in broader descriptive analyses.

B.2 Field Date Rules and Final Window

Surveys were assigned to a calendar date based on their last day of fieldwork. For rolling surveys or tracking studies, final dates were used to anchor the polls. If multiple releases were issued for overlapping field periods, a subset of these starting from the last release and

[8] Multiple results could be considered consistent with the list if all eventual major party nominees were in the poll and only minor party nominees varied. If this happened, included results were weighted in all analyses so that they counted as a single matchup.

selecting each additional poll using non-overlapping dates was retained for descriptive purposes and accuracy statistics.[9]

The “final two weeks” window was selected to match past AAPOR task force standards and ensure comparability across cycles. Historical datasets, including parallel datasets for 2016, 2018, 2020, and 2022, applied the same windowing rules.

B.3 Duplicate Handling and Source Precedence

To eliminate duplication across aggregators and sponsor repostings, the dataset applied a hierarchical precedence rule for source links:

1. FiveThirtyEight polls took precedence when a firm-contest-year match appeared across sources.
2. RealClearPolitics filled in polls not already captured by 538.
3. Roper iPoll archive was used for any remaining gaps, particularly for older polls.

If the same firm-contest combination appeared more than once with slightly different field dates or sample sizes, and referred to the same underlying survey, only the most complete and recent version was kept; otherwise, entries were treated as distinct surveys. Where date ranges overlapped and it was unclear whether versions were distinct, the record from the higher-precedence source was retained. Source precedence was designed to maximize the number of results available for each contest and the availability of poll metadata.

This rule assumes that aggregator coverage of a firm-contest pairing is consistent within a year and that repostings typically reflect the same underlying data.

B.4 Handling of Multiple Releases

In some cases, the same matchup (contest-poll) was published multiple times (e.g., because alternate sets of weights, likely voter models, or questions were used). These versions often appear as separate records in aggregator feeds, especially on FiveThirtyEight (which treats each set of estimates reported from a poll as a new row).

To avoid overweighting these polls, a weighting adjustment was applied:

- For each poll that appeared more than once and passed all other inclusion filters, an analytic weight of 1/nReports was assigned, where *nReports* is the number of qualifying reposts.
- This prevents high-volume syndicated polls from dominating means and ensures consistent treatment across aggregators.

[9] If release timing meant that two polls had a partial overlap even after following this procedure, partially overlapped polls were incorporated into the dataset, but were downweighted to account for the proportion of their data that was novel (i.e., a window with two unique days for a four-day field period was counted as half of a poll).

- When releases partially overlap (e.g., tracking waves), we partition field dates into non-overlapping segments and prorate each release by its share of unique interview days within the final window; the prorated contribution is then weighted by 1/nReports.

A total of 3,671 surveys asked about general election contests for the 2024 cycle, covering 4,666 matchups and 8,483 reported results. 2,631 of these surveys were conducted in the 2024 calendar year, with 3,575 matchups polled and 5,855 total results reported.

B.5 Contest Coverage

The final 2024 dataset includes:

Contest Type	Full cycle matchups	2024 year matchups	Last 2 week matchups	All contests polled (in last 2 weeks)
Presidential national general	1271	680	60	1 (1)
Presidential state general	1994	1609	291	50 (35)
Presidential Congressional district general	31	29	8	5 (5)
Senate general	922	845	197	32 (27)
Gubernatorial general	162	146	27	11 (7)
U.S. House general	283	262	28	100 (20)
Total general	4666	3575	611	200 (95)

A full contest-by-contest breakdown appears in Appendix E.

Appendix C. Variable Definitions and Coding Crosswalk

This appendix defines all core poll-level variables used in the report and documents how those variables were derived or reconciled across multiple metadata sources. For any given analysis, the source providing the most complete and precise coverage was used, but full harmonized codes are retained in the archive.

Polls differ in many ways, but four decisions are especially important for how accurate they turn out to be:

1. **Interview mode** – *how the interview was conducted.* A live interviewer on the phone, an automated dialer that accepts touch-tone responses, a text message that directs people to a web link, a fully online questionnaire, or a mixed approach.
2. **Sampling frame** – *where the contact list came from.* Random phone numbers (RDD), an official voter file, a postal address list (ABS), or an opt-in panel of volunteers.
3. **Weighting variables** – the levers a pollster pulls to make sure their sample looks like the electorate (age, gender, race, education, partisanship, past vote, region).
4. **Likely-voter (LV) model** – the rule that tries to separate people who will actually cast a ballot from those who enter the sample. Some use a single likelihood question; others build a multivariate turnout score.

