



AMERICAN ASSOCIATION FOR
PUBLIC OPINION RESEARCH

An Evaluation of the Methodology of the 2008 Pre-Election Primary Polls

Revised Version, April 2009

Prepared by:

The American
Association for Public
Opinion Research
(AAPOR) Ad Hoc
Committee on the 2008
Presidential Primary
Polling



March 30, 2009

To: Richard Kulka, Nancy Mathiowetz, Peter Miller
From: Ad Hoc Committee on the 2008 Presidential Primary Polling

Attached herewith is the report of our committee. The executive summary provides the key conclusions of our panel, and we urge the Council to study those points. We believe, however, that there are issues that deserve internal AAPOR consideration and actions regarding the current AAPOR standards enumerated in Best Practices for Survey and Public Opinion Research.

a) The world of survey research now uses more complicated and diverse sampling frames and selection techniques.

We inhabit a world where the sampling frames used for studies of the same population are quite diverse, where the separation between volunteering and being approached because you were selected randomly is fuzzy. The field now uses technologies where the selection of respondents is not straightforward (e.g., IVR measurement).

b) The world of survey research uses more complicated and diverse statistical adjustments for errors of non-observation.

Propensity models are increasingly used as an adjustment tool; some firms claim that their models are trade secrets, not to be disclosed. The adjustment for non-response is combined with adjustment for coverage and likely voting in ways that cannot be disconnected.

c) The current world of survey research uses sampling techniques that do not easily yield themselves to proper sampling variance estimation

Techniques that cannot assign known probabilities of selection to sampling frame elements produce sampling variance estimates with great difficulty. Extrapolation from variance estimates for simple random samples is inappropriate; more detailed guidance on disclosing such techniques and their implication would serve the membership well.

While we do not at this time have new language to propose (nor do we think it appropriate to our mission), we do urge the appointment of a committee to create a proposal to bring the disclosure standards more in tune with current practice.

An Evaluation of the Methodology of the 2008 Pre-Election Primary Polls

Prepared by the American Association for Public Opinion Research
Ad Hoc Committee on the 2008 Presidential Primary Polling

*Revised Version
April 2009*

Members of the Ad Hoc Committee:

Glen Bolger, Public Opinion Strategies

Darren W. Davis, University of Notre Dame

Charles Franklin, University of Wisconsin-Madison

Robert M. Groves, Institute for Social Research, University of Michigan

Paul J. Lavrakas, methodological research consultant

Mark S. Mellman, The Mellman Group

Philip Meyer, University of North Carolina

Kristen Olson, University of Nebraska-Lincoln

J. Ann Selzer, Selzer & Company

Michael W. Traugott (Chair), Institute for Social Research, University of Michigan

Christopher Wlezien, Temple University



An Evaluation of the Methodology of the 2008 Pre-Election Primary Polls

Table of Contents

Foreword	3
Executive Summary	5
Introduction	9
Analyzing Accuracy	12
Organization of the Committee's Work	16
How Well Did the Polls Do?	20
Explanations for Differences in Accuracy	26
Mode of Data Collection	26
Sample Frames and Respondent Selection	28
Non-coverage of Cell Phone Only Voters	31
Nonresponse Bias and the Composition of Samples	32
Trial Heat Question Wording	37
Likely Voter Definitions	44
Calling Protocols	48
Timing of Data Collection	51
Social Desirability	53
Weighting	58
Time of Decision	63
Participation of Independents	68
Allocations of 'Undecideds'	69
Ballot Order Effects	70
Conclusions	73
References	79
Appendix A: Committee Members	81
Appendix B: Charge to the Committee	82
Letter from AAPOR President Mathiowetz	83
Appendix C: AAPOR Request for Information	84
Roper Center Guidelines for Depositing Data	87
Appendix D: Hypotheses for Errors	89
Appendix Table 1	91
Appendix Table 2	92
"A Review and Proposal for a New Measure of Poll Accuracy"	93

Foreword

This report was prepared by a special committee appointed by the American Association for Public Opinion Research (AAPOR). The problem that prompted the formation of the committee was poll performance leading up to the New Hampshire primary. The members of the committee volunteered their time in this effort because of their interest in understanding what happened with estimation of candidate preference in the 2008 polls. Their expertise in various aspects of polling and survey research methods guided the analysis and write-up of the results. From the beginning, committee members decided that they would pursue their investigation empirically. But there was no opportunity to design prospective research into polling methods. The fact that the committee's work began only after questions were raised about the quality of the polling in the early primaries meant that some avenues of inquiry could not be pursued. Appropriate data to explore these matters were not available (or were not made available) to the committee by those who conducted the polls. This report represents the committee's best effort to address these issues raised by the 2008 pre-primary polls within the constraints of limited available information.

Initially, we expected the committee to work rapidly and complete its report by the annual AAPOR conference in May 2008. However, the slow response of many pollsters and a lack of cooperation from some delayed the analysis and subsequent reporting. The fact that many pollsters did not provide us with detailed methodological information about their work on a timely basis is one reason we will never know for certain exactly what caused the problems in the primary polling that we studied. It is also true that some of the more interesting questions about the causes are not amenable to post hoc analysis. While the available data allowed us to disconfirm some hypotheses and provide some tantalizing clues about what went wrong, definitive declarations about the sources of estimation errors are not possible.

Polling during an election campaign is an important element of news coverage of related events. The symbiotic relationship between campaign coverage and polling is a given in contemporary campaigns; it is hard to imagine one without the other. But polling is also a scientific data collection technique, and it is impossible to evaluate the performance of the pollsters without information about their methodology. That is why the AAPOR "Code of Professional Ethics and Practices" include a set of elements that those who conduct polls and surveys should disclose so that other professionals can evaluate the quality of the research that they conduct and the results that they disseminate. The committee's experience suggests that some firms engaged in election polling pay only lip service to

these disclosure standards, while others are operating at such a thin margin that they do not have the resources to devote to answering questions about their methods. The committee believes that professional organizations like AAPOR, the Council of American Survey Research Organizations (CASRO), and the National Council on Public Polls (NCPP) should review their published standards with an eye toward an educational program to explain them and reinforce the underlying justification for having them and to promote more effective enforcement of them.

A number of acknowledgments are due. The committee's efforts were supported by a grant from James S. Jackson, director of the Institute for Social Research (ISR) at the University of Michigan. These funds were used to support research assistance and administrative costs, and several of the analyses would not have been possible without this support. In particular, these resources were used to support a content analysis of the press coverage of polling around the New Hampshire primary; these results are discussed in the introduction to the report. This content analysis was conducted at ISR by Colleen McClain and Brian Krenz, while the analysis of the information received from the pollsters was performed by Courtney Kennedy, a graduate student in the Program in Survey Methodology at ISR, all under my supervision. Ms. Kennedy also participated in the organization, design and layout of the report, and she drafted some sections. The pollsters who provided information to the committee were given a draft to review for accuracy with regard to statements about their procedures; their comments were helpful. The final report was edited by N.E. Barr, who significantly improved the prose. As always, the contents of the report are the sole responsibility of the committee and no one else.

This version of the report incorporates some small editorial changes and supersedes the March 2009 version.

Michael Traugott
Ann Arbor, Michigan

Executive Summary

This report is the product of an investigation by the Ad Hoc Committee on the 2008 Presidential Primary Polling appointed by the American Association for Public Opinion Research (AAPOR). The committee pursued its investigation by analyzing information that AAPOR solicited from the pollsters who conducted studies in four primaries in the 2008 presidential campaign: New Hampshire, South Carolina, California, and Wisconsin. A central concept underlying the role of public opinion research and polling as a scientific enterprise is the disclosure of the methods used. While most citizens cannot make informed judgments about the quality of polling data or their interpretation, other professionals in the field can, if they have access to a minimal set of information about how the data were collected and analyzed. The failure of some pollsters to provide information on a timely basis runs counter to this principle, and it hindered the progress of the committee's work and delayed the release of this report. While the results of this investigation could not affect subsequent polling in the 2008 campaign, the committee hopes that its work will raise questions for consideration about the methodology of subsequent pre-election polling. We also hope that the report will spur timely disclosure of information to aid in future evaluations of the methodological quality of pre-election polling.

The committee developed and tested a series of hypotheses that could be tested empirically, employing information at the level of the state, the poll, and, in limited cases, the respondent. Since the analysis was conducted after data collection, it was not possible to evaluate all of the hypotheses in a way that permitted strong causal inferences. And due to the incomplete nature of the data for various measures, it was not possible to pursue all hypotheses about what might have happened, nor was it possible to pursue multivariate analyses that looked simultaneously at multiple explanatory factors. In the end, the analysis suggests some possible explanations for the estimation errors and the unlikely impact of other factors. The research also highlights the need for additional disclosure requirements such as information on likely voter models and the details of weighting algorithms, as well as the need for better education by professional associations like AAPOR, the Council of American Survey Research Organizations (CASRO), and the National Council on Public Polls (NCPP).

Polling in primary elections is inherently more difficult than polling in a general election. Usually there are more candidates in a contested primary than in a general election, and this is especially true at the beginning of the presidential selection process. For example, there were a total of 15 candidates entered in the Iowa caucuses and more than 20 names on the New Hampshire primary ballot. Since primaries are within-party events, the voters do not have the cue of party identification to rely upon in making their choice. Uncertainty in the voters' minds can create additional problems for pollsters. Turnout is usually much lower in primaries than in general elections, although it varies widely across events. Turnout in the Iowa caucuses tends to be relatively low compared to the New Hampshire primary, for example. So estimating the likely electorate is often more difficult in primaries than in the general election. Furthermore, the rules of eligibility to vote in the primaries vary from state to state and even within party; New Hampshire has an open primary in which independents can

make a choice at the last minute in which one to vote. All of these factors can contribute to variations in turnout, which in turn may have an effect on the candidate preference distribution among voters in a primary election.

The estimation errors in the polls before the New Hampshire Democratic primary were of about the same magnitude as in the Iowa caucus. But the mis-estimation problems in New Hampshire received much more – and more negative – coverage than they did in Iowa. Because of a small level of undecided voters in every poll, the estimates for each individual candidate were generally lower than the proportion of votes they received. And these underestimates tended to be greater for the first place finisher than the second place finisher. But the majority of the polls before New Hampshire suggested the wrong winner, while only half in Iowa did.

Factors that may have influenced the estimation errors in the New Hampshire pre-primary polls include:

1. Given the compressed caucus and primary calendar, polling before the New Hampshire primary may have ended too early to capture late shifts in the electorate there, measuring momentum as citizens responded to the Obama victory in the Iowa caucus but not to later events in New Hampshire.
2. Patterns of nonresponse, derived from comparing the characteristics of the pre-election samples with the exit poll samples, suggest that some groups that supported Senator Hillary Clinton were underrepresented in the pre-election polls.
3. Variations in likely voter models could explain some of the estimation problems in individual polls. Application of the Gallup likely voter model, for example, produced a larger error than their unadjusted data. While the “time of decision” data do not look very different in 2008 compared to recent presidential primaries, about one-fifth of the voters in the 2008 New Hampshire primary said they were voting for the first time. This influx of first-time voters may have had an adverse effect on likely voter models.
4. Variations in weighting procedures could explain some of the estimation problems in individual polls. And for some polls, the weighting and likely voter modeling were comingled in a way that makes it impossible to distinguish their separate effects.
5. Although no significant social desirability effects were found that systematically produced an overestimate of support for Senator Obama among white respondents or for Senator Clinton among male respondents, an interaction effect between the race of the interviewer and the race of the respondent did seem to produce higher support for Senator Obama in the case of a black interviewer. However, Obama was also preferred over Clinton by those who were interviewed by a white interviewer.

Factors unlikely to have contributed to the estimation errors in the New Hampshire pre-primary polls include:

1. The exclusion of cell phone only (CPO) individuals from the samples did not seem to have an effect. However, this proportion of citizens is going to change over time, and pollsters should remain attentive to its possible future effects.
2. Using a two-part trial heat question, intended to reduce the level of “undecided” responses, did not produce that desired effect and does not seem to have affected the eventual distributions of candidate preference.
3. The use of either computerized telephone interviewing (CATI) techniques or interactive voice response (IVR) techniques made no difference to the accuracy of estimates.
4. The use of the trial heat questions was quite variable, especially with regard to question order, but no discernible patterns of effects on candidate preference distributions were noted. While the names of the (main) candidates were frequently randomized, the committee did not receive data that would have permitted an analysis of the impact of order.
5. Little compelling information indicates that Independents made a late decision to vote in the New Hampshire Republican primary, thereby increasing estimate errors.

Factors that present intriguing potential explanations for the estimation errors in the New Hampshire polls, but for which the committee lacked adequate empirical information to thoroughly assess include:

1. The wide variation in sample frames used to design and implement samples – ranging from random samples of listed telephone numbers, to lists of registered voters with telephone numbers attached, to lists of party members – may have had an effect. Greater information about sample frames and sample designs, including respondent selection techniques, would facilitate future evaluations of poll performance.
2. Differences among polls in techniques employed to exclude data collected from some respondents could have affected estimates. Given the lack of detailed disclosure of how this was done, it is not possible to assess the impact of this procedure.
3. Some polls combined weighting to adjust for nonresponse among demographic groups with weighting that reflects likely voter models into a single set of weights for a study. This complicates the analysis of whether or how much sampling issues or likelihood of voting models are contributing to estimation error.

Finally, factors that appeared to be potential explanations for estimation errors, but for which the committee lacked any empirical information to assess include:

1. Because of attempts by some states to manipulate the calendar of primaries and caucuses, the Iowa and New Hampshire events were rescheduled to the first half of January, with only five days between the events, truncating the polling field period in New Hampshire following the Iowa caucus.
2. The order of the names on the ballot – randomly assigned but fixed on every ballot - may have contributed to the increased support that Senator Hillary Clinton received in New Hampshire.

All of the information provided to the committee is being deposited in the Roper Center Data Archive, where it will be available to other analysts who wish to check on the work of the committee or to pursue their own independent analysis of the pre-primary polls in the 2008 campaign.

Introduction

Polling in primary elections is inherently more difficult than polling in a general election. Usually there are more candidates in a contested primary than in a general election, and this is especially true at the beginning of the presidential selection process. For example, there were a total of 15 candidates entered in the Iowa caucuses.¹ Since primaries are within party events, the voters do not have the cue of party identification to rely on in selecting their choice. This level of uncertainty in the voters' minds can create additional problems for the pollsters. Turnout is usually much lower in primaries than in general elections, although it varies widely across events. So estimating the likely electorate is often more difficult in primaries than in the general election. Furthermore, the rules of eligibility to vote in the primaries vary from state to state and even within party; New Hampshire has an open primary in which independents (those with an undeclared party registration) can make a choice in which one to vote. All of these factors can contribute to variations in turnout, which in turn may have a great effect on the candidate preference distribution among voters in a primary compared to the general election.

In the 2008 primary campaign, the record of the polls in estimating outcomes differed in the Democratic and Republican events. This could be explained by a number of factors, not the least of which is that the Democratic contest was hard fought and went on for the entire calendar of events, while the Republicans had selected John McCain as their presumptive nominee by March. On the Democratic side, the percent of actual votes cast for the winner among the votes cast for the top two candidates tended to be greater than the same ratio for the winner in the final week's pre-election polls. That is, the polls generally underestimated the winner's performance on Election Day relative to the second place finisher, although analysis shows that, by this measure, the performance of the polls may have improved slightly over the course of the primary calendar.

This relationship was quite different in the Republican contests. In the early contests – and up through Super Tuesday – a similar pattern of underestimating the winner's share of the vote for the top two candidates appears. The winner's share of support for the top two candidates in the polls was generally less than his share of the actual vote for the same two candidates, though less consistently (and to a lesser extent) than on the Democratic side. In the later primaries, after McCain essentially secured the nomination, this tendency disappeared. For a while the polls tended to exhibit less bias with respect to the winner, and then they tended consistently to overestimate the winner's share of the

¹ In the New Hampshire Democratic primary, there were 19 candidates officially named on the ballot, and 19 Democrats received write-in votes; only five of these candidates received more than 1,000 votes out of almost 300,000 cast. In the Republican primary, there were 22 candidates named on the ballot, and 21 additional candidates received votes. In total, only nine candidates received more than 1,000 votes out of almost 250,000 cast.

two-candidate vote. This analysis suggests that there were factors associated with the contests themselves, including the level of competition or the uncertainty of the outcome, that seem to have had an effect on the quality of estimation.²

The impetus for this report was the performance of the polls prior to the New Hampshire primary. In the run up to the January 8 event, the pre-election polls showed Senator Barack Obama with a comfortable lead over Senator Hillary Clinton, while Senator John McCain was holding a steady lead over former governors Mitt Romney and Mike Huckabee. McCain won about as expected on the Republican side, while Clinton bested Obama by three percentage points.

Although the Republican contest ended up about as the polls showed, the mis-estimation of the Democratic race caused much consternation in the press and within the polling profession. The errors in the Democratic estimates in Iowa, discussed briefly in the beginning of this report, were about the same magnitude as in New Hampshire, but the level of criticism was much lower because several polls correctly projected an Obama victory. In New Hampshire, the pollsters got the Democratic winner wrong, and that made all the difference in the press coverage and commentary about the polls. It also determined that the pre-primary polls in New Hampshire would be a focal point for our inquiry, although other states were added to the list, as noted below.

A content analysis of the media coverage in six sources³ for the first twelve days of January 2008 showed that concerns about the portrayal of the polling industry were warranted. The coverage of polls increased in amount after New Hampshire, turned more negative than in the period leading up to the primary, and became more focused on "the polls" as a group rather than on specific estimates produced by individual polling firms, as shown in Table 1 (Traugott, Krenz, and McClain 2008).

Table 1. References to "The Polls" Increased after the New Hampshire Primary

Time period	% References to "the polls"	N
Before Iowa caucuses	38%	(164)
Between Iowa and New Hampshire	49%	(673)
After New Hampshire primaries	70%	(909)

The denominator is the total number of references that mentioned poll or polling during the time period indicated.

² Additional information about this analysis can be obtained from Christopher Wlezien, Department of Political Science, Temple University: Wlezien@temple.edu.

³ The news outlets reviewed in the media content analysis were CNN, FOX News, CBS News, the *New York Times*, the *Washington Post*, and the *Boston Globe*.

Furthermore, the most negative elements of the post-New Hampshire coverage were the references to the (lack of) accuracy of the polls, as shown in Table 2. These trends raised concerns among the professional associations whose members are pollsters and survey researchers because they understood the unusual relationship between the accuracy of pre-election polls and the image of the entire industry. Most public polling estimates do not have external validation, but pre-election polling is the special case where the results of the election itself validate the quality of the estimation. The leadership of the American Association for Public Opinion Research (AAPOR) also believed that, given the scientific basis for polling, it should be possible to explore reasons for the estimation problems.

Table 2. The Valence of Poll References in Media Coverage by Topic

Topic of reference	Positive	Neutral	Negative	N
Accuracy*	11%	13	75	(149)
Margin of error/Sampling issues	0%	78	22	(82)
Trial heat	1%	97	2	(1,064)
Any other reference	0%	99	1	(451)
TOTAL	1%	89	9	(1,746)

*Almost all of these references came after the New Hampshire primaries.

The parallels between these concerns and those expressed after the 1948 general election were obvious, and AAPOR decided to empanel a group of experts to investigate the potential explanations for the mis-estimates in the New Hampshire pre-primary polls. The committee comprised academic experts, public pollsters, and partisan pollsters who work for candidates, although not for candidates in the 2008 presidential campaign. By agreement, our work was empirical, involving only the evaluation of possible explanations that could be investigated with information about how the polls were conducted or through the analysis of data collected from them. Since the investigation was not planned ahead of time, the committee often found itself without the appropriate data with which to test some hypotheses.

The work of the committee, and hence this report, has been delayed by a slow response from many of the pollsters who collected data from the four states in which the committee focused its efforts – New Hampshire, South Carolina, Wisconsin, and California.⁴ This is quite a different situation than after the 1948 general election, when there were fewer firms engaged in public polling, the threat to the future of the industry seemed to be greater, and the polling firms were fully cooperative. In 2008, many of the firms that polled in New Hampshire had studies in the field for primaries that occurred right after that. Today, there are well-publicized standards for disclosure of information about how polls are conducted. AAPOR, an organization of individuals engaged in public opinion research; the National Council on Public Polls (NCPP), an organization of organizations that conduct public opinion research; and the Council of American Survey Research Organizations (CASRO), also an organization of organizations, have all promulgated standards of disclosure. Despite the norms, at the time this report was finalized, one-fifth of the firms from which information was requested had not provided it. For each of these four firms, we were able to retrieve some of the requested information through Internet searches, but this was incomplete at best. If additional information is received after the report’s release, the database at the Roper Center will be updated.

Analyzing Accuracy

In this evaluation of the pre-primary and caucus polls, we measured their accuracy using the statistic A ,⁵ the natural log odds ratio of the relative standing of two candidates in a poll compared to their relative standing in the actual election returns. In the case of each primary that was analyzed, the denominator of A was the same for all polls; the difference in the value of A can be attributed to differences in the numerator, which was calculated from the poll estimates themselves. Primaries are multi-candidate events, often with very large fields of competing candidates at the beginning of the presidential nomination process. We used A in two different ways to measure the relative standing of

⁴ The last pieces of information were supplied between October 15, 2008 and March 13, 2009 after a final request for information and a reminder that the responses to the request would be fully disclosed in the committee’s report. Some additional information arrived on or after February 10, 2009, after the AAPOR Standards Chair contacted several organizations about various elements of disclosure that were lacking.

⁵ This measure is described in detail in Martin, Traugott, and Kennedy (2005). A is a statistic that captures both the magnitude and the direction of estimation errors by computing the log odds ratio of the vote shares in the election for the two leading candidates compared to the vote shares in a pre-election poll of the same two candidates. Other measures of polls accuracy are available (Mitofsky 1998) but are not included here in order to simplify the analysis. A copy of the original article is appended to this report.

the top candidates in the voting compared to the estimate of their standing in the polls. For the Democrats, we consistently measured Obama's share compared to Clinton's. For the Republicans, we measured the first place finisher's share compared to the second place finisher's – a pairing that changed from event to event as clearly noted in each analysis. The closer A is to zero, the closer the odds ratio is to 1.0 and the more accurate the poll is in relation to the election outcome. In Democratic contests, negative values denote an underestimation of Obama's vote share compared to Clinton's vote share (or an overestimation of Clinton's share compared to Obama's), relative to the actual outcome of the election. Positive values indicate an overestimation of the Obama's share compared to Clinton's (or an underestimation of Clinton's share compared to Obama's). In Republican contests, negative values denote an underestimation of the winner's vote share compared to the second place finisher in the poll relative to the actual outcome of the election. A positive value indicates an overestimation of the first place finisher's share compared to the second place finisher's share.

Information presented in Table 3, based upon each firm's last poll before the relevant event, suggests that the pre-voting polls on the Democratic side in Iowa were no more accurate than those in New Hampshire, by both the average absolute value of A (0.26 in Iowa and 0.27 in New Hampshire) and the range of values (a comparison of a range from 0 to .55 in Iowa to a range of .10 to .45 in New Hampshire). Table 3 also shows the differences in the point estimates for the first and second place finishers; most of these differences are negative because the total of the votes cast add to 100%, but the poll estimates include some "undecided" voters. The Democratic polls generally did well in projecting the vote share of the second place finisher, but they were off severely in estimates for the winner's proportion. In both Iowa and New Hampshire, the polls in the Democratic race understated the winner's vote share by 9 to 10 percentage points, but came within one point of projecting the loser's vote share. On the Republican side, the polls also underestimated the winner's proportion in both Iowa and New Hampshire by almost four percentage points on average, but there was only a one percentage point difference in Iowa in the estimates of the runner up's proportion.

Table 3. Election Results and Polls in Iowa and New Hampshire, Ordered by Absolute Value of A

Iowa Democratic Caucuses						Iowa Republican Caucuses						
	Obama	Clinton	A	Difference on winner	Difference on 2nd place		McCain	Romney	Huckabee	A	Difference on winner	Difference on 2nd place
Election Result	38%	29%				Election Result	13%	25%	34%			
Selzer & Co.	32	25	0.00	-6.0	-4.0	Insider Adv.	11	24	30	-0.08	-4.0	-1.0
Zogby	31	24	0.01	-7.0	-5.0	Research 2000	8	27	34	-0.08	0.0	2.0
Strategic Vision	32	27	-0.07	-6.0	-2.0	Zogby	10	25	31	-0.09	-3.0	0.0
Research 2000	29	28	-0.21	-9.0	-1.0	Selzer & Co.	13	26	32	-0.10	-2.0	1.0
Mason-Dixon	22	23	-0.29	-16.0	-6.0	ARG	11	24	29	-0.12	-5.0	-1.0
ORC	31	33	-0.31	-7.0	4.0	LA Times	11	23	37	0.17	3.0	-2.0
LA Times	26	29	-0.35	-12.0	0.0	Strategic Vision	16	30	28	-0.38	-6.0	5.0
ARG	25	34	-0.55	-13.0	5.0	ORC	10	31	28	-0.41	-6.0	6.0
Insider Adv.	22	30	-0.55	-16.0	1.0	Mason-Dixon	13	27	23	-0.47	-11.0	2.0
Mean	27.8	28.1	-0.26	-10.2	-0.9	Mean	11.4	26.3	30.2	-0.17	-3.8	1.3
New Hampshire Democratic Primary						New Hampshire Republican Primary						
	Obama	Clinton	A	Difference on winner	Difference on 2nd place		McCain	Romney	Huckabee	A	Difference on winner	Difference on 2nd place
Election Result	36%	39%				Election Result	37%	32%	11%			
Research 2000	34	33	0.10	-6.0	-2.0	Univ. of NH	31	26	13	0.02	-6.0	-6.0
Mason-Dixon	33	31	0.13	-8.0	-3.0	Research 2000	35	29	13	0.03	-2.0	-3.0
RKM	34	31	0.16	-8.0	-2.0	Gallup	34	30	13	-0.03	-3.0	-2.0
Op. Dynamics	32	28	0.20	-11.0	-4.0	Marist College	35	31	13	-0.04	-2.0	-1.0
Suffolk	39	34	0.21	-5.0	3.0	Op. Dynamics	34	27	11	0.07	-3.0	-5.0
Marist College	34	28	0.26	-11.0	-2.0	ARG	31	24	14	0.10	-6.0	-8.0
Rasmussen	37	30	0.28	-9.0	1.0	Strategic Vision	35	27	13	0.10	-2.0	-5.0
CBS News	35	28	0.29	-11.0	-1.0	RKM	38	29	9	0.11	1.0	-3.0
ARG	40	31	0.32	-8.0	4.0	Zogby	36	27	10	0.13	-1.0	-5.0
Univ. of NH	39	30	0.33	-9.0	3.0	Mason-Dixon	32	24	12	0.13	-5.0	-8.0
Strategic Vision	38	29	0.34	-10.0	2.0	Rasmussen	32	31	10	-0.13	-5.0	-1.0
Zogby	42	29	0.44	-10.0	6.0	Suffolk	26	30	13	-0.30	-11.0	-2.0
Gallup	41	28	0.45	-11.0	5.0	Mean	33.3	27.9	12.0	0.02	-3.8	-4.1
Mean	36.8	30.0	0.27	-9.0	0.8							

Source for election results: http://nass.org/index.php?option=com_content&task=view&id=89&Itemid=223

Estimating the outcome of the Iowa caucuses⁶ is complicated by a two-step process of measuring preferences: supporters of a candidate who does not receive votes from at least 15% of the caucus delegates who voted in the first round can express a second choice in the next round. Barack Obama finished first with 37.6% of the eventual Democratic caucus delegates, John Edwards second with 29.8%, and Hillary Clinton third with 29.5%. Most of the Iowa pre-caucus polls underestimated the size of Barack Obama's margin over Hillary Clinton while getting the winner correct in about half of the estimates. On the Republican side, Mike Huckabee finished first with 34.4% support from those caucus attendees, Mitt Romney second with 25.2%, Fred Thompson third with 13.4%, and John McCain fourth with 13.0%. Here the polls consistently underestimated support for Huckabee, and three suggested that Romney would finish first.

In the New Hampshire primary, which came only five days later amid a flurry of coverage of the Obama victory in Iowa, every one of the pre-primary polls showed him in the lead, although the margin was often by a statistically insignificant amount. When the votes were counted, however, Hillary Clinton had won 39% of the vote while Barack Obama received 36% of the vote, a narrow but unexpected Clinton victory. While these polls were not far off on their estimate of support for Obama, they all underestimated support for Clinton. The range of the absolute values for A , all of which were positive, was from .10 to .45. On the Republican side, 10 out of 12 polls suggested a McCain victory, with one showing the contest between McCain and Mitt Romney very close and one suggesting that Romney was in the lead. The range of the absolute values of A went from .02 to .30. In this case, the estimates were generally predictive of the winner.

The inability of the vast majority of the pre-primary polls in New Hampshire to estimate the Clinton victory accurately, even while they estimated the Republican outcome reasonably well, produced consternation in the national media about the performance of "the polls." This in turn prompted discussion among the leadership of the American Association for Public Opinion Research (AAPOR) because political polling represents the public face of the public opinion and marketing research professions. Given the virtually unique circumstances of having a source of external validation for published survey estimates – in this case, the election results – the potential for public reaction to pre-primary polling estimates is great, and these reactions may influence the public's more general view of the industry, and thus its reputation.

⁶ The election results reported for the Iowa Democratic caucuses are the relative proportions of the delegate totals, not popular vote totals. The Iowa Republican Party releases the results of a straw poll conducted as part of their caucus, but the state Democratic Party does not.

Organization of the Committee's Work

In her charge to the AAPOR committee formed to investigate the estimation problems in the pre-New Hampshire polls (see Appendix A for a list of all committee members), then President Nancy Mathiowetz asserted: “What we learn from this review will help us to continue to improve our methodology and ensure continued accuracy.” (The formal charge to the committee is contained in Appendix B.) When organizing the work of the committee, chair Michael Traugott of the University of Michigan proposed an empirical investigation of a series of possible explanations for the problems. To learn more about what might have happened in the New Hampshire Democratic pre-primary polls, the committee agreed to expand its analysis in limited ways to include investigation of the estimations in both the Republican and Democratic pre-election polls in four states: New Hampshire, South Carolina, Wisconsin, and California. As will be explained later, the polls generally underestimated support for Barack Obama in South Carolina and overestimated support for Hillary Clinton in California. In addition, the unusually large numbers of undecided respondents in South Carolina and Wisconsin suggested that analyzing those polls would yield insightful results.

The committee decided upon a number of data elements that it would need to pursue its analysis of possible explanations for the differences between the final pre-election estimates and the actual primary outcomes in the selected states. The next step was to obtain information about each survey to support the committee’s work. President Mathiowetz took on the task of recruiting information from the 21 polling organizations that produced publicly reported estimates in any of the four states within two weeks of the primary election.⁷ Her request included information that public pollsters subscribing to the Standards for Minimal Disclosure as part of the AAPOR Code of Ethics and Professional Practice could be expected to reveal under normal circumstances.⁸ Beyond these minimal items, Mathiowetz’s layered request asked for information not part of the minimal disclosure requirements, but that would help the committee with its work. This information included a copy of the micro dataset from the survey, as well as data concerning interviewer characteristics, where applicable, and other administrative data from the data collection process such as calling information. (A copy of the disclosure request is included in Appendix C.) This request, dated March 4, 2008, came during a very busy part of the primary schedule itself, and many firms were engaged in continuing data collection that prevented them from responding immediately. And AAPOR was not in a position to offer financial assistance in the preparation of these additional materials, so the firms had to bear the cost themselves.

⁷ Some firms conducted surveys in more than one of the four states, and these multiple occurrences are counted in the 31 estimates.

⁸ For a complete description of these items, see <http://aapor.org/aaporcodeofethics>.

Table 4 shows the level of polling organization response to these requests.⁹ The rate of providing information and the specific information elements provided varied widely by firm. Despite repeated requests for information, at the time of the analysis of data for this report three firms never responded: Clemson University, Ebony/Jet, and Strategic Vision.

Table 4. Inventory of Responses to Request for Minimum Disclosure and Microdata

Polling Organization	Primary Polling States	Micro-data*	Interviewer Variables*	Question Wording	Weighting Description
American Research Group	CA, NH, SC, WI			W	Y
CBS News/New York Times	NH	Y	Y	Y	Y
Clemson University	SC			W	
Datamar	SC		n/a (IVR)	Y	Y
Ebony/Jet	SC				
Field Poll	CA	Y		Y	Y
Gallup/USA Today	NH	Y	Y	Y	Y
LA Times/CNN/Politico	CA, NH			Y	Y
Marist College	NH			Y	Y
Mason-Dixon/McClatchy/MSNBC	CA, SC			W	Y
Opinion Dynamics/FOX News	NH	Y		Y	Y
Public Policy Institute	CA	Y		Y	Y
Public Policy Polling	SC, WI		n/a (IVR)	Y	Y
Rasmussen Reports	CA, NH, SC		n/a (IVR)	Y	Y
Research 2000/Concord Monitor	NH, WI			Y	
RKM/Franklin Pierce/WBZ	NH			Y	Y
Strategic Vision	NH, WI			W	
Suffolk Univ./WHDH	CA, NH			Y	Y
SurveyUSA	CA, SC	Y	Y**	Y	Y
Univ. of New Hampshire/CNN/WMUR	NH	Y	Y	Y	Y
Zogby/Reuters/C-SPAN	CA, NH, SC			Y	W

Y: Information provided directly to the committee in response to the AAPOR request through March 2009.

W: Information obtained from materials posted on the Internet.

n/a: Information is not applicable

*This information was requested but is not required under the AAPOR Minimal Disclosure Standards.

**SurveyUSA used IVR, but they still provided the race and gender of the interviewers who read the questionnaire.

⁹ Some surveys had multiple sponsors plus a data collection firm (e.g., LA Times/CNN/Politico.com/ORC). To make the tables and text more readable, we employ a shorthand label for each survey, corresponding to the organization contacted by the committee.

Given the lack of information from several of the firms, the committee attempted to retrieve some of the minimum disclosure information from alternative sources, such as a firm's web site or news sources reporting on the study. The requested minimal disclosure items included: sponsorship (16 of the 21 firms or survey agencies provided this information), the exact wording of each question asked (16 firms provided this information directly and for 4 firms, the information was obtained from other sources), a description of the sampling frame (provided by or found for 19 of the firms), sample sizes and eligibility criteria (available for 33 surveys from 20 firms with varying levels of specificity for eligibility criteria), response rates (available for 17 surveys representing 12 firms), a description of the precision of the findings as assessed by a margin of error statement (provided by or found for all surveys), the weighting procedures (available for 17 firms), and the dates of the field period (provided by or found for all surveys except one).

The additional requested information included: a micro data file for analysis (provided for 7 surveys, including two from SurveyUSA),¹⁰ information on the characteristics of the interviewers (provided for three surveys and not directly appropriate for the four IVR surveys), and the call records (not provided by any firm, although two datasets included information on the call on which the interview was completed).

The major design features of the pre-election polls are presented in Table 5. All the polls were conducted over the telephone. Most used computer-assisted telephone interviewing (CATI), but a noticeable minority was conducted using interactive voice response (IVR) equipment. With IVR, the respondents hear a pre-recorded reading of the survey questions, and they register their answers verbally or manually on their telephone touchpad. The implications for estimation error stemming from data collection mode and other design features are investigated in the analysis below.

Summary: The committee commends the firms that did respond on a timely basis, especially those that provided micro datasets for extended analysis. Several of the firms that conducted pre-election polling in the 2008 primaries in New Hampshire, South Carolina, Wisconsin, and California were slow to disclose the details of their work to the committee. Beyond those that did not respond, others made incomplete information available in terms of AAPOR's minimal disclosure "requirements."

¹⁰ When subsequent information is presented by firm, there is only one entry for SurveyUSA as they employed the same methodology in each poll.

Table 5. Design Features of 2008 Primary Pre-election Polls

	Mode and sample design	Sample size for trial heat	Response rate	Field period	Within HH selection
New Hampshire Polls					
ARG	CATI using Landline RDD	n=600 LV	RR(1)=42%	Jan. 6-7	NA
CBS News	CATI using Landline RDD	n=323 Dem LV	Nov. RR(1)=8%; Jan. re-interview rate=77% among the likely Dem primary voters	Jan. 5-6	None (callback)
Gallup	CATI using Landline RDD + Cell RDD with cell-only screener	n=776 Rep LV, n=778 Dem LV	RR(2)=9%	Jan. 4-6	Last birthday
LA Times	CATI using Landline RDD	n=318 Rep LV, n=361 Dem LV	NA	Dec. 20-23,26	NA
Marist College	CATI using Landline RDD	n=477 Rep LV, n=636 Dem LV	RR(3)=37%	Jan. 5-6	Last birthday
Opinion Dynamics	CATI using Landline RDD	n=500 Dem LV	16%	Jan. 4-6	None
Rasmussen Reports	IVR using Landline RDD	n=1,774 Dem LV	NA	Jan. 5-7	NA
Research 2000	CATI using Landline RDD	n=400 Rep LV, n=400 Dem LV	NA	Jan. 4-5	None
RKM	CATI using RV lists	n=409 Rep, n=403 Dem	RR(5)=11%	Jan. 4-6	None (RV list)
Strategic Vision	CATI unspecified sample	n=600 Rep LV, n=600 Dem LV	NA	Jan. 4-6	NA
Suffolk University	CATI using Landline RDD	n=500 Dem LV, n=500 Rep LV	RR(1)=6%	Jan. 5-6	None
Univ. of NH	CATI using Landline RDD	n=492 Rep LV, n=599 Dem LV	between 20 - 25%	Jan. 5-6	Last birthday
Zogby	CATI using listed landlines	n=844 LV	NA	Jan. 4-6	asked for an RV in the HH
South Carolina Polls					
ARG	CATI using Landline RDD	n=600 LV	RR(1)=21%	Jan. 23-24	NA
Clemson University	CATI unspecified sample	NA	NA	Jan. 15-23	NA
Ebony/Jet	CATI using RV list	n=600 Dem LV	NA	NA	NA
Mason-Dixon	CATI using Landline RDD	n=400 Rep LV	RR(1)=32%	Jan. 14-16	NA
Public Policy Polling	IVR using RV list landlines	n=808 Rep, n=788 Dem	NA	Jan. 16 (R) Jan.	None
Rasmussen Reports	IVR using Landline RDD	n=624 LV	NA	Jan. 21	NA
SurveyUSA	IVR using Landline RDD	1,800+ oversample of 803 (women, whites, and seniors)	RR(4)=9%	Jan. 23-24	None
Zogby	CATI using listed landlines	n=816 Dem LV	NA	Jan. 24-25	asked for an RV in the HH
Wisconsin Polls					
ARG	CATI using Landline RDD	n=407 Dem LV	RR(1)=26%	Feb. 17-18	NA
Public Policy Polling	IVR using RV list landlines	n=701 Rep, n=1,411 Dem	NA	Feb. 16-17	None (RV list)
Research 2000	CATI using Landline RDD	n=400 Rep LV, n=400 Dem LV	NA	Feb. 13-14	None
Strategic Vision	CATI unspecified sample	n=800 LV	NA	Feb. 8-10	NA
California Polls					
Datamar	IVR using RV lists	n= 558 Rep RV, n=578 Dem RV	NA	Jan. 11-14	None (RV list)
Field Poll	CATI using RV lists	n= 481 Rep RV, n=511 Dem RV	RR(3)=10%	Jan. 25-Feb. 1	None (RV list)
LA Times	CATI using Landline RDD	n=518 Rep LV, n=918 Dem LV	NA	Jan. 23-27	Last birthday
Mason-Dixon	CATI using Landline RDD	n=400 Rep LV, n=400 Dem LV	RR(1)=12%	Jan. 30-Feb. 1	NA
Public Policy Institute	CATI using Landline RDD	n=348 Rep LV, n=543 Dem LV	RR(2)=12%	Jan. 13-20	Last birthday
Rasmussen Reports	IVR using Landline RDD	n=693 Rep LV, n=798 Dem LV	NA	Feb. 2	NA
Suffolk University	CATI using RV lists	n=500 Rep LV, n=700 Dem LV	RR(1)=5%	Jan. 31- Feb. 1 (R) Feb. 1-3 (D)	None (RV list)
SurveyUSA	IVR using Landline RDD	2,000 + oversample of 544 (women, whites, and seniors)	RR(4)=7%	Feb. 3-4	None
Zogby	CATI using listed landlines	n=833 Rep LVs, n=895 Dem LV	NA	Feb. 3-4	asked for an RV in the HH

NA indicates that the information was not supplied by the pollster and could not be located elsewhere.

RR indicates a response rate calculated according to an AAPOR formula. Any other response rate was supplied by a pollster but without an indication of how it was calculated.

Table entries are based upon information provided directly to the committee in response to the AAPOR request though mid March, 2009.

How Well Did the Polls Do? Evaluation of the Pre-election Polling Estimates in Five States¹¹

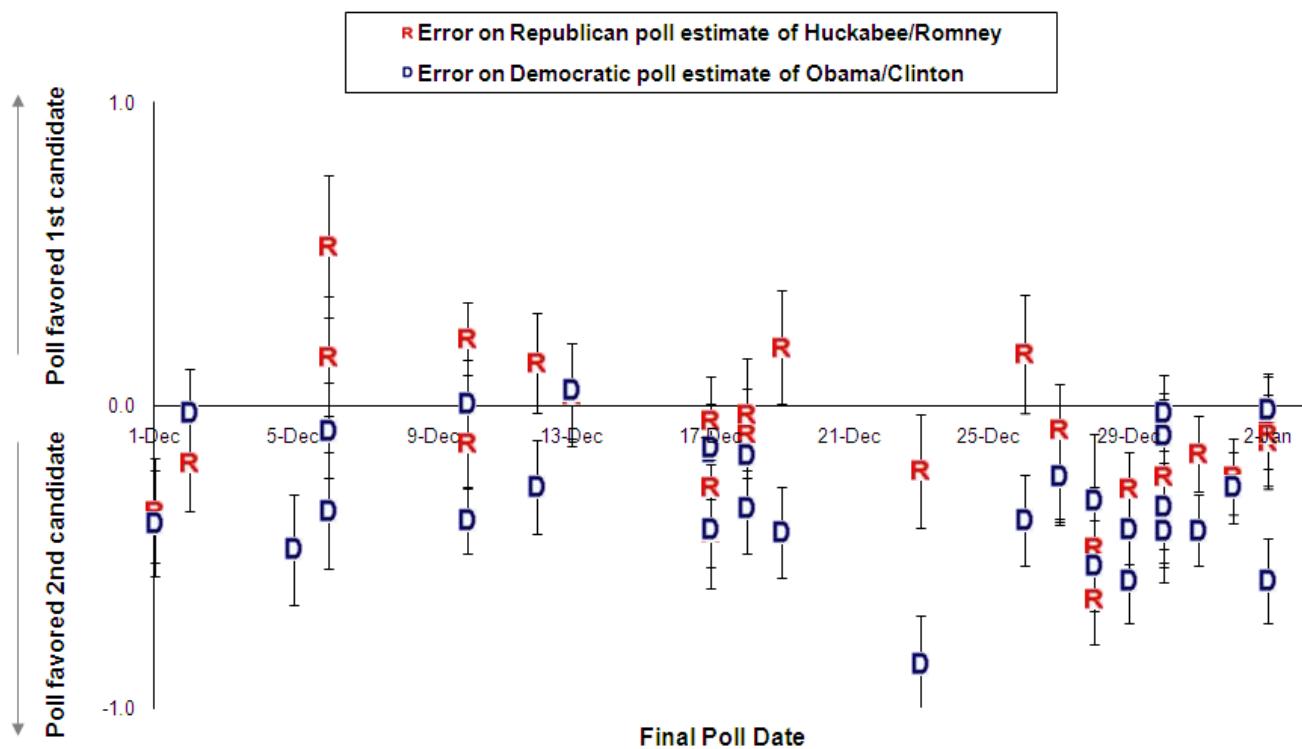
The committee began by developing a series of hypotheses or conjectures, based upon the members' own expertise as well as opinions contained in the news media or put forward by other survey professionals after the New Hampshire primary. (The list of hypotheses is contained in Appendix D.) A useful way to think initially about the estimation issues is through a visual examination of the accuracy of the polls, as measured by A , for the period of approximately *one month* preceding each election date; in later analysis, we will focus on the polls conducted in the two-week period preceding each event.

One would not necessarily expect the estimates four weeks in advance of an election with many candidates to be accurate, but this time period gives a perspective on a number of issues such as whether the estimates became more accurate as Election Day approached; whether the A values of different polls were more or less randomly distributed around a value of 0, indicating more or less correct estimation without any aggregate bias; and whether there were differences in the accuracy of the estimates produced in the Democratic and Republican contests. These graphical summaries are presented in Figures 1 through 5. By proceeding in the chronological order of the events, it is also possible to see how or whether the accuracy of the estimates might have changed over time, especially as the number of candidates in each field declined.

¹¹ The five states for which some data are presented in this report differ with respect to the rules governing who can vote in their primaries. Generally speaking, the more inclusive the primary, the more difficult the outcome is to predict. Both major parties held caucuses in Iowa, and voters were allowed to register at the caucus site. New Hampshire had semi-open primaries with Election Day registration. South Carolina had open primaries for both parties. California had semi-open primaries with no Election Day registration permitted. Wisconsin had open primaries for both parties, and Election Day registration was permitted.

Starting with Iowa as a foreground for the committee’s central focus and to provide a context for investigating accuracy (Figure 1), we employ a red “R” to indicate each estimate from a single poll for the Republican caucus and a blue “D” to indicate each estimate for the Democratic caucus. The range of the values of A for the Democratic contests was more variable and generally tended to overestimate support for Hillary Clinton relative to Barack Obama in comparison to the actual outcome. In the final week before the caucus, the polls moved toward more accurate estimation, with the exception of an estimate by the American Research Group that showed Clinton ahead by 9 percentage points. In the Republican contest, the estimates from the polls showed the same movement toward correct estimation, but all of the polls underestimated Mike Huckabee’s margin over Mitt Romney. In both contests, the polls underestimated the support for the second place candidate relative to the winner, although this error generally was reduced as caucus day approached.