To the extent possible, each poll in the archive—both 2024 and earlier years—was assigned values on those four dimensions. To do that, five information sources were combined, listed in the order of typical preference, but not used as a strict hierarchy (the best-documented source that covered most polls in a given comparison always won):

Label used in this report	What it is	What it gives us
Aggregator-coded (FiveThirtyEight tags)	Standardized “methodology” strings attached to most polls since the mid-2010s (and to many older Pollster records); available for most records.	Record-level mode and LV flags

API-coded (ChatGPT summaries)	For any poll with an accessible report, press release, topline, or crosstab file, that file was downloaded and ChatGPT was asked to produce a short plain-English methods summary; these were then matched to keywords such as “live interviewer,” “RDD,” or “address-based” with a pattern-matching dictionary; available for almost all polls since 2000.	Record-level mode, frame, weighting, and LV details
Hand-coded	A task force member read every poll in the final two-weeks subset of 2024 and assigned codes directly.	Record-level mode, frame, weighting, and LV details
Firm-survey codes	Thirty-nine organizations answered a questionnaire indicating whether they *never, sometimes,* or *always* used particular practices in their 2024 pre-election surveys.	Firm-level mode, frame, weighting, and LV details
Pew Research Center historical file	Mode and frame classifications maintained by Pew. 2000 through 2022	Firm-level mode and frame details

Because these sources rarely all exist for the same poll, the committee chose the highest-quality source that covered most records in any given analysis. For example, a comparison of 2024 mode accuracy relies on hand-coded values (complete coverage, most precise), while a long-run trend may rely on aggregator-coded and Pew values (broader coverage, still comparable). For each analysis, the source used is noted.

C.1 Key Variables Used in Analysis

Variable	Description	Source(s)	Notes
mode	Interview mode: how the respondent was surveyed	538 tags, API summaries, hand-coded, firm survey, Pew	Example categories: live phone, text-to-web, IVR, online opt-in, online probability, multiple
frame	Sampling frame: source of contact info	API summaries, hand-coded, firm survey, Pew	Example categories: RDD, voter file, ABS, opt-in panel
weight_vars	Whether poll was weighted on various demographic and political variables	Hand-coded, firm survey, API summaries	Includes: education, race, party, past vote, region; coded as binary flags
lv_model	Type of likely-voter model used (if any)	538 tags, API summaries, hand-coded	Example categories: none, self-reported, index score, registration file
partisan_affiliation	Whether pollster was party-affiliated	538, hand-coded	Categorical (3 levels): Dem-affiliated, GOP-affiliated, nonpartisan
sample_size	Number of respondents in final topline	Aggregator metadata	Some rounding or missing values

margin_error	Reported margin of error (MoE)	Aggregator metadata or API summary	Used for error normalization in Appendix H
contest_type	Type of election polled	Aggregator metadata	Categories: president (national), president (state), Senate, governor, House
field_end_date	Final day of interviewing	Aggregator metadata	Used to assign poll to field window
weight	Downweighting for reposts	Computed (1/nReports)	Used in accuracy statistics

C.2 Coding Sources and Prioritization

When multiple sources provided metadata, the following precedence rules applied:

1. Hand-coded values were used when available, especially for 2024 final-week polls.
2. If hand coding was unavailable, FiveThirtyEight tags were used.
3. When both were missing or incomplete, API summaries (from ChatGPT method descriptions) were used, matched to regex tags.
4. Firm-survey responses were used for firm-level default practices (e.g., “always weights on party”) when poll-level metadata was not available.
5. Pew classifications were used primarily for historical (2000–2022) cycles and only for mode and frame.

Each table or figure in the report indicates the source used for poll characteristic variables in that analysis.

Appendix D. Benchmark Election Results

Every accuracy number in this report is grounded in certified vote returns. Because no single data set spans every contest and year, this report combined three well-known sources, always taking the best available option in this order:

1. **Associated Press election files (AP).**
   Primary benchmark. The AP provides county-level results for presidential, Senate, House, and most gubernatorial races in 2024 and in many earlier cycles. These files are normally updated within a few weeks of state certification. County totals come from the same AP feed.

2. **Federal Election Commission canvass (FEC).**
   First fallback. The FEC publishes certified national and state returns for presidential, House, and Senate contests. It does not report gubernatorial results, and its release lags the AP by several months, but it includes a longer historical record of contests.

3. **CQ Press Voting and Elections Collection (CQ).**
   Second fallback. For contests or years where neither AP nor FEC provides coverage—mainly older cycles or gubernatorial contests—CQ's state-level returns supply the two-party vote share. CQ is used only when needed for statewide comparisons.

For each contest, the task force computed the two-party vote margin—Democratic share minus Republican share—so minor-party and write-in votes do not cloud the baseline. Poll toplines are converted the same way, producing an apples-to-apples error measure.

Appendix E. Microdata Contributions

Several polling organizations shared respondent-level datasets with the task force under data use agreements. These files enabled the geographic coverage, subgroup accuracy, and demographic diagnostics reported in Sections 2.4 through 2.6. All analyses were conducted using de-identified data, and only aggregate results are reported.

To see *where* polling errors emerged—and which voters were hardest to reach—the task force asked every prolific 2024 polling organization if they would be willing to share a de-identified copy of at least one data set. Each was also asked to complete a questionnaire about their field practices.

- **Who responded?**
  - 86 organizations were contacted; 24 pollsters ultimately shared files, yielding 86 micro-datasets that together cover 143 individual data collections.
  - 39 pollsters (including all that shared microdata files) completed the post-election methods survey.
- **What do the files contain?**
  - Respondent weights, basic demographics, 2024 vote choice, and usually self-reported 2020 turnout and vote.
  - Geography to at least the state level; three-quarters of national polls include county, ZIP, or finer identifiers.
  - Roughly two-thirds of the firms also used voter-file variables such as verified past turnout, though these were only sometimes provided to the committee.
- **Validation and use**
  - County and ZIP codes were screened for impossible values and matched to FIPS codes.
  - Only the polls with microdata feed the county-coverage and subgroup analyses; statewide accuracy statistics draw on the full archive.
  - Aggregate results only are reported—no respondent-level records leave the secure share.

E.1 Contributing Polling Organizations

Between November 2024 and February 2025, 86 polling firms were invited to participate. Of those, 24 firms shared at least one microdata file, and 39 completed the accompanying methodological questionnaire (see Appendix J). Collectively, the shared files cover 143 poll releases throughout the cycle.

E.2 Variables Included in Shared Files

Most shared datasets included the following respondent-level fields:

- Demographics: age, gender, education, race/Hispanic origin

- 2024 general-election vote choice (Harris, Trump, other, undecided)
- Survey weights (raked or poststratified, as used in the topline)
- Self-reported 2020 turnout and vote choice
- Geographic identifiers, including:
    - State (100% of files)
    - County FIPS, ZIP, or congressional district (approx. 75%)
- Voter-file appends (in ~60% of files):
    - Mode of contact
    - Turnout history (validated)
    - Modeled partisanship

E.3 Validation, Cleaning, and Use

Each microdataset was screened for:

- Usable weights and complete vote-choice fields
- Valid state, county-level, or zip code identifiers (mapped to FIPS codes)

A harmonized analysis file was created to support:

- **County coverage diagnostics** (Section 2.5)
- **Subgroup vote margin dispersion** (Section 2.6–2.7)
- **Demographic comparisons to external benchmarks** (Section 2.6)

Only polls for which usable microdata were available were included in those analyses. All state-level margin analyses and most methodological comparisons in Section 2 continue to use the full poll-level archive.

E.4 Data Access and Retention

These datasets are stored by the AAPOR task force committee. No respondent-level data will be publicly released.

Aggregate results from these files are fully reproducible using the posted code in the task force GitHub repository. See Appendix G for replication details.

Appendix F. Sub-group and geographic diagnostics

The statewide margin tells us how far off a poll was; it does not reveal where or among whom it missed. To expose those patterns this report relies on the respondent-level datasets described in Appendix E. Only polls for which a micro-dataset was shared and data collection occurred in October or November feed microdata accuracy analyses based on these data; broader accuracy statistics continue to draw on the two weeks archive.[10]

The first diagnostic asks whether polls interviewed voters in the right places. Many microdata files carry either a county FIPS code or a ZIP that can be mapped to one. Weighted respondent counts were aggregated compared with the Associated Press county vote share data introduced in Appendix D. A county in which a pollster collects too few interviews relative to its turnout receives a negative "coverage residual," while over-represented counties receive a positive one; these were then aggregated across counties based on how counties voted in 2020. Section 2.5 shows that under-sampling Republican counties and over-sampling Democratic counties explains part—but not all—of the modest Democratic overstatement that persisted in 2024.