Figure 1. 2008 Iowa Caucus Polls' Deviations from the Election Results

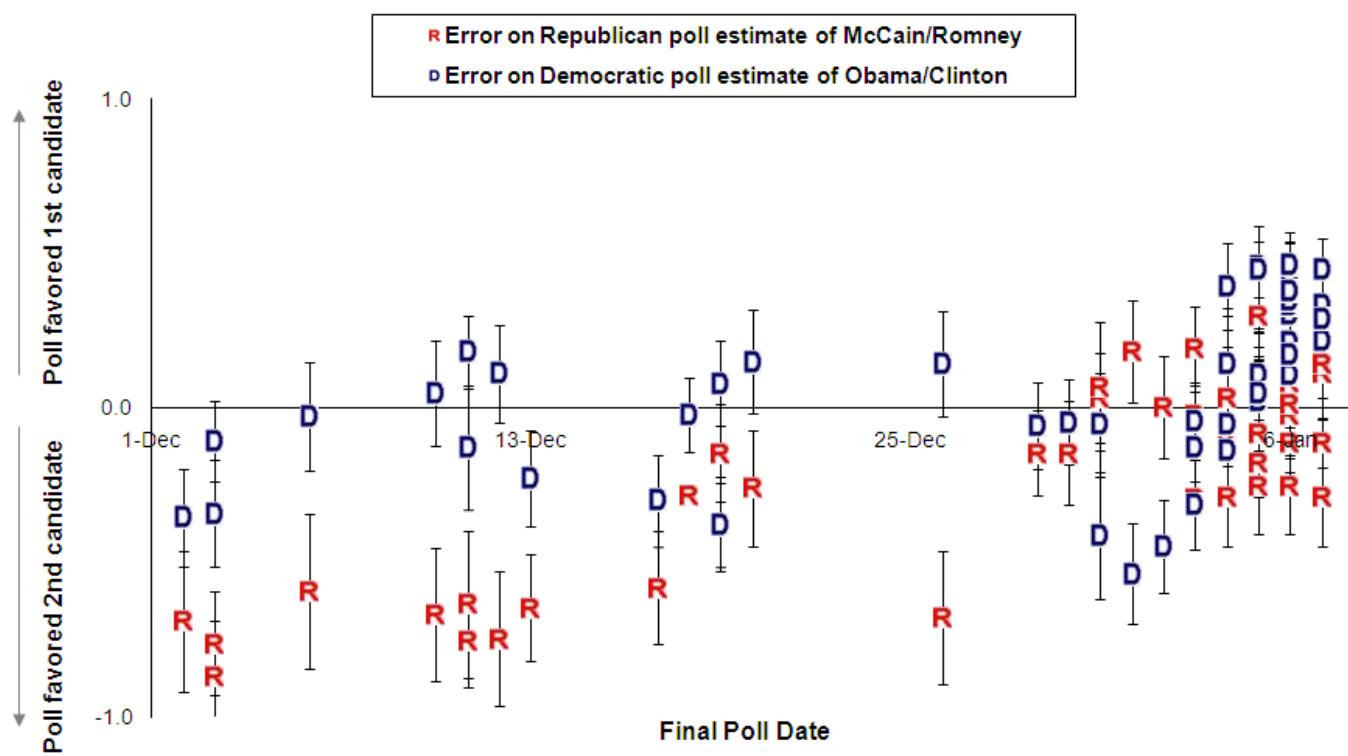


¹ Deviations from election results are calculated using A , which summarizes the degree to which a poll overstated or understated election day support for the 1st candidate relative to the 2nd candidate. Polls that predicted the exact election result (in terms of relative support for the top two candidates) have an A value of 0.0.

² The Republican and Democratic primaries were held January 3rd, 2008.

Data are presented in Figure 2 for the equivalent time series of poll estimates in New Hampshire, occurring only five days later. In this case, the problem seems relatively clear: 18 out of 21 polls that went into the field after the Iowa caucus showed Obama in the lead (with two estimates within the margin of error); one showed Clinton in the lead, early on, although the difference was within the margin of error. The figure shows that relative support for Obama increased after the Iowa caucus, but there was no associated shift in support for McCain or Romney, as some of the values of A in the last few days – generally smaller than the corresponding values in the Democratic estimates – were positive and others negative. These patterns suggest a systematic shift in the estimation of the outcome of the Democratic primaries and something more akin to random differences in the Republican estimates.

Figure 2. 2008 New Hampshire Primary Polls' Deviations from the Election Results

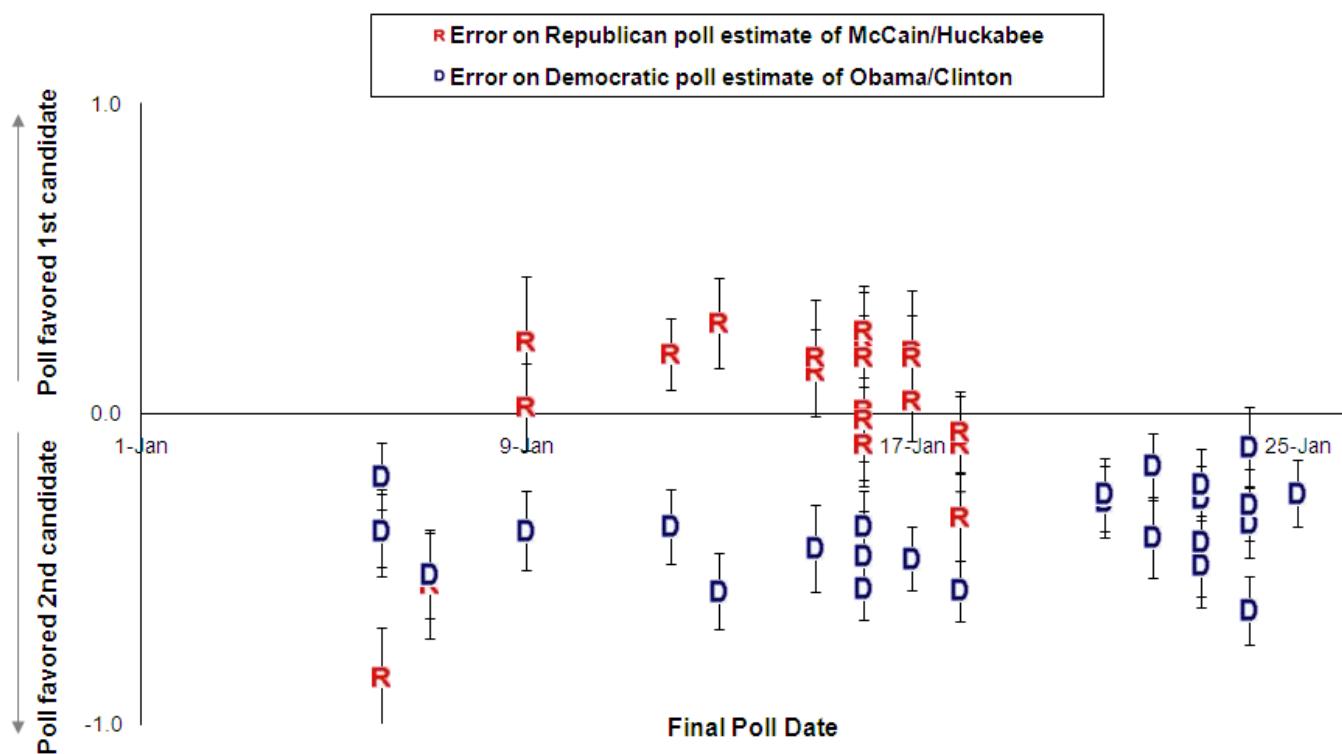


¹ Deviations from election results are calculated using A , which summarizes the degree to which a poll overstated or understated election day support for the 1st candidate relative to the 2nd candidate. Polls that predicted the exact election result (in terms of relative support for the top two candidates) have an A value of 0.0.

² The Republican and Democratic primaries were held January 8th, 2008.

Analogous data are presented in Figure 3 for the estimates produced before the South Carolina primaries. The Republicans and Democrats in this state actually held their primaries one week apart, with the Republicans voting on January 19th and the Democrats on January 26th. Fewer pre-election polls were conducted in South Carolina than in Iowa or New Hampshire. McCain won the primary by a narrow three-percentage-point victory over Huckabee, and generally the pre-election polls indicated a tight Republican contest, although one suggested Huckabee would win. Obama won a decisive victory over Clinton by 28 percentage points, and, while all of the polls indicated he was in the lead, they consistently produced underestimates of his eventual margin. And these estimates did not improve as Election Day approached.

Figure 3. 2008 South Carolina Primary Polls' Deviations from the Election Results

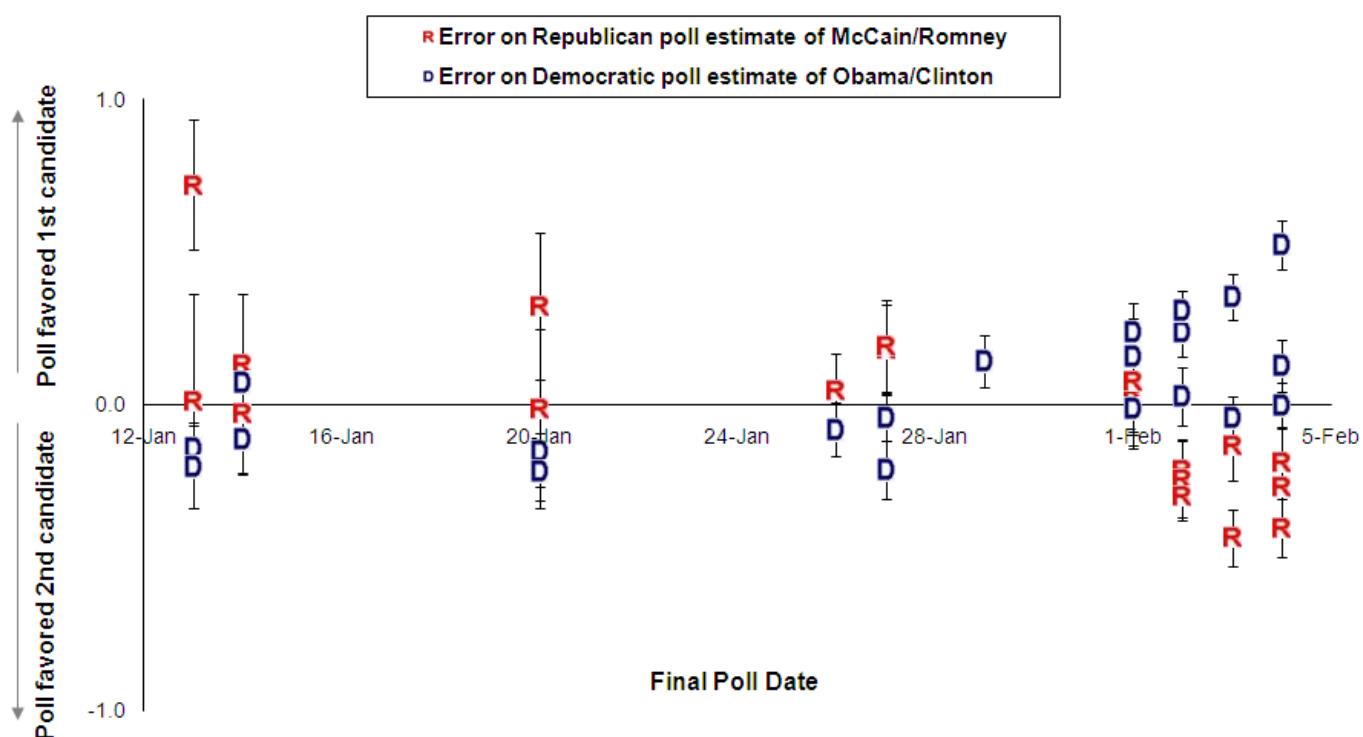


¹ Deviations from election results are calculated using A , which summarizes the degree to which a poll overstated or understated election day support for the 1st candidate relative to the 2nd candidate. Polls that predicted the exact election result (in terms of relative support for the top two candidates) have an A value of 0.0.

² The Republican and Democratic primaries were held January 19th and 26th, respectively.

Data are presented in Figure 4 that show the distribution of the pre-primary estimates for the California contests. Focusing on the final estimates produced in the week leading up to the primaries, two different patterns emerge. For the Democrats, Clinton won by a 9-point margin; but several of the polls overestimated her advantage. On the Republican side, the pre-primary estimates favored Romney – absolutely and relative to McCain, although McCain won the primary by 7 percentage points. The time series of estimates suggests that overall the polls in California became less accurate as Election Day approached.

Figure 4. 2008 California Primary Polls' Deviation from Election Results

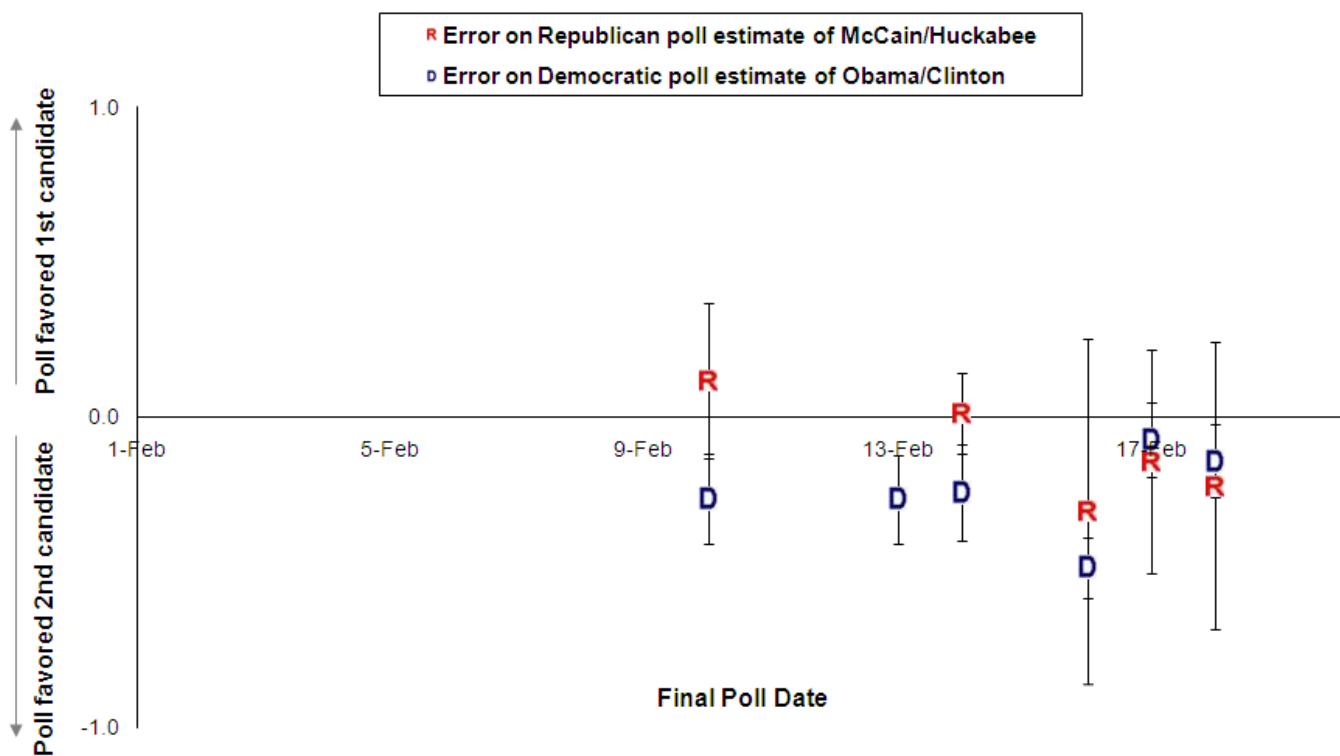


¹ Deviations from election results are calculated using A , which summarizes the degree to which a poll overstated or understated election day support for the 1st candidate relative to the 2nd candidate. Polls that predicted the exact election result (in terms of relative support for the top two candidates) have an A value of 0.0.

² The Republican and Democratic primaries were held February 5th, 2008.

In the case of Wisconsin, with its primaries held two weeks after Super Tuesday on February 19, Obama was an easy winner by 17 percentage points. Fewer polls were conducted before the Wisconsin primary than other primaries, and they all underestimated support for Obama. One factor contributing to this error may have been the Wisconsin rule permitting Election Day registration, which can make it harder for pollsters to anticipate the turnout rate accurately, overall and among key subgroups. The data presented in Figure 5 show that the estimates got closer to the eventual outcome as Election Day approached. On the Republican side, McCain defeated Huckabee by eight percentage points, and the final polls projected this within their margin of error.

Figure 5. 2008 Wisconsin Primary Polls' Deviation from Election Results



¹ Deviations from election results are calculated using A , which summarizes the degree to which a poll overstated or understated election day support for the 1st candidate relative to the 2nd candidate. Polls that predicted the exact election result (in terms of relative support for the top two candidates) have an A value of 0.0.

² The Republican and Democratic primaries were held February 19th, 2008.

Summary: An examination of the time series of estimates of caucus and primary outcomes in five states does not show a consistent pattern of accuracy in estimation. In some cases, estimation improved as Election Day approached, but in other cases it did not. In some cases, the inaccuracy of estimates seemed random, but in others there was an indication of systematic bias favoring one candidate or another. In general, the average estimates of the outcomes in the Republican races were more accurate than those in the Democratic races.

Explanations for Differences in the Accuracy of Pre-Primary Polls

In this section, we present the results of analyses related to possible explanations for estimation errors. The results are presented in line with the general sequence or chronology of decisions made in the design and implementation of any survey. That is, we begin with an examination of differences in the mode of data collection, then turn to sampling issues, wrap up with an analysis of the impact of weighting procedures and likely voter models on the final preference distributions, and end with a consideration of other external factors. The analysis will shift between comparisons of states, survey organizations, and individual survey estimates, and such changes will be made clear as they occur.

The use of this structure and the nature of the available data for analysis imply that we looked at one cause at a time for problems with accuracy. However, this should not be construed as an expectation that there was a single explanation for the problems. Any single moncausal explanation is likely to be both inaccurate and misleading. Pre-election polling is a complex process, and explanations for difficulties could derive from multiple factors such as statistical sampling theory or voter decision making or the psychology of interviewer-respondent interactions. Moreover a variety of small effects can accumulate to produce significant inaccuracies. Untangling these relationships in an ex post facto analysis such as this is virtually impossible given differences in the amount of information available for each poll estimate.

Mode of Data Collection

An initial decision in designing any survey is selecting a mode of data collection. In the case of pre-election polling, firms organize their work around a particular mode of data collection, and sponsors or clients usually end up with a particular mode as a result of selecting a firm to do the work, typically on the basis of cost. In the case of the pre-primary and caucus polls we analyzed, only two modes of data collection were used: 1) telephone interviews using a human interviewer in combination

with a computer-assisted telephone interviewing (CATI) system, and 2) telephone interviews using an interactive voice recognition (IVR) system in which digitally recorded questions were answered using a touch-tone phone.

Data are presented in Table 6 that show the average absolute value of A , the accuracy measure, for final trial heat estimates made in the final *two* weeks leading up to the primary elections in five states. The averages are computed separately for CATI polls and IVR polls because we wanted to assess whether the level of accuracy differed by mode. However, it is important to note that these polls differed on factors besides mode (e.g., field dates), which does not make this comparison straightforward.

In every state, more surveys were conducted via CATI than via IVR systems, although the differences were relatively small in South Carolina and Wisconsin compared to Iowa, New Hampshire, and California, which may be attributable to the larger number of polls conducted in the latter three states. The number of IVR polls conducted per state ranged from 1 to 3, while the number of CATI polls ranged from 3 to 13. Where only one IVR poll was conducted, the comparison might be thought of as an evaluation of a “house effect” from a single firm as much as a mode comparison.

Table 6. Accuracy of Final 2008 Primary Pre-election Polls by Mode of Data Collection^{A,B}

Number of polls in final two weeks		Mean absolute value of A	Number of polls in final two weeks		Mean absolute value of A
Iowa Democratic Caucuses			Iowa Republican Caucuses		
CATI	12	0.25	CATI	11	0.23
IVR	1	0.41	IVR	1	0.27
New Hampshire Democratic Primary			New Hampshire Republican Primary		
CATI	13	0.26	CATI	12	0.15
IVR	1	0.28	IVR	1	0.11
South Carolina Democratic Primary			South Carolina Republican Primary		
CATI	5	0.43	CATI	5	0.16
IVR	3	0.25	IVR	3	0.13
California Democratic Primary			California Republican Primary		
CATI	8	0.20	CATI	7	0.16
IVR	3	0.13	IVR	2	0.17
Wisconsin Democratic Primary			Wisconsin Republican Primary		
CATI	3	0.21	CATI	3	0.19
IVR	2	0.17	IVR	1	0.08

^A Only the final estimates from each poll are included in this analysis. No poll is included more than once, but the set of polls considered is larger than that listed in Table 5 because all polls during the final two weeks are included here.

^B Small sample sizes, particularly with respect to the small number of IVR polls, severely limit attempts to isolate an effect from mode on accuracy.

Table 6 presents data for 10 comparisons encompassing state and party. In half of the comparisons, only one IVR poll was conducted, so its final estimate is compared to an average value for multiple CATI polls. In these five comparisons, the average value of A was lower for the CATI surveys in three cases (both Iowa caucuses and the New Hampshire Democratic primary) and higher in two (the New Hampshire and Wisconsin Republican primaries). In the other 5 comparisons, where there were 2 or 3 IVR firms conducting polls, the average value of A was lower for the IVR estimates in 4 cases (South Carolina Democratic and Republican primaries, and the California and Wisconsin Democratic primaries). In one case, the California Republican primary, the average IVR A value was slightly higher than the average CATI A value.

With only 4 IVR firms producing 18 estimates in these 10 contests, it is possible to look at the average value of A for the three firms that produced multiple estimates. The lowest average absolute value of A was produced by Public Policy Polling (.147), based upon four estimates in both parties' primaries in South Carolina and Wisconsin. This was very closely followed by SurveyUSA (.147), based upon four estimates in both parties' primaries in South Carolina and California.¹² Rasmussen Reports produced nine estimates in total with an average A value of .233.

Summary: All of the final pre-primary polls were conducted by telephone, using either CATI or IVR systems. We found no evidence that one approach consistently out-performed the other – that is, the polls using CATI or IVR were about equally accurate. We caution that all of the comparisons are based on very small sample sizes and are potentially confounded with other factors that can contribute to accuracy.

Sample Frames and Respondent Selection

The issues of sample frames, respondent selection, estimating the likelihood of voting, and weighting are inextricably linked in pre-election surveys. Separate sections in this report discuss estimating likelihood and weighting, and each of these issues receives some treatment in this section as well. Pre-election pollsters are interested in obtaining information from people who will vote in the election (caucus, primary, or general), but the respondent may not even know at the time of interview whether he or she will participate. Obviously, under these circumstances there is no way to identify the population of interest in advance, that is, to obtain or construct a sample frame consisting of a list of

¹² The difference in the relative value of the A for these two firms is due to rounding.

voters prior to the event. In this way, the use of a particular frame and the definition of likelihood of voting are linked.

Table 7 indicates the kind of frame used to select a sample for the pre-primary polls in the four states on which we focus, as well as the method for determining likelihood of voting. Some firms employed a definition of likelihood to vote at the same time that they drew their sample; for others, they drew a sample from a frame and then asked questions in the survey to determine likelihood of voting. In the case of one firm, the interviewer asked to speak to “a registered voter in the household.”