The second diagnostic turns to the demographic mix. For every micro-dataset that includes basic socio-demographics, the poll's weighted share of age, gender, race or Hispanic origin, and educational attainment was computed and compared with estimates of the proportion of those groups in the electorate from four additional data sources: Catalist's validated-turnout file, the National Exit Poll, AP VoteCast, and the Pew Research Center's post-election validated-voter study.

A few caveats temper these diagnostics. Roughly half of the micro-datasets lack usable county identifiers, so those polls drop out of the geographic check. Even with these limitations, the microdata reveal patterns that a single statewide margin cannot. Together with the benchmark sources above, they clarify *where* and *among whom* polling currently struggles, setting up the explanations in Section 2.5.

[10] The last two weeks approach was used to match prior committee reports using topline data. Because there was relatively little topline survey change in October and some of the firms that provided data fell just outside the last two weeks, we opted to use this longer period.

Appendix G. Accuracy Metrics and Statistical Tests

Poll performance was assessed with six complementary measures that together capture *magnitude*, *direction*, *comparability*, and *clustering* of errors:

First, the average absolute error is the mean distance—without regard to direction—between each poll's two-party margin and the certified election margin, expressed in percentage points. This headline metric shows how far, on average, polls missed.

Second, the mean signed error retains direction: positive values signify polls leaned Democratic, negative values Republican. To unpack whether one party's support was more routinely misestimated, the task force also calculated party-specific absolute errors, averaging the absolute miss for the Democratic share and, separately, for the Republican share across all polls.

Third, recognizing that polls vary in sample size and design precision, margin-of-error units are also generated by dividing each poll's absolute error by its own published MoE. Averaging these ratios places all polls on a common scale, helping compare a small-n telephone survey with a large web panel.

Fourth, poll-to-poll dispersion measures how tightly different estimates cluster: within each contest, the contest's mean margin was subtracted from each poll's margin and the standard deviation of those residuals was calculated. A low dispersion means polls agreed closely; a high dispersion signals divergence.

Fifth, the task force explored herding, asking whether some firms' polls moved unnaturally toward the pack. The herding statistic compares each poll's error to the contemporaneous prior and subsequent poll averages; firms with polls significantly closer to the prior average than subsequent polling in the same contests constitutes convergence beyond chance.

Finally, to ensure robustness the committee tested whether these metrics vary systematically by contest type, field-date window, or methodological category.

Appendix H. Alternative Accuracy Windows and Metrics

The main report evaluates polling accuracy using the final two-week window of the 2024 campaign: polls with fieldwork ending between October 23 and November 5, and reporting two-party vote shares for both major-party nominees. That choice follows precedent from earlier AAPOR task force reports and strikes a balance between freshness and sample size. This appendix explores how the results change when different inclusion rules or error metrics are applied.

H.1 Alternative Field Windows

Table H.1 compares overall average errors using three different field-date cutoffs:

Field Window	Poll Count	Avg. Abs. Error	Signed Error (Dem – Rep)
Oct 23 – Nov 5 (main)	611	3.3 pp	+2.7 pp
Oct 30 – Nov 5 (7-day)	334	3.2 pp	+2.5 pp
Nov 2 – Nov 5 (3-day)	179	3.3 pp	+2.5 pp

The magnitude and direction of polling error were consistent across windows. Slight reductions in signed error are visible in narrower windows, as late-deciding voters and shifts in turnout patterns became clearer.

H.2 Alternative Error Metrics

In addition to the mean absolute error (MAE) and signed error reported throughout the main text, a number of additional metrics are examined here:

- **Root Mean Square Error (RMSE)**: Emphasizes large outliers
- **Share Outside Margin of Error (MoE)**: Percent of polls where the certified result falls outside the poll's reported MoE
- **Median Absolute Error**: The absolute error size of the average poll, which is not sensitive to rare outliers.
- **Median Signed Error**: The signed error size of the average poll, which is not sensitive to rare outliers

- **Correct Winner**: The proportion of polls for which the poll leader matched the election winner, with ties counted as half correct
- **Absolute Democratic Error**: The mean signed error for Democratic candidates
- **Absolute Republican Error**: The mean signed error for Republican candidates
- **Signed Democratic Error**: The mean signed error for Democratic candidates (pos. means share is overestimated; neg. means share is underestimated)
- **Signed Republican Error**: The mean signed error for Republican candidates (pos. means share is overestimated; neg. means share is underestimated)