Another distinction is that firms drawing samples of phone numbers differed in whether or not cell phone numbers were included and how information about whether the number was listed or not was used. They then selected a respondent by a variety of methods (see Table 5) after contact was made with the household or telephone number.¹³ In these cases, a series of questions at the beginning of the interview were used to determine likelihood of voting, and the interview only proceeded with a likely voter according to that firm’s definition. In other cases, the interview started with some questions that everyone answered, then likelihood was determined, and then the trial heat question(s) about candidate preference were asked only of those identified or defined as likely voters. In still other cases, firms purchased a sample of registered voters that presumably included telephone numbers and sometimes included information about past voting behavior, and they attempted to contact individuals selected on the basis of their registration status and/or past voting behavior for an interview. These lists were purchased from commercial firms or, in one case, supplied by a political party. Another issue, discussed in greater detail in the next section, has to do with the inclusion or exclusion of cell phone numbers. Very few of the polling firms included a special sample of cell phone numbers in their primary polling.

Summary: The specific impacts of the use of particular sampling frames and methods of respondent selection on the accuracy of estimations remain difficult to assess. This is because of the joint relationship between the use of certain frames, respondent selection after contact, the determination of voting likelihood, and post-survey weighting procedures, as discussed in greater detail below.

¹³ It is unclear how the firms using IVR techniques that purchased a list that included a subset of identified registered voters insured that they were speaking to the correct respondent.

Table 7. Sampling Frame and Respondent Selection

Sampling Frame		Method to Identify Likely Voters
New Hampshire Polls		
ARG	Landline RDD	Likelihood determined by questions
CBS News	Landline RDD callback	Likelihood measured through questions and then a regression technique to construct weights.
Gallup	Landline RDD + Cell RDD	Likelihood determined through questions used to form an index
Marist College	Landline RDD	Likelihood determined by questions
Opinion Dynamics	Landline RDD	Likelihood determined by questions
Rasmussen Reports	Landline RDD	Likelihood determined by questions
Research 2000	Landline RDD	Screened for likelihood (Registered, habitual voter, definitely/ probably vote in the primary)
RKM	RV list	Likelihood determined by questions
Strategic Vision	No information available	No information available
Suffolk University	Landline RDD	"Wide screens" to get first time voters
Univ. of NH	Landline RDD	Likelihood determined by questions
Zogby	Directory of listed landlines	Likelihood determined by questions
South Carolina Polls		
ARG	Landline RDD	Likelihood determined by questions
Clemson University	No information available	Likelihood based on screening for past primary voting and planning to vote in this one
Ebony/Jet	Voter file from SC Dem. Party	Likelihood determined by questions
Mason-Dixon	Landline RD	Likelihood determined by questions
Public Policy Polling	RV list	Likelihood determined by frame variables on RV list and survey questions. They ask people to hang up if they do not intend to vote in the primary that they are calling about.
Rasmussen Reports	Landline RDD	Likelihood determined by questions
SurveyUSA	Landline RDD	Likelihood determined by questions
Zogby	Directory of listed landlines	Likelihood determined by questions
Wisconsin Polls		
ARG	Landline RDD	Likelihood determined by questions
Public Policy Polling	RV list	Likelihood determined by frame variables on RV list and survey questions. They ask people to hang up if they do not intend to vote in the primary that they are calling about.
Research 2000	Landline RDD	Screened for likelihood (Registered, habitual voter, definitely/ probably vote in the primary)
Strategic Vision	No information available	No information available
California Polls		
Datamar	RV list	Likelihood determined using frame variables on the RV list
Field Poll	RV list	Likelihood determined by questions
LA Times	Landline RDD	Likelihood measured by questions ascertaining registered, intend to attend a primary, vote history (but 1 st time voters are included)
Mason-Dixon	Landline RDD	Likelihood determined by questions
Public Policy Institute	Landline RDD	Likelihood determined by questions
Rasmussen Reports	Landline RDD	Likelihood screening by questions
Suffolk University	RV list	Likelihood determined by questions
SurveyUSA	Landline RDD	Likelihood determined by questions
Zogby	Directory of listed landlines	Likelihood determined by questions

Non-coverage of Cell Phone Only (CPO) Voters

At the time of the New Hampshire primary, about 14.5% of American adults lived in a household with a cell phone but no landline, according to the National Health Interviews Survey (Blumberg and Luke 2008). We adopt the conventional short-hand of “cell phone only” (CPO) to describe this population. Cell phone only voters were excluded from all but 1 of the 13 New Hampshire primary polls studied in this report. Only the Gallup Organization included a sample of CPO adults in its poll.

During the primary season, it was widely believed that the omission of CPO individuals did not have a sizable impact on estimation of candidate preference. An analysis of exit poll data from the 2004 presidential election suggested that post-survey demographic adjustments, in particular an iterative-proportional fitting technique (raking) to population control totals for age, effectively eliminated coverage error in pre-election polls (Keeter 2006). In January 2008 a report from the Pew Research Center concluded that including cell phone interviews did not substantially change any key survey findings (Keeter 2008). And in May 2008 two other studies indicated that including a sample of cell phones had minimal effects on primary election trial heat estimates (Jones 2008; ZuWallack, Piehl, and Holland 2008).

By the time of the general election, however, views about the exclusion of this group had changed. The Pew Research Center estimated in September 2008 that this omission could underrepresent support for Obama by 2 percentage points in general election trial heats (Keeter, Dimock and Christian 2008). Similarly, Gary Langer (2008c), director of polling for ABC News, reported that Obama had a 6-point margin (50% to 44%) over McCain among likely voters when CPO interviews were included and a 4-point margin (49% to 45%) when they were excluded. Differences for full-sample estimates were generally quite small, however, leading Langer to conclude that the effect from adding cell phone only interviews was negligible.

Looking back on the New Hampshire pre-primary polls, exclusion of CPO individuals does not appear to have been an important factor in the estimation errors. Several teams of researchers independently concluded that CPO exclusion influenced estimates of support for Clinton versus Obama by at most 1 or 2 percentage points, and the direction was not always consistent. Pew had a statistically non-significant finding that Obama benefited from the exclusion of CPO individuals, while ORC Macro and CNN had a statistically non-significant finding that Clinton benefited. Furthermore,

Gallup, which included a cell phone only sample, performed no better in New Hampshire than the other polls in estimating the election outcome.¹⁴

Summary: Only one survey organization (Gallup) fielded a random-digit dialing (RDD) sample of cell phone cases. Other New Hampshire polls appear to have excluded CPO individuals. We found some evidence suggesting that this coverage gap influenced estimates in the general election, presumably as young people were motivated to turn out in support of Barack Obama, but we found no strong evidence suggesting that the gap influenced primary estimates in any meaningful way. In particular, non-coverage of CPO individuals in pre-primary polls does not appear to have been an important factor in the New Hampshire Democratic primary poll errors.

Nonresponse Bias and the Composition of Responding Samples

Another hypothesized cause of the error in the pre-election polls is nonresponse bias. That is, voters who could not be located or declined to participate in the surveys may have favored different candidates than those who did participate in the surveys. Demographic weighting that aligns the characteristics of the survey sample to characteristics of the entire voting population is generally thought to remove much of the bias from nonresponse. Weighting, however, is an imperfect technique and relies on assumptions that may not be valid in the context of these particular primary elections.

In an op-ed in the *New York Times* (2008) after the primary, Andrew Kohut, president of the Pew Research Center, posited that the pre-election polls for the New Hampshire Democratic primary missed the mark because of nonresponse bias. Kohut noted that poorer, less educated whites are less likely to participate in surveys than other voters and that these whites have more unfavorable views of blacks than respondents who participate in surveys. By this line of reasoning, the survey estimates would still be subject to nonresponse bias even after weighting adjustments were made for the educational and racial composition of the sample.

¹⁴ The Gallup poll was in fact the least accurate in estimating the New Hampshire Democratic Primary winner. This raised the question as to whether including the CPO sample might actually have increased the error in Gallup's estimate rather than decreased it. The committee was unable to test this hypothesis directly as it requires the availability of two sets of weights – one for the full sample estimate and a separate one for a landline-only sample estimate. Lacking the latter weight, we compared the unweighted full sample estimate with the unweighted landline sample estimate. The results were identical (37% Obama, 36% Clinton), suggesting that the addition of the CPO sample data had minimal effects on the final weighted estimate. It is also the case that the difference in vote preference between the CPO respondents (41% Obama versus 25% for Clinton) and the landline respondents (37% Obama, 36% Clinton) was probably not sizable enough to move the estimates by a meaningful amount given that the CPO cases represented 7.2% of the entire sample, and there was no post-survey adjustment for telephone service.

The difficulty in assessing this hypothesis and others related to nonresponse bias is that we generally know little if anything about those who do not participate in surveys. The 2008 primary polls were no different, and none of the organizations contacted by the committee provided call records for non-responding cases. Given this lack of information about how much effort was devoted to contacting original sample elements, we cannot rigorously test any hypotheses about how pre-election poll errors could be attributable to nonresponse bias. We were able, however, to conduct an indirect assessment of the levels of nonresponse bias in the polls by comparing characteristics of the respondents in the pre-election samples to those in the exit poll samples.

The best information available to us on this issue comes from the National Election Pool (NEP) exit polls funded by major news outlets and conducted by Edison Media Research and Mitofsky International. The exit polls provide demographic, attitudinal, and behavioral characteristics of those who voted in the primaries. The exit poll data are weighted to actual vote totals, and the resulting aggregate estimates are the best available information on the characteristics of the participating electorate.¹⁵ In the following analysis we make inferences about pre-election survey nonrespondents by comparing the personal demographic characteristics of the survey samples to the same characteristics of the voters as measured by the exit polls in the New Hampshire primary.

This analysis has two important limitations that should be kept in mind when considering the results. First, differences we observed between survey samples and the exit polls are not just due to nonresponse bias, but they are also due to non-coverage and possibly measurement differences. Error from non-coverage and error from nonresponse may not affect survey results in the same way. Unfortunately, we have no good way to disentangle these two error sources in this analysis.¹⁶ The second important point is that the exit polls have some problems of their own. People refuse to cooperate with the exit polls just as they refuse to cooperate with pre-election polls. The researchers who release the exit poll estimates do attempt to correct for this by aligning the exit poll results to the actual vote division. It is still possible, however, that residual nonresponse bias affects the accuracy of the exit poll figures. We use the exit poll data here because it is the best available information on the composition of the electorate.

¹⁵ There are a variety of different weighting schemes used in the pre-primary polls themselves, and they will be discussed in greater detail in a separate section that follows.

¹⁶ One possible way to separate nonresponse error from noncoverage error would be to compare the error properties of surveys using a landline RDD design from those using a landline RDD + cell sample design. We did not perform this analysis because there are so few surveys representing each design and these surveys differ in a number of respects besides sample design. Any attempt to isolate the effect from sample design would be confounded by other differences between surveys.

In Table 8, differences between the demographic distributions of four New Hampshire Democratic pre-election primary polls and the primary exit poll are reported. On several dimensions – gender, race, marital status, and party identification – the pre-election survey samples resembled the electorate as measured in the exit polls. On other dimensions the survey samples look a bit different, providing some possible explanations for the errors in the polls.

The reader will recall that pre-election polls in the New Hampshire Democratic primary understated support for Hillary Clinton. Table 8 shows that two of the surveys slightly under-represented households with at least one labor union member; a measure of union membership was not available for the other surveys. According to the exit poll, Clinton won union households by a 40% to 31% margin. This suggests that the pre-election polls may have missed some Clinton support by underestimating the size of the union vote. We see a similar result looking at education. In the exit poll, Clinton won voters with less than a college education by a 48% to 30% margin. Three of the pre-election polls underrepresented this group and thus appear to have missed some Clinton support. This could be consistent with the Kohut hypothesis. The CBS News poll, which was a re-interview survey,¹⁷ overrepresented respondents with a high school education or less, so their error would have come from some other sources.

One difference between the pre-election polls and the exit poll leads to a counter-intuitive result. Table 8 shows that all four pre-election polls in New Hampshire overrepresented registered Democrats. According to the exit poll, registered Democrats comprised 52% of the actual voters, but they comprised roughly 57 to 61% of the survey samples. Clinton won this group by a 43% to 32% margin. It is curious that the surveys over-represented this pro-Clinton group but still underestimated her support.

In sum, we find some evidence that the pre-election polls may have missed Clinton support by underestimating union households and non-college graduates and overstated Clinton support by overestimating registered Democrats. We reiterate, however, that the exit poll is itself an imperfect benchmark, so the conclusions are more suggestive than definitive.

In addition to the exit poll analysis, the committee sought to investigate nonresponse error by testing for a relationship between response rate and accuracy. Response rate is a poor indicator of

¹⁷ The CBS News poll re-interviewed New Hampshire registered voters first interviewed in November of 2007 to measure the amount of individual change.

nonresponse error (Curtin, Presser, and Singer 2000; Groves 2006; Keeter et al. 2000; Merkle and Edelman 2002), but survey organizations are required to disclose it under AAPOR's minimum disclosure standards. Our attempt to carry out this analysis was hindered by two factors. Critically, only 8 of the 13 survey organizations releasing estimates for the New Hampshire Democratic primary disclosed their response rate (see Table 5). Furthermore, the rates disclosed were not all calculated in the same fashion. One was reported as a range; three are based on AAPOR response rate (RR) calculation 1; one each is based on RR (2), RR (3), and RR (5); and one is based on an unknown calculation.¹⁸ In light of these limitations, we did not pursue a rigorous analysis of the relationship between response rate and accuracy. The correlation between A (absolute value) and response rate for all of the primary polls disclosed was not significant (Pearson $r = 0.317, p=.22$).

Summary: We found some evidence that nonresponse error contributed to the estimation errors in the New Hampshire Democratic primary polls, but we lack data to confirm this rigorously. According to the exit polls, the pre-election polls underestimated the size of two pro-Clinton groups (union households and those with less than a college education). On the other hand, the pre-election poll samples were similar to weighted exit poll results in terms of gender, race, marital status, and party identification.

¹⁸ Both the AAPOR definitions of disposition codes and response rate calculations, as well as an Excel spreadsheet for making response rate calculations can be found at: <http://aapor.org/resources>.

Table 8. Sample Composition of Exit Poll vs. Pre-election Polls in the NH Democratic Primary

N	Exit Poll 1,955 voters	CBS News		Univ. of NH		Opinion Dynamics		Gallup	
		323 LVs	diff.	599 LVs	diff.	500 LVs	diff.	778 LVs	diff.
Gender									
Male	43%	44%	1	41%	-2	47%	4	42%	-1
Female	57%	56%	-1	59%	2	53%	-4	58%	1
Age									
18-24	11%	7%	-4	7%	-4	6%	-5	12%	1
25-29	7%	15%	8	2%	-5	4%	-3	4%	-3
30-39	15%	14%	-1	12%	-3	8%	-7	12%	-3
40-49	23%	15%	-8	21%	-2	17%	-6	25%	2
50-64	31%	34%	3	36%	5	35%	4	34%	3
65 or Over	13%	15%	2	22%	9	30%	17	14%	1
Education									
H.S. Grad or less	20%	30%	10	18%	-2	19%	-1	18%	-2
Some College	27%	39%	12	21%	-6	22%	-5	26%	-2
College Graduate	30%	14%	-16	33%	3	35%	5	23%	-7
Postgraduate	23%	17%	-6	28%	5	24%	1	33%	10
Party Identification									
Democrat	54%	54%	0	54%	0	54%	0	51%	-4
Republican	3%	3%	0	4%	1	2%	-1	1%	-2
Independent	44%	43%	-1	42%	-2	44%	0	49%	5
Party Registration									
Indep/Unregistered	48%	38%	-10	40%	-8	36%	-12	43%	-5
Regist. Democrat	52%	61%	9	60%	8	61%	9	57%	5
Regist. Republican		1%	1			3%	3	0%	0
Vote by Income*									
Under \$15,000	5%	5%	0			5%	0		
\$15-30,000	9%	14%	5			11%	2	11%	-3
\$30-50,000	18%	26%	8			20%	2	16%	-2
\$50-75,000	24%	23%	-1			24%	0	21%	-3
\$75-100,000	16%	16%	0			17%	1	19%	3
\$100,000+	28%	16%	-12			23%	-5	32%	4
Union household									
Yes	20%	16%	-4			15%	-5		
No	80%	84%	4			85%	5		
Ideology									
Liberal	56%	40%	-16					44%	-12
Moderate	36%	52%	16					46%	10
Conservative	8%	8%	0					10%	2
First Time Primary Voter									
Yes	19%	14%	-5					9%	-11
No	81%	86%	5					92%	11
Married									
Yes	61%	60%	-1					67%	6
No	39%	40%	1					34%	-6
Religious Service Attendance									
More than weekly	4%			4%	0	8%	4		
Weekly	14%			20%	6	22%	8		
Monthly	10%			13%	3	15%	5		
Few times a year	33%			28%	-5	24%	-9		
Never	37%			35%	-2	31%	-6		
Race									
White	95%	98%	3						
African-American	1%	1%	0						
Asian	1%	0%	-1						
Other	2%	1%	-1						

Figures reflect the poll sample composition after weighting.

*Gallup used different response options, so categories were collapsed in order to make equivalent comparisons.

Trial Heat Question Wording

In the pre-primary surveys, the respondents are asked a “trial heat” question about their preferred candidate in the election. This question can be asked in several forms, both in terms of the stem question wording and the response categories offered. We had information on 13 different question wordings used in the Democratic primary and 11 different questions used in the Republican primary in New Hampshire, assembled from information that the polling organizations provided or that could be located from other sources. We also had information on four different questions used in the South Carolina Democratic and Republican primaries, seven questions used in the California Democratic and Republican primaries, four questions used in the Wisconsin Democratic primary, and three questions used in the Wisconsin Republican primary.¹⁹ This information is presented in Table 9.

One of the ways that such questions might differ is in whether the names of some or all of the candidates are mentioned, providing a form of recognition of the candidates for those who may not have been paying much attention to the campaign. In New Hampshire, the survey conducted by RKM did not have the interviewer read any of the names of the candidates, while the survey conducted by Suffolk University told the respondent that there were 22 names on the ballot but offered the names of only the “eight major candidates” listed alphabetically.

Another difference is whether the order of the names in the trial heat question corresponds to the order of the names on the ballot. For example, the New Hampshire ballot arranges candidates’ names alphabetically starting with a randomly selected letter of the alphabet. (See the section on *Ballot Order Effects* below.) However, only in the Zogby and Suffolk University trial heat questions for New Hampshire’s Democratic primary were the candidates’ names arranged in the same order that they appeared on the ballot. In South Carolina, the candidates were listed alphabetically on the ballot, and only one of the polls presented the names in that order to all respondents. In California, ballot order was randomized and rotated across districts. In the Wisconsin Democratic primary, the order of the candidates was assigned at random, and none of the questions offered the names in that order.

Still another difference is whether the CATI and IVR pre-primary surveys randomly rotated the names of the candidates in the heat trial question – a method for counteracting recency effects (Krosnick and Alwin 1987; Holbrook, Krosnick, Moore, and Tourangeau, 2007). A recency effect is a

¹⁹ The Field Poll had slightly different question sequences for the Democratic and Republican pre-primary polls. Early in their polling on these races, when the field of candidates in each contest was large, respondents were asked which of the candidates they could vote for and then which was their first choice. By the time the primaries approached, the Democratic field had narrowed, and only one question about a choice among three candidates was asked. The Republican field was still relatively large, so the earlier sequence was used to preserve an ability to comment on the trend in support based upon the first choices.

cognitive bias that makes respondents more likely to select the option they heard last. As shown in Table 9, most of the New Hampshire Democratic primary polls randomized the order of candidate names across interviews, but four polls did not. The trial heat question from the Zogby, Suffolk University and *LA Times* polls presented Obama's name after Clinton's, which may have made Obama somewhat more salient in respondents' minds as they formulated their answer. These polls had Obama leading by 13, 5 and 2 percentage points, respectively. The fourth poll using a uniform name order was conducted by Research 2000, which presented Clinton's name after Obama's. If a recency effect occurred, we might expect this poll to be more accurate than the others because a possible recency effect would break in favor of Clinton and offset the general trend of Clinton support being underestimated in the New Hampshire polls. This was the case, as Research 2000 had Obama leading by just 1 percentage point (a statistical dead heat).

Unfortunately, this analysis is suggestive but not definitive; we do not have an experimental manipulation of name order to assess its impact. The results regarding response order and estimation error are in the expected direction, but there are far too few data points to establish any causal relationship. We find these results intriguing, but much more data are required to address this issue properly.

Table 9. Trial Heat Question Wording Information Available from Pollsters and Other Sources

Poll	Question Wording	Rotate names?	One-step force?*
New Hampshire Democratic Primary			
ARG	If the 2008 Democratic presidential primary were being held today between (candidates), for whom would you vote?	Yes	Yes
CBS News	If the New Hampshire Democratic presidential primary was held today, which candidate would you vote for? (candidates)	Yes	Yes
Gallup	Suppose the Democratic primary for president were being held today. If you had to choose among the following candidates, which candidate would you vote for? (candidates) As of today, to which Democratic candidate do you lean most?	Yes	No
LA Times	If the January Democratic primary for president were being held today and the candidates were New York Senator Hillary Clinton, Connecticut Senator Chris Dodd, former North Carolina Senator John Edwards, Delaware Senator Joe Biden, New Mexico Governor Bill Richardson, Illinois Senator Barack Obama, Ohio Representative Dennis Kucinich and former Alaska Senator Mike Gravel, for whom would you vote: Clinton, Dodd, Edwards, Biden, Richardson, Obama, Kucinich or Gravel?	No	No
Marist College	If the New Hampshire Democratic Presidential Primary were held today, whom would you support if the candidates are Hillary Clinton, John Edwards, Bill Richardson, Dennis Kucinich, Barack Obama? [IF NEEDED: If you had to decide today...]	Yes	No
Mason-Dixon	If the 2008 New Hampshire Democratic primary were held today, which one of the following candidates would get your vote? (candidates)	Yes	Yes
Opinion Dynamics	If the New Hampshire Democratic presidential primary were held today, would you vote for? (candidates) [IF DON'T KNOW: As of today, which of these candidates are you leaning towards?]	Yes	No

Research 2000	If the 2008 Democratic primary for president were held today, which of the following candidates would you vote for? John Edwards, Barack Obama, Hillary Clinton, Bill Richardson, Dennis Kucinich, or Mike Gravel?	No	Yes
RKM	If the Democratic presidential primary election were held today, who would you vote for? If you were forced to choose right now, which candidate would you be most likely to vote for?	names not read	No
Strategic Vision	If the 2008 Democratic presidential primary were held today between (candidates), for whom would you vote?	Yes	Yes
Suffolk University	There are 22 Democratic candidates for president listed on your New Hampshire primary ballot. Of these, the major eight candidates listed alphabetically on your ballot are (candidates). For whom will you vote or toward whom would you LEAN at this time? Joe Biden, Hillary Clinton, Christopher Dodd, John Edwards, Mike Gravel, Dennis, Kucinich, Barack Obama, or Bill Richardson?	No	No
University of NH	I'm going to read you the names of the candidates who are running for the Democratic nomination. If the Democratic primary for president were held today, which of the following would you support for the Democratic nomination ... Hillary Clinton ... John Edwards ... Mike Gravel ... Dennis Kucinich ... Barack Obama ... Bill Richardson ... or someone else?" ROTATE RESPONSES ²⁰	Yes	Yes
Zogby	If the Democratic primary for president were held today, for whom would you vote? Joe Biden, Hillary Clinton, Christopher Dodd, John Edwards, Michael Gravel, Dennis Kucinich, Barack Obama, or Bill Richardson?	No	No

New Hampshire Republican Primary

CBS News	If the New Hampshire Republican presidential primary was held today, which candidate would you vote for? (candidates)	Yes	Yes
Gallup	Suppose the Republican primary for president were being held today. If you had to choose among the following candidates, which candidate would you vote for? (candidates) As of today, to which Republican candidate do you lean most?	Yes	No
LA Times	If the Republican primary for president were being held in your state today and the candidates were former New York City Mayor Rudy Giuliani, former Arkansas Governor Mike Huckabee, California Congressman Duncan Hunter, Arizona Senator John McCain, former Massachusetts Governor Mitt Romney, former Tennessee Senator Fred Thompson, or Texas Representative Ron Paul for whom would you vote: Giuliani, Huckabee, Hunter, McCain, Romney, Thompson or Paul?	No	No
Marist College	If the New Hampshire Republican presidential primary were held today, whom would you support if the candidates are Rudy Giuliani, Mike Huckabee, Duncan Hunter, John McCain, Ron Paul, Mitt Romney, Fred Thompson? [IF NEEDED: If you had to decide today...]	Yes	No
Opinion Dynamics	If the New Hampshire Republican presidential primary were held today, would you vote for? (candidates) [IF DON'T KNOW: As of today, which of these candidates are you leaning towards?]	Yes	No
Research 2000	If the Republican Primary for President were held today, which of the following candidates would you vote for? Rudy Giuliani, Ron Paul, John McCain, Mitt Romney, Mike Huckabee, Fred Thompson, or Duncan Hunter?	No	Yes
RKM	If the Republican presidential primary election were held today, who would you vote for? If you were forced to choose right now, which candidate would you be most likely to vote for?	names not read	No
Strategic Vision	If the 2008 Republican presidential primary were held today between (candidates), for whom would you vote?	Yes	Yes

²⁰ This wording differs slightly from the one that appears in the final University of New Hampshire press release.