Metric	All Contests	National Presidential	State Presidential	Senate	Governor	US House
Mean Absolute Error (pp)	3.3	2.6	3.0	3.3	5.0	7.0
RMSE (pp)	4.7	2.8	4.2	4.0	9.2	11.4
Share Outside MoE	34%	48%	29%	33%	60%	100%
Median Absolute Error (pp)	2.6	2.1	2.3	2.7	3.3	6.7
Mean Signed Error (pp)	+2.7	+2.5	+2.6	+2.4	+3.8	+4.9
Median Signed Error (pp)	+2.1	+2.1	+2.1	+1.6	+1.7	+5.7
Correct Winner	71%	30%	72%	76%	96%	80%

Absolute Democratic Error (pp)	1.6	1.2	1.4	1.8	2.7	3.1
Absolute Republican Error (pp)	2.9	2.3	2.5	2.9	4.0	6.2
Signed Democratic Error (pp)	-0.04	+0.2	+0.1	+0.3	+0.3	-1.1
Signed Republican Error (pp)	-2.7	-2.3	-2.4	-2.7	-3.5	-6.0

RMSE values are slightly higher than mean absolute error, reflecting a small number of large-miss polls, particularly in low-polling states. The share of polls with actual results outside their reported MoE varies by mode and sample size (not shown), but is somewhat smaller than the historical norm. Patterns for median error indicate that a few large errors made a big difference in the averages. And the smaller errors for Democratic candidates may indicate that error comes more from an underprovision of Republican respondents rather than an overprovision of Democratic ones.

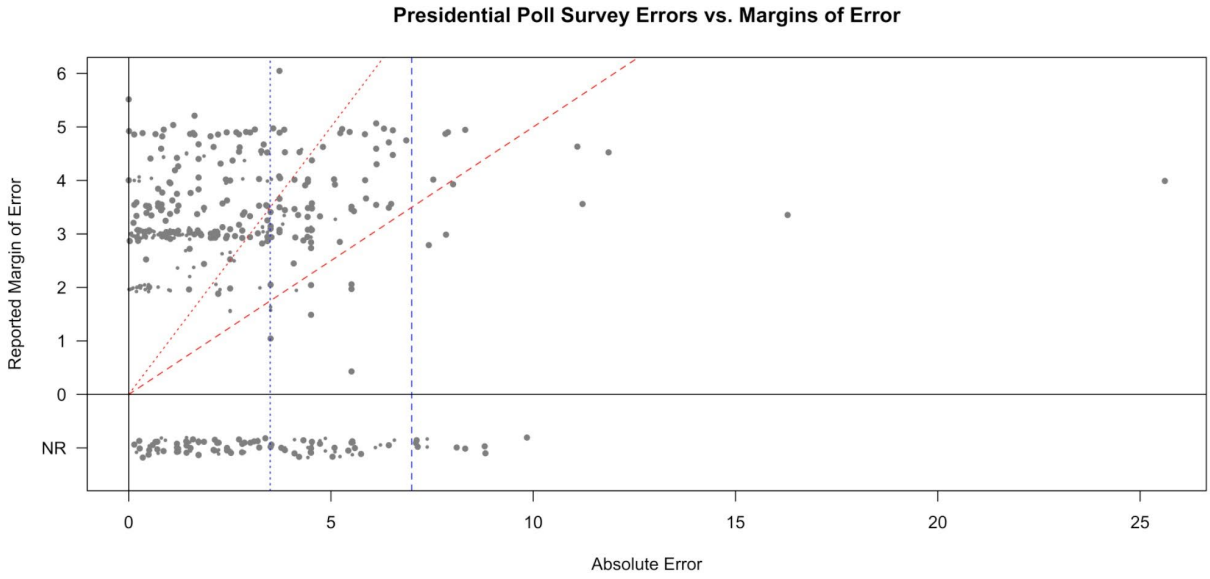

Figure H2.1: Distributions of reported margins of error in 2024 Presidential Polls. NR indicates that a margin of error for that poll was not reported. Dashed angled red lines indicate which results were more or less than 1 and 2 times their reported margins of error. Dashed vertical blue lines indicate which results were more or less than 1 and 2 times the overall average reported margin of error.

Appendix I. Full Accuracy Tables by Contest and State

This appendix provides complete accuracy statistics by contest type and state for all final-window polls (October 23–November 5) in 2024, along with selected parallel metrics from earlier cycles. These tables underpin summary statistics in Sections 2.1, 2.2, and 2.3 and support historical comparisons presented in Figure ES-1.

I.1 2024 Accuracy by Contest Type

Contest Type	Poll Count	Avg. Abs. Error	Signed Error (D–R)	RMSE	Median Error
Presidential (National)	60	2.6 pp	+2.5 pp	3.2 pp	+2.5 pp
Presidential (State)	291	3.0 pp	+2.6 pp	4.0 pp	+2.2 pp
Presidential (Swing States)	190	2.3 pp	+2.0 pp	2.9 pp	+1.7 pp
Presidential (Non-Swing States)	101	4.2 pp	+3.6 pp	5.6 pp	+3.3 pp
U.S. Senate	197	3.3 pp	+2.4 pp	4.2 pp	+2.2 pp
Gubernatorial	27	5.0 pp	+3.8 pp	7.2 pp	+2.3 pp
House	28	7.0 pp	+4.9 pp	9.0 pp	+3.9 pp

I.2 2024 Accuracy by State (Presidential Contests)

State	Poll Count	Avg. Abs. Error	Signed Error (D–R)	Certified Margin	Poll Avg. Margin
Arizona	28	3.5 pp	+3.3 pp	-5.5 pp	-2.3 pp
Florida	10	6.4 pp	+6.4 pp	-13.1 pp	-6.8 pp
Georgia	23	1.2 pp	+0.6 pp	-2.2 pp	-1.6 pp
Michigan	34	2.5 pp	+2.4 pp	-1.4 pp	1.0 pp
Nevada	19	3.1 pp	+2.5 pp	-3.1 pp	-0.6 pp
North Carolina	22	1.9 pp	+1.8 pp	-3.2 pp	-1.4 pp
Ohio	11	3.6 pp	+3.6 pp	-11.3 pp	-7.8 pp
Pennsylvania	35	1.9 pp	+1.6 pp	-1.7 pp	0.0 pp
Wisconsin	29	1.9 pp	1.6 pp	-0.9 pp	0.7 pp

I.3 Historical Accuracy by Year and Contest Type (Final Window)

Year	Contest Type	Poll Count	Avg. Abs. Error	Signed Error	Notes
2016	Presidential (State)	501	5.7 pp	+3.3 pp	Baseline for recent cycles
2020	Presidential (State)	375	4.8 pp	+4.0 pp	
2022	Senate	150	4.8 pp	–1.2 pp	Midterm benchmark
2024	Presidential (State)	291	3.0 pp	+2.6 pp	Current year
2024	Senate	197	3.3 pp	+2.4 pp	Current year

I.4 Notes on Table Construction

- Signed errors are calculated as the reported poll margin minus the certified election margin, such that positive values indicate a Democratic lean.
- Contest-level values are weighted using the reposting adjustment described in Appendix B.4.
- When multiple polls existed for the same contest, they are treated independently unless issued by the same sponsor and fielded over identical dates.

A full .csv version of these tables will be posted to the task force GitHub repository upon publication.

Appendix J. Methodology Survey Instrument and Responses

To supplement metadata gathered from aggregators and public releases, the task force distributed a brief post-election questionnaire to polling organizations that conducted public pre-election surveys in 2024. The goal was to gather additional information about firms' typical design practices—especially when poll-level detail was incomplete or ambiguous.

J.1 Survey Outreach and Participation

- 86 organizations were invited to participate.
- 39 completed the full questionnaire.
- All firms that submitted microdata (see Appendix C) also completed the survey.

Firms were asked to answer based on their typical approach in 2024, with options to indicate whether specific practices were "Never," "Sometimes," or "Always" used.

J.2 Survey Instrument

The pollster questionnaire covered six domains used in the analyses:

1. **Sample types and sampling methods** (whether 2024 pre-election polls used probability, non-probability, or blended samples; where respondents were sourced—e.g., online panels, voter-file lists, RDD, ABS, direct outreach).
2. **Modes of data collection** (telephone live interviewer, IVR, text-to-web, text surveys, online panels, mail, face-to-face, ads, AI responses).
3. **Weighting and quotas/stratification** (whether weighting was used; demographic vs. political controls; specific variables; use of quotas/strata and which variables were quotaed).
4. **Likely-voter (LV) modeling** (whether LV models were produced; classification vs. probability scoring; features included).
5. **Voter-file usage** (whether files were used for sampling, matching, weighting, LV estimation; primary vendor).
6. **Targeted recruitment** (any steps taken to increase representation of likely Trump voters; brief descriptions).

The full, formatted questionnaire appears in the Online Appendix.