Suffolk University	There are 21 Republican candidates for president listed on your New Hampshire primary ballot. Of these, the major eight candidates listed alphabetically on your ballot are (candidates). For whom will you vote or toward whom would you LEAN at this time? Rudy Giuliani, Mike Huckabee, Duncan Hunter, Alan Keyes, John McCain, Ron Paul, Mitt Romney, or Fred Thompson?	No	No
University of NH	I'm going to read you the names of the candidates who are running for the Republican nomination. If the Republican primary for president were held today, which of the following would you support for the Republican nomination? ... Rudy Giuliani ... Mike Huckabee ... Duncan Hunter ... John McCain ... Ron Paul ... Mitt Romney ... Fred Thompson ... or someone else?	Yes	Yes
Zogby	If the Republican primary for president were held today, for whom would you vote? Rudy Giuliani, Mike Huckabee, Duncan Hunter, John McCain, Ron Paul, Mitt Romney, Tom Tancredo, or Fred Thompson?	No	No

South Carolina Democratic Primary

Mason-Dixon	If the 2008 South Carolina Democratic primary were held today, which one of the following candidates would get your vote? (candidates)	Yes	Yes
PPP	The Democratic candidates still running for President are Hillary Clinton, John Edwards, and Barack Obama. If the primary was today, who would you vote for? If you would vote for Hillary Clinton, press 1. If for John Edwards, press 2. If for Barack Obama, press 3. If you're undecided, press 4.	No	Yes
Survey USA	If the Democratic Primary for President of the United States were today, would you vote for? (candidates)	Yes	Yes
Zogby	If the Democratic primary for president were held today, for whom would you vote? Hillary Clinton, Barack Obama, John Edwards, or Mike Gravel?	No	No

South Carolina Republican Primary

Mason-Dixon	If the 2008 South Carolina Republican primary were held today, which one of the following candidates would get your vote? (candidates)	Yes	Yes
PPP	The Republican candidates still running for President are Rudy Giuliani, Mike Huckabee, John McCain, Ron Paul, Mitt Romney, and Fred Thompson. If the primary was today, who would you vote for? If you would vote for Rudy Giuliani, press 1. If for Mike Huckabee, press 2. If for John McCain, press 3. If for Ron Paul, press 4. If for Mitt Romney, press 5. If for Fred Thompson, press 6. If you're undecided, press 7.	No	Yes
Survey USA	If the Republican Primary for President of the United States were today, would you vote for? (candidates)	Yes	Yes
Zogby	If the Republican primary for President were held today, for whom would you vote? Rudy Giuliani, Mike Huckabee, Duncan Hunter, John McCain, Mitt Romney, or Fred Thompson?	No	No

California Democratic Primary

Datamar	If the primary election were held today, for whom would you vote? The candidates are Joe Biden, U.S. Senator from Delaware; Chris Dodd, U.S. Senator from Connecticut; Bill Richardson, Governor of New Mexico, Mike Gravel, former U.S. Senator from Alaska; Dennis Kucinich, Congressman from Ohio; John Edwards, former U.S. Senator from North Carolina; Barack Obama, U.S. Senator from Illinois; Hillary Clinton, U.S. Senator from New York.	No	Yes
Field Poll	If the Democratic presidential primary election were being held today and the choices were (candidates), which candidate would you vote for?	Yes	Yes
LA Times	If the presidential primary were held today, please tell me which candidate you would be most likely to vote for? (candidates)	Yes	Yes
PPIC	If the Democratic primary for president were being held today, and these were the candidates, who would you vote for? (candidates)	Yes	Yes

Suffolk	There are two active candidates for President on your ballot <Barack Obama and Hillary Clinton>. For whom will you vote, or toward whom would you lean at this time?	No	No
SurveyUSA	If the Democratic Primary for President of the United States were today, would you vote for? (candidates)	Yes	Yes
Zogby	If the Democratic primary for president were held today, for whom would you vote? Hillary Clinton, John Edwards, Mike Gravel, or Barack Obama?	No	No

California Republican Primary

Datamar	If the primary election were held today, for whom would you vote? The candidates are Mike Huckabee, former Governor of Arkansas; Duncan Hunter, California Congressman; Tom Tancredo, Colorado Congressman; Ron Paul, Texas Congressman; Mitt Romney, former Governor of Massachusetts; Fred Thompson, former U.S. Senator from Tennessee; John McCain, U.S. Senator from Arizona; Rudy Giuliani, former Mayor of New York City.	No	Yes
Field Poll	I am going to read back the names of the candidates you said you would have at least some chance of voting for. Of these persons, who would be your first choice if the California Republican primary election for President were being held today? (AS APPLICABLE: Rudy Giuliani, Mike Huckabee, John McCain, Ron Paul, Mitt Romney)	No	Yes
LA Times	If the presidential primary were held today, please tell me which candidate you would be most likely to vote for? (candidates)	Yes	Yes
PPIC	If the Republican primary for president were being held today, and these were the candidates, who would you vote for? (candidates)	Yes	Yes
Suffolk	There are four active candidates for President on your ballot (candidates). For whom will you vote or toward whom would you lean at this time?	No	No
SurveyUSA	If the Republican Primary for President of the United States were today, would you vote for? (candidates)	Yes	Yes
Zogby	If the Republican primary for president were held today, for whom would you vote? Rudy Giuliani, Mike Huckabee, John McCain, Ron Paul, or Mitt Romney?	No	No

Wisconsin Democratic Primary

ARG	If the 2008 Democratic presidential primary were being held today between (candidates), for whom would you vote?	Yes	Yes
PPP	The Democratic candidates still running for President are Hillary Clinton and Barack Obama. If the primary was today, who would you vote for? If you would vote for Hillary, press 1. If for Barack Obama, press 2. If you're undecided, press 3.	No	Yes
Research 2000	If the Democratic Primary for President were held today, which of the following candidates would you vote for: Barack Obama or Hillary Clinton?	No	Yes
Strategic Vision	Who is your choice for the Democratic Presidential nomination in 2008? Barack Obama or Hillary Clinton?	No	Yes

Wisconsin Republican Primary

PPP	The Republican candidates still running for President include Mike Huckabee, John McCain, and Ron Paul. If the primary was today, who would you vote for? If Mike Huckabee, press 1. If John McCain, press 2. If Ron Paul, press 3. If you're undecided, press 4.	No	Yes
Research 2000	If the Republican Primary for President were held today, which of the following candidates would you vote for: Ron Paul, John McCain, or Mike Huckabee?	No	Yes
Strategic Vision	Who is your choice for the Republican Presidential nomination in 2008? John McCain, Mike Huckabee, or Ron Paul?	No	Yes

*Surveys marked Yes for one-step force administered a single trial heat question with no follow-up for undecided respondents. Surveys marked No either probed the undecided for their best guess or asked a follow-up question about the candidate to whom they were leaning.

Still another difference in the trial heat questions used by polling firms is whether candidate preference is asked as a single question or in a two-step sequence in which those who respond as “Undecided” or “Don’t know” to the first question are asked in a follow-up question to name the candidate toward which they are leaning. Surveys that used the two-step question should ultimately produce a lower level of “Undecided” voters because the respondents were provided with two chances to give a candidate’s name. The data in Table 10 show that six of the New Hampshire survey firms used a single question, and seven used a two-question sequence. Information is provided in Table 10 on the level of “Undecideds” in each survey. Surprisingly, the *average* level of “Undecideds” was no higher for the polls using single-question format (6%) than for polls using the two-question format (7%). However, this can be attributed to the fact that two of the firms using the two-part trial heat question, Opinion Dynamics and the *Los Angeles Times*, had significantly higher levels of undecided responses than any other survey. Excluding them, the average level for the two-question sequence would have been 5.2%.

Table 10. Accuracy of NH Dem. Primary Polls by Use of Forced Choice

Poll	One-step force?	A	% Undecided
American Research Group	Yes	0.32	2%
CBS News	Yes	0.29	9%
Mason-Dixon	Yes	0.13	8%
Research 2000	Yes	0.10	3%
Strategic Vision	Yes	0.34	6%
Univ. of New Hampshire	Yes	0.33	6%
Mean		0.25	6%
Gallup	No	0.45	2%
LA Times	No	0.13	12%
Marist College	No	0.26	4%
Opinion Dynamics	No	0.20	12%
RKM	No	0.16	7%
Suffolk University	No	0.21	8%
Zogby	No	0.44	5%
Mean		0.26	7%

While the primary polls differed in their trial heat question wording, they were nearly uniform on another dimension. None of the poll estimates studied for this report featured an allocation of undecided voters in the published trial heat estimates.²¹ This means that the poll-based point estimates of candidate vote share are expected to be systematically lower than the actual election vote share because they will not add to 100%. *Relative* shares of the vote and the margin between the top two candidates, however, should be largely unaffected by whether or not undecided voters were allocated, as the level of undecided was low for each poll. The decision not to allocate probably had little if any impact on the accuracy of the estimates. To be sure, if all the New Hampshire Democratic primary polls had allocated the share of undecided voters entirely to Senator Clinton, they would have come closer to projecting the actual result; but this outcome obviously could not have been known in advance. Even under this hypothetical allocation, several of the polls still would have projected an Obama victory in New Hampshire. Furthermore, when allocation of undecided voters is performed, most often it is assumed that each of the top two or three candidates will receive some proportion of the undecided vote, rather than one candidate receiving all of it.

Summary: We found no compelling evidence to suggest that the wording of the trial heat questions contributed to the New Hampshire polling errors in any meaningful way. Most polls randomized the order of the candidates' names, but we were unable to evaluate independently whether this had any impact on respondents' expression of support for candidates. Levels of undecided voters were generally low in most polls, with some exceptions; and average levels of "undecided" responses were similar for surveys using either the one- or the two-question format to ascertain candidate preference. Polls that used a follow-up question for undecided respondents performed no differently in terms of accuracy than polls that did not.

²¹ The one exception is the University of New Hampshire poll. Their press release included estimates with no allocation (39% Obama, 30% Clinton) as well as estimates with undecided voters allocated. With allocation, the point estimates for the top two candidates each increased by two percentage points (41% Obama, 32% Clinton), and the difference between them remained the same.

Likely Voter Definitions

One problematical reality of pre-election polling is that not all persons interviewed for a survey will, in fact, vote. Survey researchers attempt to account for this in a number of ways. One is to ask questions such as “How often do you usually vote?” and “Do you know the location of your polling place?” – answers to which help pollsters predict which of their respondents are likely to vote. Another method is to draw samples from registered voter lists that have some form of voting history attached; a likely voter may be defined as one who has voted in a previous primary, either for president or for some other office. Models of likely voters are quite variable across survey organizations, and they are sometimes considered proprietary. Whatever the method of identifying a likely voter, those deemed unlikely to vote are typically excluded from estimates of candidate preference in the electorate or are assigned a relatively small weight when estimation is performed.

Developing an effective likely voter model is particularly difficult in primary contests where the electorate can change in important ways from one election year to another. Anticipating the profile of the voting electorate was especially challenging in 2008 given the significant increase in turnout relative to recent primary elections. In 2004 some 219,787 New Hampshire voters cast ballots in the Democratic primary, but in 2008 this figure increased 31%, to 287,527. In other states, the increase in turnout was even more dramatic. For example, nearly twice as many South Carolina voters attended the Democratic primary in 2008 as did in 2004. And almost two million more voters participated in the 2008 than in the 2004 California primary, which featured an earlier election date. The increase in turnout in 2008 for four early Democratic contests is shown in Table 11 and suggests that the use of historical likely voter models might not have worked as effectively in 2008 as it has in the past.

Table 11. Voter Turnout in 2004 and 2008 Democratic Primaries/Caucuses

	2004	2008	change
New Hampshire	219,787	287,527	+31%
South Carolina	292,422	532,468	+82%
California	3,107,629	5,066,993	+63%
Wisconsin	826,250	1,113,285	+35%

The New Hampshire exit poll is the only one in which any questionnaires included an item about whether the respondent was voting for the first time in a primary. Because analysis of the exit poll data shows that first-time voters were more likely to support Obama over Clinton (47% to 37%) than were those who had voted in previous primaries (33% to 38%), knowing the correct proportions of first-time and previous primary voters in the sample could affect estimation. And the estimated 19% of self-reported first-time voters in 2008 would not have been picked up in a likely voter model that was based on prior voting. This kind of complexity makes likely voter modeling a probable suspect when researchers seek to explain errors in pre-election polls.

Two distinct approaches are used to account for likelihood of voting in pre-election surveys (Traugott and Tucker, 1984), one involving the construction of a likelihood index and the other involving the calculation of likelihood weights for each respondent. In the first case, a series of variables are combined to form an index of likelihood of voting, with only cases that fall in an acceptable range, based on cutoff points on the index, included in the analysis. In the second approach, a weight is calculated for each respondent based upon their likelihood of voting, ranging from close to 0 for the least likely to approaching 1.0 for the most likely. All weighted cases are included in the candidate preference distribution. Several firms reported which questions they used to create their likely voter models, but few went so far as to report which of these two statistical approaches they used. When the nonresponse weights are combined with weights for the likelihood of voting, it is difficult if not impossible to assess the relative contribution of each factor to estimation. For the datasets provided to the committee, we can attempt to infer which approach was used based on the distribution of the weight factors used in the trial heat estimate. The topic of weighting is considered separately later in this report. At this point, we consider weights only to learn more about the likely voter models used.

Data in Table 12 summarize the distribution of the values of the weights assigned to respondents deemed likely voters in the datasets provided to the committee. The average weights across the surveys are very similar, approximating 1.00 with a range from .76 to 1.12, while the ratio of the weighted number of likely voters to the actual number of likely voters ranges from .76 (CBS News) to 1.12 (Field Poll). CBS News employs regression-based weighting for likelihood estimation that incorporates information from all of the cases in analysis, as well as weights that account for differential response; these weights range from 0.06 to 6.30. The Gallup Organization uses an index-based system to account for likelihood to vote, with weights ranging from 0.34 to 2.33.

Table 12. Sample Size and Weights for 2008 Democratic Primary Trial Heat Polls

Polling Organization	Unweighted Number of LVs	Weighted Number of LVs	Ratio of Weighted to Unweighted N	Average weight	Smallest weight	Largest weight
CBS News (NH Primary)	322	244	0.76	0.79	0.06	6.30
Gallup (NH Primary)	778	662	0.85	0.85	0.34	2.33
U.N.H. (NH Dem Primary)	595	594	1.00	1.00	0.21	2.37
Field Poll (CA Primary)	511	571	1.12	1.12	0.36	3.77
Op. Dynamics (NH Primary)	500	500	1.00	1.00	0.55	1.30
PPIC (CA Primary)	644	543	0.84	0.84	0.40	7.44
SurveyUSA (CA Primary)	1,132	872	0.77	0.77	0.16	6.11

The average, minimum, and maximum weights reported here are based only on respondents identified in each survey as likely voters.

Some polling organizations used unusual terminology, or common terminology in unusual ways, to describe their approaches to sample weighting. Some IVR pollsters, for example, may have less control over the details of their interviewing in the sense that they cannot simply “turn off” their computerized data collection when they complete the exact number of interviews for which they have a contract. So they can sometimes end up with more cases than they use. Furthermore, these additional cases appear, as would be expected, among those who are most likely to be at home when the phone rings, such as women, older people, or whites. One IVR firm, SurveyUSA, reweights its samples to bring such groups (which they describe as “oversamples”) back into their appropriate proportions in the population²²; however another IVR firm, Datamar, uses an algorithm to discard or delete extra cases at random. Datamar considers the details of its approach to be proprietary and did not disclose them.²³

The effect of the likely voter models for two firms in the New Hampshire and California primaries for which we received micro-level datasets, is presented in Table 13. It reports the unweighted trial heat estimate based on the full sample, or all respondents asked the trial heat question, and the weighted trial heat estimate based only on the likely voters. Most datasets provided to the committee could not be used for this analysis because the polling organization did not disclose how likely voters were defined, or because likely voters were identified mid-interview and the trial heat questions were only administered to them. As a result, only three comparisons could be made.

²² SurveyUSA provided a Power Point version of a presentation at the 2005 AAPOR annual conference about its weighting algorithms. That Power Point is available at the Roper Center site that contains all of the information provided by the pollsters in response to AAPOR’s request for disclosure.

²³ Personal communications with the committee staff.

Gallup's likely voter estimate is clearly farther from the election outcome than the full sample estimate. In the full sample estimates, Obama led Clinton by 5 percentage points, while the weighting for likelihood and other factors produced an Obama advantage of 13 percentage points. The Gallup Organization uses the index construction method to measure likelihood to vote, which has been shown to introduce error variance into pre-election polls during general elections (Erikson, Panagopoulos and Wlezien, 2004). The application of their likely voter model increased Obama's proportion slightly, but decreased Clinton's support by a greater amount. Internal analysis at Gallup²⁴ led their editor-in-chief, Frank Newport, to conclude that a faulty likely voter model was the single biggest factor in their underestimation of support for Clinton in New Hampshire. The Gallup likely voter model included measures of enthusiasm and attention to the race, dimensions found in higher levels among Obama supporters following the Iowa primary. It is quite possible that for the earlier Gallup polls in New Hampshire, disproportionate numbers of Clinton supporters were classified as "unlikely voters" and were inappropriately dropped from the Gallup estimates (Erikson and Wlezien 2008). It is important to note, though, that even Gallup's full sample estimate would have indicated the wrong winner.

Table 13. Full Sample vs. Likely Voter Estimates

	n	Obama	Clinton	Other	Und.	Total	A	
NH Democratic Primary		36%	39%	24%		100%		
Gallup: LV estimate	778	41%	28%	29%	2%	100%	0.45	
Gallup: Full sample estimate	1,224	39%	34%	25%	3%	100%	0.22	
		Obama	Clinton	Other	Und.	Total	A	
CA Democratic Primary		n	43%	52%	5%	100%		
PPIC: LV estimate	644		28%	43%	18%	11%	100%	
PPIC: Full sample estimate	791		26%	45%	16%	13%	100%	
		McCain	Romney	Huckabee	Other	Und.	Total	A
CA Republican Primary		n	42%	35%	12%	11%	100%	
PPIC: LV estimate	392		29%	17%	10%	30%	14%	100%
PPIC: Full sample estimate	485		29%	15%	10%	30%	16%	100%

All sample sizes in this table are unweighted.

²⁴ This analysis was reported by Gary Langer of ABC News (Langer 2008b).

Analogous comparisons of full sample and likely voter estimates are reported for the Public Policy Institute of California's (PPIC) surveys in that state's primaries. For the California Democratic primary, the difference in estimates was only two percentage points; this reduced the overestimation of the margin between Clinton and Obama, as reflected in the value of A being closer to zero. On the Republican side, moving to a likely voter subset slightly reduced the substantial underestimate of Romney's support but did not change the underestimate of McCain's support. Overall, the value of A was improved in the Republican race by the move to a likely voter subsample.

Summary: The likely voter model appears to explain much of the error in the Gallup poll in New Hampshire, but we find no compelling evidence that it explains errors in the other pre-primary polls for which we had appropriate data to analyze. In fact, outside of the Gallup poll in New Hampshire, the use of a likely voter model did not change estimates of candidate support very much in relation to the candidate preference distribution for the entire sample in the other polls.

Calling Protocols

Errors in the pre-election estimates could also have stemmed from decisions about how the data were actually collected. All of primary polls were conducted by telephone, but these polls varied in the number of call attempts they made to each case. Studies have shown that increasing the number of call attempts in a survey can change the partisan composition of the sample (Traugott 1987; Keeter et al. 2000, 2006), and it might change the proportions of likely voters or the candidate preference distribution as well.

The committee was severely constrained in its ability to investigate any relationship between the number of calls attempted and the degree of error in the survey estimates. Only CBS News and Gallup provided datasets for their New Hampshire polls with the number of call attempts made on each case. The CBS News New Hampshire Democratic primary estimates are based on a re-contact survey rather than a fresh sample, which limits the generalizability of the results. CBS News conducted a maximum of six calls, and Gallup conducted a maximum of five calls. In both surveys, more than 94% of respondents provided data on the first or second attempt, undoubtedly a consequence of each firm trying to achieve a large enough sample size in a brief field period so that campaign events would not have an appreciable effect on candidate preferences. These distributions are shown in Table 14.

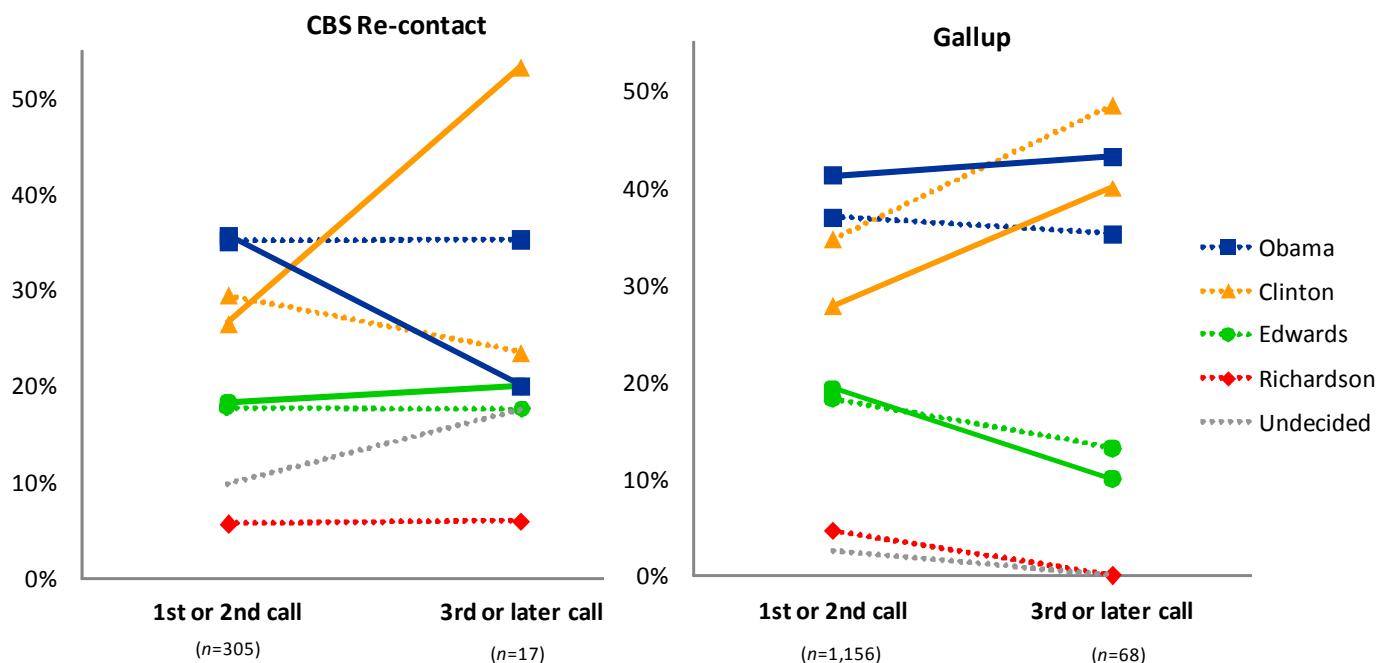
Table 14. Number of Call Attempts to Obtain an Interview for Two N.H. Democratic Pre-Primary Polls

	Number of interviews	
	<u>CBS Re-contact</u>	<u>Gallup</u>
1st call	752	1,762
2nd call	225	962
3rd call	39	149
4th call	5	12
5th call	3	3
6th call	<u>2</u>	<u>--</u>
	N=1,026	N=2,888

In our assessment of candidate preference by call effort, we found no statistically significant difference in the preference distribution for those who were interviewed on the first call or the second call. The small number of cases in the third or later call category limits the statistical power of this test, but the results are shown in Figure 6. The Gallup survey's unweighted estimates (dashed lines) indicate that respondents requiring 3 or more calls favored Clinton over Obama (49% to 35%), while those requiring 1 or 2 calls favored Obama (37% to 35%). When weighted estimates (solid lines) are considered, the difference in candidate preference by level of call effort is similar – with the hard-to-reach favoring Clinton and the easy-to-reach favoring Obama – though less dramatic.

The findings from the CBS News survey are more mixed, reflecting the fact that only 17 of the 322 re-contacted respondents required 3 or more calls. The weighted CBS estimates show the same pattern as the Gallup data, with harder-to-reach respondents favoring Clinton by a 2-to-1 margin, and the easy-to-reach favoring Obama. The unweighted CBS estimates, however, show the opposite: those requiring three or more calls were slightly more favorable toward Obama than those requiring one or two calls. The callback design of the CBS study and the small sample size in the “higher effort” group makes these data less suitable for this type of analysis than the Gallup data, which come from fresh RDD landline and cell phone samples. However, in both surveys, the more difficult to reach respondents were more likely to favor Clinton.

Figure 6. Vote Preference in New Hampshire Democratic Primary for Respondents Interviewed on the First or Second Call versus Respondents Interviewed after Three or More Calls²⁵



Analysis of the Gallup data offers some indication that the primary pollsters may have achieved more accurate estimates had they implemented a more rigorous (and expensive) data collection protocol. Had the surveys used, say, an 8 or 10 call maximum rule rather than a 5 or 6 maximum, it appears they may have tapped into Clinton support in New Hampshire that was missing from their final estimates. That said, this analysis is limited in statistical power and potentially confounded with real changes in preferences. The CBS News and Gallup surveys were fielded January 5th - 6th and January 4th - 6th, respectively. On average, those requiring more than two calls were most likely interviewed on the last or penultimate day of interviewing, while those requiring fewer calls were interviewed earlier. Changes in candidate preference at the individual level during this January 4th - 6th period might be misclassified in the Figure 6 analysis as differences between groups (i.e., those requiring low- versus high-effort calling). While such individual change may have occurred, we suspect it does not swamp the differences between the low- versus high-effort groups. We say this in part because Senator Clinton's emotional moment at the New Hampshire diner – which is speculated to have influenced some undecideds and tepid Obama supporters – did not occur until January 7th, after both studies had ended data collection. It is also important to keep in mind how the primary calendar influenced data collection. The New Hampshire primary was held just five days after the Iowa

²⁵ Dashed lines denote unweighted estimates; solid lines denote weighted estimates.

caucuses. It is unlikely that any “bounce” Obama received after Iowa would have dissipated by the time interviewing was underway for final New Hampshire polls.

Summary: We found a modest indication that the primary pollsters may have achieved slightly more accurate estimates in New Hampshire had they implemented a more rigorous (and expensive) data collection protocol. Respondents reached in three or more calls were more likely to support Clinton. In polls with very short field periods, the sample tends to be comprised of respondents contacted on the first few attempts, complicating assessments of the impact of interviews collected with more effort to contact respondents. This raises the prospect that sample management during the field period could have affected accuracy, with more prolonged effort producing better estimation. However, these results are based on only two surveys, one of which was unusual because it was a call-back study.

Timing of Data Collection

Another design feature often thought to affect the accuracy of pre-election surveys is the length of time between data collection and Election Day. Research on this topic has yielded mixed results (DeSart and Holbrook 2003). Some studies find that polls fielded closer to Election Day are more accurate (Crespi 1988); others find a null or even negative relationship (Lau 1994; Martin et al. 2005). We tested for a relationship between poll timing and accuracy in the Iowa, New Hampshire, South Carolina, and California primaries. The results, presented in Table 15, emanate from a richer analysis because the field dates of surveys are commonly reported by polling firms, as per AAPOR guidelines. We summarized the relationship between timing and accuracy with simple bivariate correlations.

Table 15. Correlation Between Poll Accuracy and Days Until Election^{1,2}

	Number of polls (n)	Correlation: Accuracy and Days Until Election	Correlation: Accuracy (abs) and Days Until Election
IA Democratic Caucus	29	0.18	-0.17
IA Republican Caucus	28	0.58	-0.04
NH Democratic Primary	46	-0.40	-0.30
NH Republican Primary	45	-0.80	0.81
SC Democratic Primary	26	-0.20	0.20
SC Republican Primary	21	-0.52	0.64
CA Democratic Primary	21	-0.64	-0.02
CA Republican Primary	21	0.68	0.07
WI Democratic Primary	5	-0.84	0.84
WI Republican Primary	4	0.98	0.98

¹Based on surveys fielded within one month of Election Day

²Smaller values of accuracy denote less deviation from the actual election results

These are the correlations between the absolute values of the accuracy scores (A) and the number of days out from the election. Simply put, if polls taken closer to the election were more accurate, then we would expect to observe a positive correlation: accuracy scores near zero are better than those farther from zero in either a positive or negative direction. For example, in the South Carolina Republican primary, the later polls were more reflective of McCain's three-point victory than earlier polls, resulting in positive correlation (.64).

The results in Table 15 demonstrate that the relationship between timing and accuracy varied greatly by election. In the problematical New Hampshire Democratic primary, there is a negative correlation (-.30 for the absolute value of A) between accuracy and temporal distance from the election. This quantifies the pattern in Figure 2 where we see that the final polls fared slightly worse on average than those fielded earlier (before the Iowa caucuses). We found a similar though weaker relationship (-.17), computed on the same basis, between poll timing and accuracy in the Iowa Democratic caucuses, and we found essentially no relationship in the Iowa Republican caucus or the California Democratic and Republican primaries. The only races in which polls conducted later were noticeably more accurate are the New Hampshire Republican primary, the South Carolina Republican and Democratic primaries, and the Wisconsin Democratic and Republican primaries – and most of these events were later in the calendar, when the field of candidates was smaller. However, in the most problematic races, particularly the Iowa Democratic caucus and the New Hampshire Democratic primary that preceded it, the final polls did not seem to improve as Election Day approached.

We attempted to test this further by using the micro-level datasets provided to the committee. Three of the New Hampshire datasets contained a variable for the interview date (CBS and Gallup).²⁶ We merged cases containing relevant common variables from these datasets and used a logistic regression model to test whether the timing of the interview relative to the election had a significant relationship with vote preference for Hillary Clinton, while controlling for other factors. Specifically, we estimated a logistic regression with vote preference for Clinton as the dependent variable and number of days until election, gender, age, education, survey firm, interviewer demographics, and Democratic Party affiliation as the independent variables.

This approach is limited in two important ways. First, the key independent variable has a very narrow range of values because the interviewing dates for these three surveys were between January 4th

²⁶ The University of New Hampshire dataset also contained the interview date, but it did not contain all of the other predictors in the model. A reduced model was estimated so that cases from all three surveys were included, and the results did not change appreciably. There was still no significant effect from the number of days until the election.

and 7th. Second, any observed effect from the number of days until the election will be confounded to some extent by other factors such as ease of contact, as reported earlier. The regression analysis suggests that the number of days until the election did not have a significant effect on the likelihood of favoring Clinton after controlling for the other factors. The estimated model parameters are provided in Appendix Table 2. This null finding does not rule out the possibility that vote preferences changed in the days leading up to the New Hampshire primary, but we find no support for a shift toward Clinton during the January 4–7 time period.

Summary: We found that the timing or field periods of the polls may have contributed to estimation errors in the early Democratic events in Iowa and New Hampshire, though we lack the full set of relevant data for proper evaluation. The timing of the New Hampshire primary, so closely following the Iowa caucus, appears to have contributed to instability in or changing preferences and, in turn, in the poll estimates.

Social Desirability

Some election observers speculated that the New Hampshire polls overestimated support for Obama because some respondents told interviewers that they would vote for him but actually voted for Clinton (Nichols 2008; Robinson 2008). Such intentional misreporting in voter preference polls is attributed to latent bigotry or intolerance in conjunction with an inclination to provide socially desirable responses.²⁷ However, in the New Hampshire pre-primary polls, the estimation error did not derive from overestimating support for Obama – which could have been driven by latent racism among respondents – but from underestimating support for Clinton. Therefore, latent misogyny cannot explain the errors in the New Hampshire polls because it would have had the opposite observed effect: the polls would have overstated support for Clinton rather than understated it.

Several compelling pieces of evidence suggest that the New Hampshire estimation errors were probably not caused by the “Bradley effect” – or the tendency for respondents to report a preference for a black candidate (Obama) but vote instead for a white opponent. A meta-analysis by Hopkins

²⁷ There is a dispute about whether and to what extent a “Bradley effect” ever existed. In terms of the original 1982 gubernatorial election in California, Tom Bradley received more votes than George Deukmejian at the polls but lost in the absentee balloting by a much larger amount in the first election when parties could organize efforts to make absentee ballots available to voters. The Republicans outmaneuvered the Democrats in this regard. But in 1989, in relation to the election returns, there appeared to be an over-report of support for David Dinkins in the pre-election polls for the New York mayor’s race and in an exit poll estimating support for L. Douglas Wilder in the Virginia governor’s race. (Traugott and Price, 1992) See also a discussion by Lance Tarrance during the 2008 general election campaign (http://www.realclearpolitics.com/articles/2008/10/the_bradley_effect_selective_m.html).

(2008) indicates that while the Bradley effect did undermine some state-level polls in previous decades, there is no evidence for such an effect in recent years. In the 2008 general election, the very accurate final poll estimates of Barack Obama's fairly decisive victory over John McCain dispelled suspicion that the Bradley effect was at play during the final weeks of the fall contest. There is also a conspicuous lack of evidence for a Bradley effect in the primary contests outside of New Hampshire. Of the 81 polls conducted during the final 30 days of the Iowa, South Carolina, California, and Wisconsin contests, the vast majority (86%) over-estimated Clinton's relative vote share, while just 14% over-estimated Obama's relative vote share. This finding is based on the signed direction of A for each survey.²⁸ Furthermore, as reported in Table 3, poll estimates of Obama's vote share in New Hampshire were quite accurate – it was only Clinton's share that was consistently underestimated.

However, it is still possible that intentional misreporting occurred during the lead up to the New Hampshire Democratic primary because of the interaction between the race of the interviewer and the race of the respondent. If social desirability influenced respondents' answers, we would expect to observe more support for the African-American candidate when the interviewer was African American than when the interviewer was not, based upon an assumption that the respondent could correctly infer the race of the interviewer over the telephone. If respondents were answering truthfully, we would expect to find no statistically significant difference between the vote preferences recorded by African-American interviewers and interviewers of other races.

The interviewer effects approach has substantial drawbacks. To the extent that respondents misreport regardless of interviewer race, this test will underestimate a social desirability effect. Also, interviewing staffs in the United States tend to be comprised mostly of Caucasians; consequently, the number of interviews conducted by African Americans is often low, yielding low statistical power for the test. And interviewers were not assigned randomly to cases so that the race of interviewer effects could be discerned without confounds. Results from race-of-interviewer effects analysis should be interpreted with these factors in mind.

In October, Gary Langer (2008a) of ABC News reported results from his analysis in the general election of the relationship between the race of the interviewer and the respondent. He found no evidence of racially motivated misreporting. Other pollsters also found no such evidence. As

²⁸ The reader should remember, as shown in Table 3, that the inclusion of undecided voters in the candidate preference distribution implies that individual candidate support levels will be underestimated. In New Hampshire, for example, 22 out of 22 polls in the month leading up to the primary underestimated support for Obama, and 18 out of 22 polls underestimated support for Clinton.

mentioned above, however, this does not rule out the possibility that racially motivated misreporting occurred during the primaries.

CBS News polling director Kathleen Frankovic (2008) used panel data to test whether racial attitudes affected the New Hampshire polls. She noted that voters who are concerned that their candidate preference may be socially unpopular could contribute to polling error in two ways. They could misreport their true preference or they could decline to be interviewed altogether. The CBS panel data was used to evaluate the latter hypothesis – were voters opposing Obama less likely to be interviewed in New Hampshire? Frankovic’s analysis suggests that the answer is “No.” The January response rate for those who supported Obama in November was similar to the January response rate for those who supported Clinton in November (74% and 68%, respectively). This difference is in the expected direction, but the magnitude is not large enough to explain the error in the polls. CBS News post-stratified their January sample to account for this difference in response rates.

While informative, this analysis has an important limitation. The test is based on people who already agreed to participate in a survey. This step may have filtered out many of those who would decline a survey request for fear of offending someone with their candidate preference. This limitation could explain the lack of a large difference in the January response rates. The CBS News analysis, therefore, does not rule out the possibility that Obama supporters were more likely to respond than those who did not support him.

The committee was also able to conduct analysis on the topic. Three survey organizations provided data to the committee that could be used to test for a race-of-interviewer effect in the New Hampshire Democratic primary. Gallup, CBS News, and the University of New Hampshire included the race of the interviewer in their survey dataset; but only CBS News included race of the respondent. Based upon the 2006 U.S. Census estimate of the proportion of the New Hampshire population that is white (95.8%), we assumed in our analysis that all of the survey respondents were white.²⁹ We combined data from the three surveys to increase statistical power and compared reported vote preference among respondents who spoke with a white interviewer to respondents who spoke with an African-American interviewer. The results are displayed in Table 16. In the pooled analysis, Obama led Clinton 36% to 35% when the interviewer was white, and he led 43% to 29% when the interview was black. This finding is in the direction of a social desirability effect and is statistically significant. Using just the CBS News dataset, we performed the same analysis looking only at white respondents.

²⁹ Available at <http://quickfacts.census.gov/qfd/states/33000.html>.

We find that black interviewers recorded higher support for Obama than white interviewers. Although the effect is not quite statistically significant by standard levels due to small sample size ($p=.13$), it is quite noticeable and in the expected direction. We also tested for this effect in a multivariate setting. The race of the interviewer was a significant predictor of vote preference for Clinton when controlling for other factors in the logistic regression presented in Appendix Table 2.

Table 16. NH Democratic Primary Preferences by Race of Interviewer

----- All respondents ¹ -----		----- White respondents only ² -----		
	White Interviewer	Black Interviewer	White Interviewer	
Clinton	35	29	Clinton	31
Obama	36	43	Obama	33
Other	<u>29</u>	<u>28</u>	Other	<u>36</u>
	100%	100%		100%
	(n=1,768)	(n=304)	(n=202)	(n=50)

¹These figures come from combining cases from the UNH, CBS, and Gallup surveys. Respondent race was not reported in the UNH or Gallup datasets, so results are reported for all respondents. Pearson $\chi^2=6.58$, $df=2$, $p=0.04$

²These figures come from the CBS survey, which reported respondent race. Pearson $\chi^2=4.02$, $df=2$, $p=0.13$

Overall, these findings suggest that the misreporting of candidate preference due to racial sensitivity to black interviewers may have contributed to the overstatement of support for Obama relative to Clinton in the New Hampshire Democratic primary polls. It could also be the case that a social desirability effect was at play even when the interviewer was white. Absent individual-level vote data to append to these datasets, we are unable to test that hypothesis rigorously.

We used these same pooled datasets to test whether vote preferences reported in the polls were influenced by the gender of the interviewer. Just as social desirability pressure may have led some Clinton supporters to report they would vote for Obama, so might some Obama supporters report they would vote for Clinton if they were interviewed by a female interviewer. We did not find strong evidence that interviewer gender influenced responses. The bivariate findings are presented in Table 17. Among male respondents, the Obama lead was 14 percentage points for male interviewers and 17

percentage points for female interviewers, indicating no effect from interviewer gender. Among female respondents, Clinton had a 3 percentage point lead when the interviewer was male and a 6 percentage point lead when the interviewer was female. This finding for female respondents is in the direction of a social desirability effect and is statistically significant. Interviewer gender is only marginally significant ($p=.079$) in the multivariate model. Any misreporting favoring Clinton in New Hampshire does not, however, help to explain why her support was underestimated in the polls.

Table 17. NH Democratic Primary Preferences by Sex of Interviewer

<i>----- Male respondents¹ -----</i>		<i>----- Female respondents² -----</i>		
	<u>Male Interviewer</u>	<u>Female Interviewer</u>	<u>Male Interviewer</u>	
Clinton	29	25	Clinton	35
Obama	43	42	Obama	32
Other	<u>28</u>	<u>33</u>	Other	<u>33</u>
	100%	100%		100%
	(n=346)	(n=444)	(n=641)	(n=715)

Source: Combined results from the UNH, CBS, and Gallup surveys

¹Pearson $X^2=3.34$, $df=2$, $p=0.19$

²Pearson $X^2=7.68$, $df=2$, $p=0.02$

Both the race-of-interviewer and gender-of-interviewer analyses are limited because they rely solely on survey responses. Ideally, we would be able to compare survey responses about candidate preference with actual voting behavior, and the interviewers would have been randomly assigned to cases rather than in conjunction with their schedules or other factors.

Summary: We found mixed evidence for social desirability effects on polling errors. Social desirability pressures may explain a small proportion of the error but probably no more than that. In a pooled analysis of three New Hampshire surveys, we find that support for Obama is significantly greater when the interviewer is black than when he or she is white. In the same analysis, however, Obama is still favored over Clinton among respondents interviewed by a white interviewer, and the number of interviews taken by black interviewers was too small to affect the overall estimates.

Weighting

In the common application of the technique, pre-election pollsters use weighting to align the demographic characteristics of their sample with known characteristics of the voting population after the interviewing is completed, usually based on Census data or information about the characteristics of registered voters.³⁰ Constructing a survey weight for primary pre-election polls is complicated by several factors. As with nearly all telephone surveys of the U.S. public, responding samples contain disproportionate numbers of women, seniors, and whites (among other demographic characteristics). Furthermore, the demographic and party identification characteristics of a primary electorate can shift substantially from election to election, making it difficult to identify appropriate parameter estimates for weighting. For example, what proportion of the New Hampshire Republican primary voters will be registered Republicans versus registered Independents, or others? In effect, the selection of weighting variables and the construction of the weights themselves are akin to building a likely voter model.

In 2008, primary pollsters addressed the weighting issue in a number of ways. Some procedures were implemented at the sampling stage, while others were implemented after data collection was completed. We discuss sampling and post-survey adjustment procedures together because, essentially, they were used as two different tools to accomplish the same task: achieving the appropriate levels of representation in the poll for certain groups that most commonly included women/men, young/old, white/black, Hispanic/non-Hispanic, and registered or self-identified Independents, Democrats, and Republicans, depending upon the state and available data from it.

Table 18 presents a summary of the procedures implemented for each poll studied by the committee. At the sampling stage, polls using registration-based sampling (RBS) made use of the information on the frame. Most of the RBS pollsters drew their samples based on what they knew about each person's voting history, party registration, and/or demographics at some time in the past. At the post-survey adjustment stage, pollsters used several different procedures that appear related to the mode of administration. Two of the three IVR pollsters who were willing to discuss their methodology described how they deleted cases to make their samples more representative.³¹ They randomly deleted a subset of cases from demographic groups, such as older women, who were overrepresented in the

³⁰ In some cases, pre-election pollsters are now combining weights to account for sample representation with adjustments made for likelihood of voting. For example, they assign unlikely voters a weight of zero to delete them from the analysis. There is additional discussion of such procedures in the section on Likely Voter Definitions, above.

³¹ This information was provided in follow up telephone conversations and not in their original submissions of materials for their polls.

responding sample. None of the CATI pollsters used this deletion technique. Instead, the CATI pollsters generally used an iterative-proportional fitting technique (e.g., raking) to align the responding sample with population control totals for several relevant dimensions.

The committee is not in a position to evaluate fully the relative merits of case deletion versus post-stratification weighting. With full disclosure of the specific procedures employed, the application of weights has an advantage in that a secondary analyst can compare the weighted distribution to the unweighted distribution to assess the impact of the weights. And if such an analyst does not agree with the algorithm or its results could in principle apply their own population-based nonresponse adjustment and assess its relative efficacy compared to the original. Deleting a case could have roughly the same effect on point estimates as assigning it a weight close to zero under certain assumptions, such as a simple random procedure for deletion; but the absence of such a case from the publicly available dataset does not provide a mechanism for assessing alternative models for deletion. Table 6 shows that the IVR polls (some of which use deletion) performed at least as well as the CATI polls (which do not use deletion). In terms of poll accuracy, we found no discernable difference between these adjustment techniques. But given the potential confounding of mode and adjustment technique, we could not evaluate the relative merits of these two adjustment techniques without additional information.

An issue of serious concern, however, is whether or not these procedures are appropriately reflected in the margins of error. If cases are deleted, the margins of error should be increased accordingly. None of the pollsters who used the case deletion approach disclosed the number of cases that were deleted. This information is critical to understanding the magnitude of the adjustment and implications for the variance of the survey estimates. Similarly, for post-stratified surveys, the variance in the weights should be reflected in the margin of error.