J.3 Aggregate Results (Selected Items)

Unless noted otherwise, percentages are of completed responses (N=39; though one of these only completed part of the survey). Items with conditional denominators are labeled.

Weighting practices (all polls unless noted):

Practice	"Always"	"Sometimes"	"Never"
Any adjustment/weighting applied	81.6%	10.5%	7.9%
Demographic weighting applied (of some adjustment/weighting)	94.7%	5.3%	—
Weight on age (of some adjustment/weighting)	92.1%	5.3%	2.6%
Weight on education (of some adjustment/weighting)	92.1%	5.3%	2.6%
Weight on party ID (of some adjustment/weighting)	26.3%	34.2%	39.5%
Use of past vote in weights (e.g., 2020) (of some adjustment/weighting)	31.6%	42.1%	26.3%
Political variables in weights (party/past vote/turnout) (of some adjustment/weighting)	68.4%	28.9%	2.6%

Likely-voter (LV) modeling:

Practice	"Always"	"Sometimes"	"Never"
Use of a likely-voter model	76.3%	0.0%	23.7%

LV model specification (among LV modelers; n=29):

- Classified as voter/non-voter: 48.3% (14 of 29)
- Assigned a probability of voting: 37.9% (11)
- Both classification and probability: 10.3% (3)
- Something else: 3.4% (1)

Features used in LV models (among LV modelers; n=29; select-all—"some" or "all"):

- Self-reported likelihood: 89.7% (26)
- Voter-file past voting (match): 62.1% (18)
- Demographics of prior voters: 48.3% (14)
- Self-reported enthusiasm / other self-reports: 41.4% (12 each)
- Self-reported past vote: 37.9% (11)
- Candidate preference: 2020 13.8% (4); 2022 3.4% (1); 2024 6.9% (2)

Sampling, modes, and voter-file usage (select-all—"some" or "all"):

- Sample types used: Probability 57.9% (22); Non-probability 34.2% (13); Blend 15.8% (6); "It varies" 10.5% (4).
- Where respondents were found: Voter-file list-based phone 60.5% (23); Online non-probability panels 47.4%(18); Online probability panels 28.9% (11); ABS from voter file 13.2% (5); RDD 10.5% (4); ABS from USPS 5.3% (2); direct email 13.2% (5); online ads 10.5% (4); river 2.6% (1); other 7.9% (3).
- Modes used: Live telephone 63.2% (24); Text-to-web 55.3% (21); Online non-prob panel 42.1% (16); Online probability panel 34.2% (13); Text survey 21.1% (8); IVR 10.5% (4); Mail 5.3% (2); Face-to-face 2.6% (1); Online ads 7.9% (3); AI responses 0.0% (1 "Not sure").
- Phone coverage (among phone users; n=25): Both cell & landline 96.0% (24); cell-only 4.0% (1).
- Voter-file usage: Sampling 65.8% (25); Weighting 57.9% (22); Matching respondents 55.3% (21); Estimating LV 52.6% (20); None 18.4% (7).
- Vendors (single choice + write-in): L-2 28.9% (11); Aristotle 13.2% (5); i360 5.3% (2); Catalist 2.6% (1); Other31.6% (12; e.g., TargetSmart, DataTrust, Bonfire, mixed-vendor setups); blank 18.4% (7).

Targeted recruitment: Took steps to increase representation of likely Trump voters 23.7% (9); No 76.3% (29).

Notes: Percentages may not sum to 100% due to rounding and multi-select items. "Always/Sometimes/Never" rows use N=38 unless labeled (LV-modeler rows use n=29).

J.4 Use of Survey Results in the Report

These aggregate responses were used in two main ways:

1. To assign default design characteristics to firms when poll-level metadata was missing (e.g., assuming a firm that “always” uses party weighting likely did so across their polls).
2. To characterize broader field practices and contextualize variability in methodological choices (e.g., in Section 3.3’s discussion of innovation and convergence).

The raw anonymized response data are retained by the task force but will not be published to preserve respondent confidentiality.

Appendix K. Replication Files and GitHub Link

All cleaned poll-level data, coding scripts, and statistical output underlying this report will be made publicly available upon publication through the task force's GitHub repository. These materials are intended to support replication, secondary analysis, and public transparency.