Table 18. Description of Sampling and Adjustment Procedures

Poll	Sampling Procedures	Mode	Adjustment Procedures
ARG	Landline RDD	CATI	No weighting procedures were used
CBS News	Callback of Nov '08 landline RDD sample	CATI	The weight was based on the population weight from the original Nov. poll, post-stratified to the Nov. poll RV distributions on age, sex, likelihood of voting, education, and candidate choice.
Clemson	?	CATI	?
Datamar	RV list. The sample was selected based on registration date, voting history, and party registration.	IVR	Data were adjusted for age, gender, and political party using random deletion.
Ebony/Jet	?	CATI	?
Field Poll	RV list with registered Republican sample augment	CATI	Data were weighted to CA registered voter population estimates for party registration, region, age, gender, race/ethnicity and permanent absentee voter status.
Gallup	Landline RDD + Cell RDD with cell-only screener	CATI	Data were weighted to NH adult population estimates for gender, age, and region. Weights were adjusted for #phone lines, #cell phones, #adults in the HH, #adults sharing cell phone.
LA Times	Landline RDD	CATI	Data were weighted to CA adult population estimates for age, education, ethnicity, gender and party registration*region.
Marist	Landline RDD. The sample was selected in proportion to the voter registration in each county.	CATI	Data were weighted to NH adult population estimates for gender, income, age, and region based on turnout in comparable elections. Both actual election data and exit poll results were used.
Mason-Dixon	Landline RDD	CATI	No weighting procedures were used
Opinion Dynamics	Landline RDD	CATI	Data were weighted to NH adult population estimates for age, ethnicity, and gender.
PPIC	Landline RDD	CATI	Data were weighted to estimates of the CA adult population for party registration, region, gender and age.
Public Policy Polling	RV list landlines. Sample was selected based on voting history.	IVR	Data were adjusted for gender using random deletion and weighted for race in the SC Democratic poll. Data were weighted for gender in the SC Republican poll. Data were weighted for race and adjusted for gender and age using random deletion in the WI Democratic poll. Data were weighted for gender and adjusted for age using random deletion in the WI Republican poll.
Rasmussen	Landline RDD	IVR	Data were weighted to NH adult population estimates for age, race, gender, party, and "other factors."
Research 2000	Landline RDD	CATI	?
RKM	RV list	CATI	No weighting or estimating procedures were used
Str. Vision	?	CATI	?
Suffolk	Landline RDD (NH) RV list (CA)	CATI	No weighting procedures were used. Quotas were set to demographics by county from last two "like" presidential primary elections where there was no incumbent president on either primary ballot.
SurveyUSA	Landline RDD	IVR	Data were weighted to CA adult population estimates for gender, age, race, ethnicity, and region.
Univ. of NH	Landline RDD	CATI	Data were weighted to NH adult population estimates for gender and region. Weights were adjusted for #phone lines and #voters in HH.
Zogby	Listed landlines	CATI	Data were weighted to NH adult population estimates for party, age, gender, and religion.

In addition to presenting this overview of procedures, we sought to evaluate empirically the effect of weighting in the primary polls by comparing weighted trial heat estimates to unweighted estimates where possible. The results are shown in Table 19. In general we found the unweighted estimates were very similar to the weighted estimates, and the few differences do not lead the weighted estimates to be consistently more accurate than the unweighted estimates. For example, in the New Hampshire Democratic primary, the unweighted University of New Hampshire estimates were 38% Obama and 33% Clinton, for an A value of 0.21. The weighted estimates were 39% Obama and 30% Clinton for an A value of 0.33. The reader may recall that higher A values reflect a greater deviation between the survey estimates and the actual election result. The results in Table 19 suggest that the weights applied to several of the primary polls were not particularly effective in reducing errors in the estimates of candidate preference. But these weights were often combined with the likely voter model. This makes it difficult to segregate the effects due to weighting from the effects due to likely voter estimation. This seems to be especially true for Gallup, whose unweighted estimates were the closest to the Election Day outcome – though they still showed a slight (37% to 36%) Obama lead – but whose weighted estimates were furthest from the final result.

Summary: We found strong evidence that faulty weighting techniques explain some (but not all) of the polling errors. In three of the four surveys suitable for examination, the weighted New Hampshire estimates were less accurate than the unweighted estimates. Because all four unweighted estimates still had Obama in the lead, however, weighting cannot explain all of the error in the polls.

Table 19. Accuracy of Weighted versus Unweighted Estimates

	Obama	Clinton	Other	Und.	Total	A	
NH Democratic Primary							
Election Result	36%	39%	24%		100%		
Univ. of NH <i>weighted</i>	39%	30%	25%	6%	100%	0.33	
Univ. of NH <i>unweighted</i>	38%	33%	23%	6%	100%	0.21	
CBS <i>weighted</i>	35%	28%	28%	9%	100%	0.29	
CBS <i>unweighted</i>	35%	29%	25%	10%	100%	0.25	
Opinion Dynamics <i>weighted</i>	32%	28%	28%	12%	100%	0.20	
Opinion Dynamics <i>unweighted</i>	33%	28%	28%	12%	100%	0.23	
Gallup <i>weighted</i>	41%	28%	29%	2%	100%	0.45	
Gallup <i>unweighted</i>	37%	36%	25%	2%	100%	0.10	
	Obama	Clinton	Other	Und.	Total		
CA Democratic Primary							
Election Result	43%	52%	5%		100%		
Survey USA <i>weighted</i>	42%	52%	2%	4%	100%	-0.04	
Survey USA <i>unweighted</i>	41%	52%	5%	2%	100%	-0.06	
PPIC <i>weighted</i>	28%	43%	18%	11%	100%	-0.25	
PPIC <i>unweighted</i>	26%	43%	19%	13%	100%	-0.33	
Field Poll <i>weighted</i>	34%	36%	12%	18%	100%	0.12	
Field Poll <i>unweighted</i>	31%	38%	13%	18%	100%	-0.04	
	McCain	Romney	Huckabee	Other	Und.	Total	A
NH Republican Primary							
Election Result	37%	32%	11%	20%		100%	
Univ. of NH <i>weighted</i>	31%	26%	13%	25%	5%	100%	0.02
Univ. of NH <i>unweighted</i>	32%	26%	13%	24%	5%	100%	0.05
Gallup <i>weighted</i>	34%	30%	13%	20%	3%	100%	-0.03
Gallup <i>unweighted</i>	36%	29%	12%	20%	3%	100%	0.06
	McCain	Romney	Huckabee	Other	Und.	Total	A
CA Republican Primary							
Election Result	42%	35%	12%	11%		100%	
Survey USA <i>weighted</i>	39%	38%	11%	10%	2%	100%	-0.17
Survey USA <i>unweighted</i>	39%	37%	11%	10%	2%	100%	-0.16
PPIC <i>weighted</i>	29%	17%	10%	30%	14%	100%	0.33
PPIC <i>unweighted</i>	29%	18%	9%	31%	13%	100%	0.26
Field Poll <i>weighted</i>	32%	24%	13%	16%	15%	100%	0.09
Field Poll <i>unweighted</i>	30%	24%	12%	17%	17%	100%	0.03

Time of Decision

Two survey organizations polling in New Hampshire - Gallup and CBS News - conducted panel studies in which the same sample of voters was interviewed twice. These types of studies can be highly informative because they allow researchers to evaluate changes in preference the level of the individual voter, not just aggregate change. The two New Hampshire panel studies were designed quite differently. The Gallup study is a “before and after” panel in that the sample was interviewed immediately before the January 8th primary (January 4-6, 2008) and immediately after (January 11-29, 2008). The CBS News study involved two surveys with the same sample that were both conducted before the primary, although they were spaced several months apart (November 2-12, 2007 and January 5-6, 2008).

The re-interview rates in both studies were quite high. The original Gallup sample included 1,179 respondents planning to vote in the Democratic primary and 1,180 who said that they planned to vote in the Republican primary. The Gallup post-election survey included interviews with 818 of them (69%) who reported voting in the Democratic primary and 800 (68%) who reported voting in the Republican primary. In the CBS News study, three-quarters (77%) of the likely Democratic primary voters identified in November completed the January re-interview.

The CBS News study measured change in candidate preferences between November 2007 and early January 2008. The results, presented in Table 20, suggest significant shifts in support during that time. Among those supporting Obama just days before the primary, only half had been supporting him in November. Those self-reported late-comers to the Obama campaign included significant numbers of former Clinton supporters.

The composition of Clinton supporters looked quite different. The vast majority (87%) of those supporting Clinton in early January had been planning to vote for her since November. The CBS News study found no evidence that Clinton captured significant numbers of supporters from any of the other major candidates during that time. However, the study did not measure any changes in preferences that took place in the last two or three days of the campaign. Results from a trial heat follow-up question suggest that this gap in measurement may have been important. When asked if their mind was made up or it was too early to say for sure whom they would vote for, almost half (46%) of John Edwards supporters said that it was too early. This suggests that many Edwards’ supporters might have anticipated his third place finish and were strongly considering voting for one of the top two contenders.

Table 20. Profiles of N.H. Democratic Primary Candidates' Supporters in January 2008

	If the primary were held today (January 5/6), Which candidate would you vote for?			
	<u>Clinton</u>	<u>Obama</u>	<u>Edwards</u>	<u>Richardson</u>
Supported since Nov. 07	87	50	45	58
Supported other candidate in Nov. 07	8	42	40	8
Formerly supported Clinton	--	19	24	0
Formerly supported Obama	1	--	0	0
Formerly supported Edwards	0	3	--	0
Formerly supported Richardson	1	3	13	--
Formerly supported other candidate	6	17	3	8
Undecided in Nov. 07	<u>4</u>	<u>9</u>	<u>16</u>	<u>33</u>
	100%	100%	100%	100%
(It's 2-3 days before the election...) Is your mind made up or is it too early to say for sure?				
Mind is made up	83	71	54	72
Too early to say for sure	<u>17</u>	<u>29</u>	<u>46</u>	<u>28</u>
	100%	100%	100%	100%
Sample size	(87)	(95)	(48)	(18)

Source: CBS callback study of NH Democratic primary likely voters

The Gallup study began where the CBS study left off, albeit with a different group of New Hampshire voters. Results from the Gallup post-election interview are presented in Table 21. There is some evidence that Clinton won the vote of many Edwards supporters. Some 6% of Clinton voters said that they supported Edwards in the pre-election poll. Similarly, about 4% of Obama voters had recently come from the Edwards camp. For his part, Edwards looks to have attracted some former Obama and Clinton supporters. Unfortunately, sample sizes in both of these studies are too small to say with confidence whether many of these shifts within the electorate were significant.

The impact of this movement toward Clinton on the self-reported vote distribution is shown in Table 22, in which the data are not weighted. The level of self-reported support for Obama (about 37%) did not change substantially from the pre-election to the post-election survey. The proportion endorsing Clinton, however, increased by 4 percentage points. Taken together, the CBS News and the Gallup panels provide some evidence for a real, sizable shift in preferences to Obama in the final weeks of the campaign and a separate shift to Clinton in the final days. This changing terrain complicated the task of the pollsters and increased their difficulty in producing accurate projections.

Table 21. Post-election Profiles of N.H. Democratic Primary Voters

	<i>For whom did you vote?</i>			
	<u>Clinton</u>	<u>Obama</u>	<u>Edwards</u>	<u>Richardson</u>
Supported this candidate before the primary	84	90	83	85
Supported other candidate before the primary	14	9	15	11
Formerly supported Clinton	--	3	4	0
Formerly supported Obama	6	--	8	7
Formerly supported Edwards	6	4	--	4
Formerly supported Richardson	1	1	3	--
Formerly supported other candidate	1	1	0	0
Undecided before the primary	<u>2</u> 100%	<u>1</u> 100%	<u>2</u> 100%	<u>4</u> 100%
Sample size	(292)	(276)	(120)	(27)

Source: Gallup New Hampshire re-interview study.

Table 22. Candidate Preferences Before and After N.H. Primary

	<i>Which candidate would you vote for?</i>	<i>For whom did you vote?</i>	<i>Change</i>
	<u>Pre-election</u>	<u>Post-election</u>	
Hillary Clinton	35.4	39.5	+4.1
Barack Obama	37.7	37.3	-0.4
John Edwards	18.1	16.2	-1.9
Bill Richardson	4.5	3.6	-0.9
Dennis Kucinich	2.2	2.2	0.0
Mike Gravel	0.1	0.1	0.0
Other	0.1	0.3	0.2
None/no opinion	<u>1.9</u> 100.0%	<u>0.8</u> 100.0%	-1.1
Sample size	(740)	(740)	

Source: Gallup New Hampshire re-interview study. Figures are unweighted and based only on respondents who (1) reported before the election that they intended to vote in the Democratic Primary and (2) reported after the election that they voted in the Democratic primary.

The Gallup re-interview also provides some empirical insights on why some Democratic voters decided during the final days of the campaign to vote for Clinton. The survey explored voter reactions to the January 5th Democratic debate, Clinton's emotional campaign appearance on January 7th, and get-out-the-vote (GOTV) efforts conducted January 7th and 8th. Each of these three events appears to account for a small increase in Clinton's support at the close of the campaign.

Among Democratic primary voters, viewership of the final debate was quite broad. About two-thirds (65%) reported watching the debate, and an additional 23% reported hearing or reading news coverage about it. Only a small fraction (4%) of those who watched or saw news coverage said that they changed their candidate preference because of the debate.

Coverage of Clinton's January 7th campaign appearance was followed closely by voters as well. Eight in ten (82%) of Democratic primary voters said that they had seen the video of the campaign appearance, and most viewers said that their reaction was positive or neutral. The Gallup re-interview also asked Clinton voters whether a number of different considerations factored into their decision to vote for her. One of the considerations was the video of that campaign appearance, and 15% reported that this was a factor in their vote.

Another speculation about a cause for late decisions was a highly publicized campaign event in which Senator Clinton responded in an uncharacteristically emotional way to a voter's question. The day before the primary election, a woman in a Portsmouth asked Clinton how she stays so "put together" on the campaign trail.³² Up until that point, Clinton had a reputation for being somewhat steely and hard-driving, but the question triggered a reflective, emotional response. This exchange revealed a more human side of Clinton that may have appealed to some Democratic voters, including some tepid Obama supporters. However, because most of the New Hampshire polls ended data collection prior to this event, these polls would have missed any related last-minute shift to Clinton.

The only information that we have about when people decided on their choice comes from self-reports of that decision in the exit polls. These are not easy data to evaluate because the exit poll respondents may have had difficulty in responding to or interpreting this question. For example, if a New Hampshire voter had been an early supporter of Hillary Clinton and then reconsidered his or her preference after Iowa but ended up voting for her on Election Day, would they have responded that they decided months ago or in the last day or two? On average, the pre-election polls suggested only

³² The exact question that prompted the response was "How did you get out the door every day? I mean, as a woman, I know how hard it is to get out of the house and get ready. Who does your hair?"

about 3% of likely voters were undecided, but about 17% of voters told exit poll interviewers that they made their candidate choice on Election Day.

Table 23 shows proportions of respondents that were reportedly late deciders in the primary elections. Between 17% and 19% of those who voted in the New Hampshire Democratic or Republican primary, respectively, said they made up their mind on the day they voted. This was only slightly more than those who said they made up their minds at the same point in the same 2000 events (15% and 14%, respectively). Almost 40% said they made up their minds during the final three days of the 2008 campaign (38% of those voting in the Democratic primary and 39% in the Republican primary). This was the same as in the Democratic primary in 2004 (35%), but higher than the proportion of late deciders in either primary in 2004 (both 26%).

Table 23. Time of Decision in Recent Primaries

	<i>When did you make up your mind?</i>	
	<u>Today</u>	<u>Final 3 days</u> ¹
2008 NH Democratic Primary	17%	38%
2008 NH Republican Primary	19%	39%
2004 NH Democratic Primary	--	35%
2000 NH Democratic Primary	15%	26%
2000 NH Republican Primary	14%	26%
2008 SC Democratic Primary	--	21%
2008 SC Republican Primary	--	--
2004 SC Democratic Primary	19%	33%
2000 SC Republican Primary	9%	19%
2008 CA Democratic Primary	--	26%
2008 CA Republican Primary	--	27%
2004 CA Democratic Primary	14%	24%
2000 CA Republican Primary	10%	20%
2000 CA Democratic Primary	12%	23%
2008 WI Democratic Primary	--	21%
2008 WI Republican Primary	--	26%
2004 WI Democratic Primary	18%	37%

Source: NEP/NES Exit Poll (Entrance Poll in IA)

¹Includes respondents reporting that they decided "today"

A closer look at the late deciders in the 2008 New Hampshire Democratic primary exit poll does not show enough late movement to Clinton to explain the error in the polls. According to analysis by Gary Langer of ABC News (2008b), the 17% of voters who said they made up their mind on the last day went narrowly for Clinton (39% to 36%) – a margin too small to explain fully the overestimation of support for Obama. Another 21% reported that they decided in the last three days, and they split narrowly for Obama (37% to 34%), again insufficient to explain the estimation problems in the pre-election polls.

Summary: We found that decision timing – in particular, late decisions – may have contributed significantly to the error in the New Hampshire polls, though we lack the data for proper evaluation. The fact that Clinton's "emotional moment" at the diner occurred after nearly all the polls were complete adds fuel to this speculation, as does the high proportion of New Hampshire Democratic primary voters who said they made up their minds during the final three days. It is also true that the percentages of voters reporting they made up their minds on Election Day or in the three preceding days are not substantially different from historical levels.

Participation of Independents Assuming an Obama Victory

Another hypothesis for the error in the New Hampshire polling estimates concerns the relative proportion of self-described Independents in the survey sample and in the electorate.³³ Some have speculated that Independents who liked both Obama and McCain could have been under the impression from the polls that Obama had locked up the Democratic race, and so they decided to participate in the Republican primary and support McCain. This hypothesis suggests that: (1) Independents should comprise a *larger* segment of the Democratic primary electorate in the pre-election polls than in the exit poll because at the time that they were interviewed they had not decided to abandon Obama to support McCain, and (2) Independents should comprise a *smaller* segment of the Republican primary electorate in the pre-election polls than in the exit poll for the same reason. We find mixed support at best for this hypotheses based on the four pre-election surveys that are reported in Table 24.

³³ New Hampshire has a party registration system whereby citizens must register as a Democrat, Republican or Undeclared. The Undeclared registrants may opt into either party's presidential primary. This formal designation is not what is typically measured in the pre-election polls, however; that is self-reported party identification. Andrew Smith of the University of New Hampshire reports that the Secretary of State counted 121,515 Undeclared voters in the Democratic primary (42.1% of all voters) and 75,522 Undeclared voters in the Republican primary (31.3% of all voters).

Table 24. Estimated Percentage of NH Primary Voters Who Were Independents

	<u>% Democratic primary voters who were Independents</u>	<u>% Republican primary voters who were Independents</u>	<u>Sample Sizes</u>
Exit Poll	44%	37%	(1955, 1520)
CBS News	45%	--	(323)
Univ. of New Hampshire	41%	34%	(599, 492)
Opinion Dynamics	39%	--	(500)
Gallup	43%	30%	(778, 776)

Each of these surveys ended data collection on January 6th, two days prior to the primary election. According to the New Hampshire exit poll, 44% of Democratic primary voters were Independents, compared to 37% of Republican primary voters, so they formed a slightly larger proportion of the Democratic than the Republican primary electorate. The pre-election surveys estimated that this figure would be between 39% and 45% on the Democratic side, a fairly narrow range that encompasses the exit poll reading. On the Republican side, the pre-election polls estimated that 30% (Gallup) or 34% (University of New Hampshire) of voters would be Independents, lower than the 37% indicated in the exit poll. These estimates for the Republican race suggest that some Independents classified as Obama voters before the election may have ended up McCain voters at the polls. This effect, however, is not especially large and is not conclusive evidence for an Independent shift.

Summary: We found little compelling information to suggest that Independents, by deciding in disproportionate numbers and at the last minute to vote in the Republican rather than the Democratic primary, contributed to the New Hampshire polling errors. The proportion of Independent voters in the Democratic primary pre-election polls is comparable to the proportion in the exit poll. Also, the differences between the exit poll and the pre-election surveys are not large enough to explain the errors in the polls.

Allocations of Those Who Remained Undecided

In some pre-election surveys, pollsters allocate respondents who remain “undecided” to yield proportions reflecting support for each candidate that add to 100%. Pollsters may use a number of different allocation approaches, any of which may have an effect on the accuracy of their estimates. However, only one of the pollsters who collected data in the contests under study, the University of

New Hampshire Survey Center, used an allocation method. Without allocation, they showed a 9 percentage point lead for Obama over Clinton, and with allocation, they showed the same lead.³⁴ Hence this cannot explain differences between the final pre-election estimates and the outcome of the elections.

Ballot Order Effects

Another possible explanation for differences between the pre-primary poll estimates and the election outcomes concerns measurement in the election itself, rather than in the polls. Political methodologists have documented a small but non-trivial bias in favor of candidates listed first on election ballots (Miller and Krosnick 1998), one that appears to be robust in primary elections (Ho and Imai 2008). This bias is a version of a primacy effect (the opposite of the recency effect discussed in the Trial Heat Question Wording section), which is a cognitive bias leading people to select options presented near the top of a list when the list is presented visually, as on a ballot.

Jon Krosnick of Stanford University wrote an op-ed for the ABC News website explaining how a ballot order effect may have influenced the results of the New Hampshire Democratic primary (2008) and contributed significantly to discrepancies between the pre-election surveys and the election outcome in New Hampshire. Krosnick noted that, unlike previous primaries in the state, the 2008 contest featured the same ordering of candidate names on all ballots: an alphabetical list starting with the randomly drawn letter Z. Consequently, Joe Biden was listed first on every ballot, closely followed by Hillary Clinton, and Barack Obama was listed near the bottom of 21 candidate names. Krosnick estimates that Clinton received at least a three percentage point boost from her position near the top of the ballot order.

In other early primaries, ballot order does not explain differences between the election outcome and the pre-election polls. One reason for this may be that the list of candidates on the New Hampshire ballot was more than twice as long as the list in other primaries (21 names versus 8 names). Recency effects are generally thought to increase in size with list length. Ballot order rules, candidate name orderings, and election outcomes are presented in Table 25.

³⁴ See the University of New Hampshire press release for this result: http://www.unh.edu/survey-center/news/pdf/primary2008_demprim10708.pdf.

In South Carolina and Wisconsin, Clinton was placed near the top of the list and Obama was listed near the bottom. Had there been a strong ballot order effect, we would expect Clinton to have done better on Election Day in these states than in the polls. In fact, the reverse occurred. Obama's actual margin of victory in South Carolina and Wisconsin was substantially greater than the margin suggested by the polls. This result does not rule out the possibility that Clinton benefitted from her higher ballot position in these states, but it suggests that any such effect was swamped by other factors. In California, the ordering was randomized and then rotated across districts, so there is no reason to believe that ballot order affected Election Day support for Clinton or Obama in that primary.