K.1 Repository Contents

The public repository will include:

- **Poll-level dataset**: All 2024 polls meeting inclusion criteria (n = 2,631), including harmonized metadata and repost weights
- **Scripts** to:
    - Clean and deduplicate aggregator data
    - Apply inclusion filters
    - Code methodological features from summaries
    - Merge benchmark vote returns
    - Generate all figures and tables in the report
- **Historical poll accuracy files** (2016–2022), processed under the same 14-day final window rules
- **Variable codebook** (matching Appendix B)
- **Reproducibility log** showing file interdependencies and runtime ordering

GitHub URL:

https://github.com/umisrcps/aapor_preelection_2024

K.2 Microdata and Restricted Files

Microdata files contributed by polling organizations will not be made public by the task force. These files are stored on a secure drive and were accessed only by task force members performing aggregation.

Aggregate results derived from these files—such as subgroup accuracy plots, county coverage residuals, and demographic weighting comparisons—are fully reproducible from summary statistics contained in the public scripts.

K.3 Use and Citation

The replication materials are provided under a CC-BY 4.0 license. Users are free to reuse, adapt, and publish derivative work with appropriate citation:

AAPOR Task Force on 2024 Pre-Election Polling: An Evaluation of the 2024 General Election Polling;
https://aapor.org/wp-content/uploads/2025/10/AAPOR-Task-Force-on-2024-Pre-Election-Polling_Report.pdf

Appendix L. Public Narrative and Context

While the primary focus of this report is on the empirical accuracy of public pre-election polls, the environment in which those polls were conducted—and interpreted—cannot be ignored. Polling exists not only as a technical endeavor but as a visible public artifact, shaped by media framing, campaign strategy, and public trust. This appendix provides a brief overview of how recent cycles have shaped perceptions of polling and the narratives that emerged in the lead-up to 2024.

L.1 Historical Background: A Crisis of Confidence

The 2016 and 2020 presidential elections fundamentally changed how polling was viewed by the public and political elites. Although national polls in both years were reasonably accurate in aggregate, state-level errors—especially in key swing states—led to widespread claims that "polling is broken." These critiques came not only from partisans whose candidates underperformed expectations, but also from media outlets, data journalists, and academics who questioned pollsters' assumptions, response rates, and weighting strategies.

After 2020 in particular, polling organizations faced intense scrutiny. The American Association for Public Opinion Research's (AAPOR) 2020 Task Force Report concluded that error was largest in the estimates of Republican vote share and could not be fully explained by design decisions. That finding raised alarms about systematic nonresponse, turnout modeling, and the adequacy of voter-file-based sampling. The narrative of failure—regardless of its nuance—stuck.

L.2 Media Coverage and Forecasting Ecosystem

The rise of poll aggregators, forecast models, and data-driven journalism further complicated the public's understanding of polling accuracy. Outlets such as FiveThirtyEight, The Economist, and RealClearPolitics produced real-time forecasts using poll-based models, often blending polling data with structural indicators like fundraising, past vote, and economic fundamentals.

This ecosystem helped temper overreaction to any single poll, but it also created the perception that *polls and models were interchangeable*. When a forecast missed, polling was often blamed—even if the model's assumptions about turnout, undecided voters, or aggregation methods were the true source of error.

Moreover, the proliferation of partisan and low-transparency polls further eroded public trust. Some firms released selective toplines, withheld crosstabs, or used modeling strategies that diverged from mainstream practice—all while being included in public averages. This contributed to skepticism, especially among political elites.

L.3 Polling in 2024: A High-Scrutiny Environment

Heading into 2024, many observers viewed polling with suspicion. Editorials and op-eds frequently questioned whether pollsters had learned anything since 2020. A particularly volatile summer—marked by a presidential dropout, replacement at the top of the Democratic ticket, and a near-miss assassination attempt—made conditions even more challenging.

Yet despite these headwinds, most 2024 public polling was notably accurate. This result surprised many skeptics—though trust remains fragile. While pollsters succeeded in describing the final margins, there was still wide disagreement in subgroup estimates, voter composition assumptions, and turnout projections. These tensions suggest that polling's public credibility will remain a contested terrain—even in cycles when toplines get it mostly right.

L.4 Implications for the Future

Understanding public perception is critical because it shapes how polling is funded, reported, and used. A single high-profile polling failure can reduce public cooperation, increase regulatory pressure, or discourage transparency. By contrast, cycles like 2024—with demonstrable improvements and empirical rigor—offer an opportunity to reset expectations and clarify what polling can (and cannot) do.

The findings in this report support a more optimistic view of election polling, but they also highlight the importance of clear communication, disclosure, and realistic framing of uncertainty. As polling continues to evolve, so too must the way it is presented and understood.