The ballot order analysis presented here is purely observational and therefore may be limited in its generalizability. We do not find evidence for a ballot order effect for the top two candidates in the South Carolina, Wisconsin, or California primaries, but this does not rule out the possibility of such an effect in New Hampshire, especially given the much longer list of candidates on the New Hampshire ballot. Unfortunately, we have no way to test directly for a ballot order effect in New Hampshire. We can only infer the consequences, as Krosnick does, from similar elections in which ballot order was varied across precincts. Absent more compelling data, we find no reason to discount Krosnick's hypothesis.

Summary: Ballot order is one of the possible explanations for some of the estimation errors, and it may explain some of the error in the New Hampshire Democratic primary polls. Krosnick's analysis of recent New Hampshire primaries suggests a 3 percentage point effect from ballot order (2008). Clinton was listed near the top and Obama near the bottom on every ballot, which is consistent with greater support for Clinton in the returns than in the pre-election polls. This conclusion is based upon an observation and not experimental evidence to evaluate ballot order effects.

Table 25. Ballot Order, Estimates, and Actual Result of the NH Democratic Primary

State	Ballot order rule*	Order of names	Election Result	Mean of Final Poll Estimates
Iowa	N/A (caucus)	N/A (caucus)	Obama +9	Obama +0
New Hampshire	Alphabetically starting with randomly drawn letter	Biden Caligiuri Capalbo Clinton Crow Dodd Edwards Gravel Hewes Hughes Hunter Keefe Killeen Koon Kucinich LaMagna Laughlin Obama Richardson Savior Skok	Clinton +3	Obama +7
South Carolina	Alphabetically	Biden Clinton Dodd Edwards Gravel Kucinich Obama Richardson	Obama +28	Obama +11
California	Randomized and rotated across districts	(8 names)	Clinton +9	Clinton +6
Wisconsin	Randomized	Kucinich Clinton Biden Gravel Dodd Obama	Obama +17	Obama +7

* In NH, SC, and WI, the same ballot order was used state-wide.

Sources: <http://www.sos.nh.gov/rsa656.htm>; <http://www.sos.nh.gov/presprim2008/dpresbelk.htm>; <http://www.scstatehouse.gov/code/t07c013.htm>; http://www.sos.ca.gov/admin/press-releases/prior/2007/DB07_084.pdf; <http://elections.state.wi.us/docview.asp?docid=13359&locid=47>

Conclusions

The committee evaluated a series of hypotheses that could be tested empirically, employing information at the level of the state, the poll, and, in limited cases, the respondent. Since the analysis was conducted after data collection, it was not possible to evaluate all of the hypotheses in a way that permitted strong causal inferences. And given the incomplete nature of the data for various measures, it was not possible to pursue all hypotheses about what might have happened, nor was it possible to pursue multivariate analyses that looked simultaneously at multiple explanatory factors. In the end, however, the analysis suggests potential explanations for the estimation errors and the unlikely impact of other factors. The research also highlights the need for additional disclosure requirements and the need for better education by professional associations like AAPOR, the Council of American Survey Research Organizations (CASRO), and the National Council on Public Polls (NCPP).

Polling in primary elections is inherently more difficult than polling in a general election. Usually there are more candidates in a contested primary than in a general election, and this is especially true at the beginning of the presidential selection process. For example, there were a total of 15 candidates entered in the Iowa caucuses and more than 20 names on the New Hampshire primary ballot. Since primaries are within-party events, the voters do not have the cue of party identification to rely on in making their choice. Uncertainty among voters can create additional problems for pollsters. Turnout is usually much lower in primaries than in general elections, although it varies widely across events. Turnout in the Iowa caucuses tends to be relatively low compared to the New Hampshire primary, for example. So estimating the likely electorate is often more difficult in primaries than in the general election. Furthermore, the rules of eligibility to vote in the primaries vary from state to state and even within party; New Hampshire has an open primary in which independents can make a choice at the last minute in which one to vote. All of these factors can contribute to variations in turnout, which in turn may have an effect on the candidate preference distribution among voters in a primary compared to the general election.

The estimation errors in the polls before the New Hampshire Democratic primary were of about the same magnitude, as measured by the statistic A , as in the Iowa caucus. But the misestimation problems in New Hampshire received much more – and more negative – coverage than they did in Iowa. Because of a small level of undecided voters in every poll, the estimates for each individual candidate were generally lower than the proportion of votes they received. And these underestimates tended to be greater for the first place finisher than the second place finisher. But the majority of the polls before New Hampshire suggested the wrong winner, while only half in Iowa did.

All of the committee's conclusions are summarized briefly in Table 26. Factors that may have influenced the estimation errors in the New Hampshire pre-primary polls include:

- Respondents who required more effort to contact seemed more likely to support Senator Clinton, but most interviews were conducted on the first or second call, favoring Senator Obama.
- Patterns of nonresponse, derived from comparing the characteristics of the pre-election samples with the exit poll samples, suggest that some groups that supported Senator Hillary Clinton were underrepresented in the pre-election polls.
- Variations in likely voter models could explain some of the estimation problems in individual polls. Application of the Gallup likely voter model, for example, produced a larger error than their unadjusted data. While the "time of decision" data do not look very different in 2008 compared to recent presidential primaries, about one-fifth of the voters in the 2008 New Hampshire primary said they were voting for the first time. This influx of first-time voters may have had an adverse effect on likely voter models.
- Variations in weighting procedures could explain some of the estimation problems in individual polls. And for some polls, the weighting and likely voter modeling were comingled in a way that makes it impossible to distinguish their separate effects.
- Although no significant social desirability effects were found that systematically produced an overestimate of support for Senator Obama among white respondents or for Senator Clinton among male respondents, an interaction effect between the race of the interviewer and the race of the respondent did seem to produce higher support for Senator Obama in the case of a black interviewer. However, Obama was also preferred over Clinton by those who were interviewed by a white interviewer.

Factors unlikely to have contributed to the estimation errors in New Hampshire include:

- The exclusion of cell phone only (CPO) individuals from the samples did not seem to have an effect. However, this proportion of citizens is going to change over time, and pollsters should remain attentive to its possible future effects.
- The use of a two-part trial heat question, intended to reduce the level of "undecided" responses, did not produce that desired effect and does not seem to have affected the eventual distributions of candidate preference.

- The use of either computerized telephone interviewing (CATI) techniques or interactive voice response (IVR) techniques made no difference to the accuracy of estimates.
- The use of the trial heat questions was quite variable, especially with regard to question order, but no discernible patterns of effects on candidate preference distributions were noted. While the names of the (main) candidates were frequently randomized, the committee did not receive data that would have permitted an analysis of the impact of order.
- Little compelling information indicates that Independents made a late decision to vote in the New Hampshire Republican primary, thereby increasing estimate errors.

Factors that present intriguing potential explanations for the estimation errors in the New Hampshire polls, but for which the committee lacked adequate empirical information to thoroughly assess include:

- The wide variation in sample frames used to design and implement samples – ranging from random samples of listed telephone numbers, to lists of registered voters with telephone numbers attached, to lists of party members – may have had an effect. Greater disclosure about sample frames and sample designs, including respondent selection techniques, would facilitate future evaluations of poll performance.
- Differences among polls in techniques employed to exclude data collected from some respondents could have affected estimates. Given the lack of detailed disclosure of how this was done, it is not possible to assess the impact of this procedure.

Finally, factors that appeared to be potential explanations for estimation errors, but for which the committee lacked any empirical information to assess include:

- Because of attempts by some states to manipulate the calendar of primaries and caucuses, the Iowa and New Hampshire events were rescheduled to the first half of January, with only five days between the events, truncating the polling field period in New Hampshire following the Iowa caucus.
- Given the calendar, polling before the New Hampshire primary may have ended too early to capture late shifts in the electorate there, measuring momentum as citizens responded to the Obama victory in the Iowa caucus but not to later events in New Hampshire such as the restaurant interview with Senator Hillary Clinton.

- The order of the names on the ballot – randomly assigned but fixed on every ballot - may have contributed to the increased support that Senator Hillary Clinton received in New Hampshire.

All of the information provided to the committee is being deposited in the Roper Center Data Archive, where it will be available to other analysts who wish to check on the work of the committee or to pursue their own independent analysis of the pre-primary polls in the 2008 campaign.

Table 26. Summary of Empirical Evidence for Explanations of the 2008 New Hampshire Democratic Pre-Primary Election Poll Errors

Explanation	Support for the explanation	Opposition to the explanation	Conclusion
Non-coverage of cell phone only voters	Conflicting, non-significant findings that excluding cell phone only voters influenced primary trial heat estimates	Jones (2008), Keeter (2008), and ZuWallack et al. (2008) found minimal effects on primary trial heat estimates from excluding cell phones.	There is no compelling reason to suggest this contributed to the errors in any meaningful way.
Nonresponse error that was not corrected by weighting	The pre-election polls underestimated the size of two pro-Clinton groups: union households and non-college graduates. Also, CBS and Gallup respondents who required 3 or more call attempts tended to be more supportive of Clinton than those easier to reach.	The pre-election poll samples are similar to weighted exit poll results in terms of gender, race, marital status, and party identification.	This may have contributed to the error in substantial ways, but the evidence is inconclusive.
Recency effect in poll question wording	Research 2000 did not randomize the order of the candidate names and consistently listed Clinton after Obama. Clinton's estimated vote share was somewhat larger (and more accurate) in the Research 2000 poll than in other NH polls. Zogby also did not randomize and consistently listed Clinton before Obama. Clinton's estimated vote share as lower (and less accurate) in the Zogby poll than in other NH polls.	Most NH polls randomized the order of candidate names.	There is no compelling reason to suggest this contributed to the errors in any meaningful way.
Forced-choice questions yielded inaccurate data	The mean percentage of undecided voters was 6% for polls using a one-step force and 7% for polls not using it.	The mean absolute value of A was the same for polls that used a forced choice question and polls that did not	There is no compelling reason to suggest this contributed to the errors in any meaningful way.
Faulty likely voter model	Gallup concluded that misspecification of their likely voter model was the single biggest reasons for the error in their NH poll.	The likely voter model can explain at most only half of the error in the Gallup estimates. In the CBS survey, the likely voter model appeared to have no effect on the accuracy of the estimates.	There is compelling evidence that this explains much of the error in the Gallup poll, but no evidence that it explains errors in the other polls.
Polls ended data collection too early to detect a late shift in preferences	Clinton's emotional moment at the NH diner occurred after nearly all the polls had finished, so they would have missed any shift in preferences triggered by this event. Some 38% of NH Dem primary voters said they made up their mind during the final 3 days, according to the exit poll. The Gallup panel study indicates a detectable shift in preference toward Clinton in the final days.	The percentage of voters making up their minds on election day is not radically different from historical levels.	There is compelling evidence that this was an important contributor to the error.
Social desirability ("Bradley effect")	In pooled analysis of three NH surveys, we find support for Obama is significantly greater when the interviewer is black than when he or she is white.	In the pooled analysis, Obama still has the lead among respondents interviewed by a white interviewer, so this cannot explain all of the error in the polls. A meta-analysis (Hopkins 2008) finds no support for racially-motivated misreporting in contemporary election surveys. Langer (2008a) finds no difference in support for Obama recorded by black or white interviewers in the general election.	The evidence is mixed. This may explain a small proportion of the error but probably no more than that.

Faulty weighting algorithm	In three of the four surveys suitable for examination, the weighted NH estimates were less accurate than the unweighted estimates.	All four unweighted estimates still had Obama in the lead, so weighting cannot explain all of the error in the polls.	There is good evidence to suggest that faulty weighting algorithms explains some (but not all) of the error.
Independents made late decision to vote in Republican primary	In the exit poll, 37% of Republican primary voters were Independent. In the pre-election surveys, this figure was lower.	The proportion of Independent voters in the Democratic primary pre-election polls is comparable to the proportion in the exit poll. Also, the differences between the exit poll and the pre-election surveys are not large enough to explain the errors in the polls.	There is little compelling reason to suggest this contributed to the errors in any meaningful way.
Ballot order effect	Analysis of recent NH primaries suggests a 3-point effect from ballot order (Krosnick 2008). Clinton was listed near the top and Obama near the bottom on every ballot, which is consistent with greater support for Clinton in the returns than in the pre-election polls.	This explanation cannot account for the severity of mis-estimation in several of the NH polls.	This is one of the most likely explanations.

References

- Blumberg, Stephen J. and Julian V. Luke. 2008. "Wireless Substitution: Early Release of Estimates from the National Health Interviews Survey, July-December 2007." Web article posted for the National Center for Health Statistic at <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200805.htm>.
- Crespi, Irving. 1988. *Pre-election Polling: Sources of Accuracy and Error*. New York: Russell Sage Foundation.
- Curtin, Richard, Stanley Presser, and Eleanor Singer. 2000. "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly* 64:413-28.
- DeSart, Jay and Thomas Holbrook. 2003. "Campaigns, Polls, and the States: Assessing the Accuracy of Statewide Presidential Trial-Heat Polls." *Political Science Quarterly*, 56(4): 431-439.
- Erikson, Robert S., Costas Panagopoulos, and Christopher Wlezien. 2004. "Likely (and Unlikely) Voters and the Assessment of Campaign Dynamics." *Public Opinion Quarterly* 68(4):588-601.
- Erikson, Robert S. and Christopher Wlezien. 2008. "Likely Voter Screens and the Clinton Surprise in New Hampshire." Web article posted for Pollster.com at http://www.pollster.com/blogs/likely_voter_screens_and_the_c.php.
- Frankovic, Kathleen. 2008. "N.H. Polls: What Went Wrong?" Web article posted for CBS News at <http://www.cbsnews.com/stories/2008/01/14/opinion/pollpositions/main3709095.shtml?source=RSS&attr=3709095>.
- Groves, Robert M. 2006 "Non-response Rates and Non-response Bias in Household Surveys." *Public Opinion Quarterly* 70:646-675.
- Ho, Daniel E. and Kosuke Imai. 2008. "Estimating Causal Effects of Ballot Order from a Randomized Natural Experiment: The California Alphabet Lottery, 1978 – 2002." *Public Opinion Quarterly*, 72:216-240.
- Holbrook, Allison, Jon A. Krosnick, David Moore, and Roger Tourangeau. 2007. "Response Order Effects in Dichotomous Categorical Questions Presented Orally: The Impact of Question and Respondent Attributes." *Public Opinion Quarterly* 71:325-348.
- Hopkins, Daniel J. 2008. "No More Wilder Effect, Never a Whitman Effect: When and Why Polls Mislead about Black and Female Candidates." Poster presented at the Meeting of the Society for Political Methodology.
- Jones, Jeff. 2008. "Cell Phones in Primary Pre-Election Surveys." Paper presented at the Annual Conference of the American Association for Public Opinion Research.
- Keeter, Scott. 2006. "The Impact of Cell Phone Noncoverage Bias in the 2004 Presidential Election." *Public Opinion Quarterly* 70:88-98.
- Keeter, Scott. 2008. "The Impact of 'Cell-Onlys' on Public Opinion Polling: Ways of Coping with a Growing Population Segment" Web article posted for the Pew Research Center at <http://people-press.org/report/391/>.
- Keeter, Scott, Michael Dimock, and Leah Christian. 2008. "Cell Phones and the 2008 Vote: An Update" Web article posted at <http://pewresearch.org/pubs/964/>.
- Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert M. Groves, and Stanley Presser. 2000. "Consequences of Reducing Non-response in a Large National Telephone Survey." *Public Opinion Quarterly* 64:125-48.
- Keeter, Scott, Courtney Kennedy, Michael Dimock, Jonathan Best, and Peyton Craighill. 2006. "Gauging the Impact of Growing Non-response on Estimates from a National RDD Telephone Survey." *Public Opinion Quarterly* 70:759-779.

- Kohut, Andrew. 2008. "Getting It Wrong." *The New York Times*, January 10. Available at <http://www.nytimes.com/2008/01/10/opinion/10kohut.html?ref=opinion>.
- Krosnick, Jon A. 2008. "Clinton's Favorable Placement on Ballots May Account for Part of Poll Mistakes." Web article posted for ABC News at <http://abcnews.go.com/PollingUnit/Decision2008/story?id=4107883>.
- Krosnick, Jon A. and Duane Alwin. 1987. "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement." *Public Opinion Quarterly* 51:210-219.
- Langer, Gary. 2008a. "Dissecting the 'Bradley Effect'." Web article posted for ABC News at <http://blogs.abcnews.com/thenumbers/2008/10/the-bradley-eff.html>.
- Langer, Gary. 2008b. "A New Hampshire Post-Mortem." Web article posted for ABC News at <http://blogs.abcnews.com/thenumbers/2008/02/a-new-hampshire.html>.
- Langer, Gary. 2008c. "Cell-Onlies: Report on a Test." Web article posted for ABC News at <http://blogs.abcnews.com/thenumbers/2008/09/cell-onlies-rep.html>.
- Lau, Richard R. 1994. "An Analysis of 'Trial-Heat' Polls during the 1992 Presidential Election." *Public Opinion Quarterly* 62, 2-20.
- Martin, Elizabeth A., Michael W. Traugott, and Courtney Kennedy. 2005. "A Review and Proposal for a New Measure of Poll Accuracy." *Public Opinion Quarterly* 69: 342-369.
- Merkle, Daniel, and Murray Edelman. 2002. "Non-response in Exit Polls: A Comprehensive Analysis." In *Survey Non-response*, ed. Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, pp. 243-58. New York: Wiley.
- Mosteller, Frederick. 1948. *The Pre-Election Polls of 1948*. New York: Social Science Research Council.
- Miller, Joanne. M., and Jon A. Krosnick. 1998. "The Impact of Candidate Name Order on Election Outcomes." *Public Opinion Quarterly* 62, 291-330.
- Mitofsky, Warren J. 1998. "Was 1996 a Worse Year for Polls than 1948?" *Public Opinion Quarterly*, 62: 230-249
- Nichols, John. 2008. "Did 'The Bradley Effect' Beat Obama in New Hampshire?" Web article posted for The Nation and available at http://www.thenation.com/blogs/state_of_change/268328.
- Robinson, Eugene. 2008. "Echoes of Tom Bradley." *The Washington Post* January 11, A17.
- Traugott, Michael W. Robert M. Groves, James M. Lepkowski. 1987. "Using Dual Frame Designs to Reduce Non-response in Telephone Surveys." *Public Opinion Quarterly* 51: 522-539.
- Traugott, Michael W., Brian Krenz, and Colleen McClain. 2008. "Press Coverage of the Polling Surprises in the New Hampshire Primary." Paper presented at the Annual Conference of the Midwest Association for Public Opinion Research.
- Traugott, Michael W. and Vincent Price. 1992. "Exit Polls in the 1989 Virginia Gubernatorial Race: Where Did They Go Wrong?" *Public Opinion Quarterly* 52: 245-253.
- Traugott, Michael W. and Clyde Tucker. 1984. "Strategies for Predicting Whether a Citizen Will Vote and Estimation of Electoral Outcomes." *Public Opinion Quarterly* 48: 330-343.
- ZuWallack, Randal, Jeri Piehl, and Keating Holland. 2008. "Supplementing a National Poll with Cell Phone Only Respondents." Paper presented at the Annual Conference of the American Association for Public Opinion Research.

Appendix A

Members of the AAPOR Special Committee on 2008 Presidential Primary Polling

Glen Bolger, a partner and co-founder of Public Opinion Strategies.

Darren W. Davis, Professor of Political Science at the University of Notre Dame.

Charles Franklin, Professor of Political Science at the University of Wisconsin and co-developer of Pollster.com.

Robert M. Groves, Director, the University of Michigan Survey Research Center, Professor of Sociology at the University of Michigan, Research Professor at its Institute for Social Research, and Research Professor at the Joint Program in Survey Methodology at the University of Maryland.

Paul J. Lavrakas, a methodological research consultant.

Mark S. Mellman, CEO of The Mellman Group.

Philip Meyer, Professor Emeritus in Journalism at the University of North Carolina.

Kristen Olson, Assistant Professor of Survey Research and Methodology and Assistant Professor of Sociology at the University of Nebraska-Lincoln.

J. Ann Selzer, President of Selzer & Company.

(Chair) Michael W. Traugott, Professor of Communication Studies and Senior Research Scientist in the Center for Political Studies at the Institute for Social Research at the University of Michigan.

Christopher Wlezien, Professor of Political Science and Faculty Affiliate in the Institute for Public Affairs at Temple University.

Appendix B

Charge to the Committee

1. Examine the available information concerning the conduct and analysis of the pre-election, entrance and exit polls related to the 2008 primaries, including, but not limited to, press releases, post-election hypotheses, and evaluations conducted by the respective polling organizations as well as the news media.
2. To the extent possible, interview those involved in the New Hampshire primary pre-election and exit polls to gather additional information for the committee.
3. Synthesize and report on the findings from the various polling organizations. The report will include a summary of the factors that have been considered to date as well as recommendations and guidelines for future research. The report is tentatively scheduled for release in early April, 2008.
4. Present the findings from the report in a public forum, hosted by the Kaiser Family Foundation at its Barbara Jordan Conference Center in Washington, D.C. The forum is tentatively scheduled for Spring, 2008.
5. To facilitate research on all of the possible factors that may have contributed to the New Hampshire polling process this year, request all sample, survey and ancillary data associated with the polls leading up to and following the New Hampshire primary, including the New Hampshire exit poll data. The request will be broad in nature – for example, so as to inform hypotheses concerning nonresponse, sample and call-record data will be included.
6. The Roper Center has generously offered to serve as the archivist for the data associated with the ad hoc committee. The Roper Center is keenly sensitive to the risks associated with these datafiles and the potential for exposure of confidential information. These data will be archived and maintained separately from the general access archives in a secure environment. Scholars interested in analyzing these restricted datasets will complete an Application for Restricted Data Use. Approval of the application will be made if the research purpose meets the criteria outlined by the ad-hoc committee. Further limitations to the researcher will include the destruction of the data files after a designated period of time, as outlined in the application.
7. Establish seed-funding for support of additional research on the New Hampshire and other primary pre-election and exit poll data. This may include support of the work of the ad hoc committee to undertake analysis or the work of individual scholars interested in conducting research on the topic.
8. Beyond the public forum in Spring, 2008, the findings of the ad hoc committee will be disseminated on the AAPOR web site and as part of a special panel discussion at the AAPOR 63rd Annual Conference.

Letter to Survey Firms from President Mathiowetz

I am writing to you as President of the American Association for Public Opinion Research (AAPOR) with regard to polling you have conducted in [STATE] as part of the 2008 presidential election campaign. As you may be aware, AAPOR has named an Ad Hoc Committee on 2008 Presidential Primary Polling. The task of that committee is to evaluate the methodology of the pre-election primary polls and the way they are being reported by the media and used to characterize the contests in pre-election press coverage. Although originally formed in response to the disparity between the pre-election polls and the outcome of the Democratic contest in New Hampshire, the mission statement of the committee has been expanded to include examination and archiving of primary data conducted in states other than New Hampshire.

The variation between the pre-election polls and the final outcome of the elections in, for example, New Hampshire (Democrats), South Carolina (Democrats), and California (Republicans), has raised questions concerning the profession as a whole as a reflection of the quality of estimates of candidate standing in those contests. The horserace aspect of polling, albeit only a small part of the work our profession does, offers an immediate and visible validation of survey estimates. In this way, the image of the entire industry is affected by the quality of the estimates that are made at the end of political campaigns. We are a profession that benefits from our collective understanding of the sources of errors that impact our estimates. After the 1948 Presidential election, the pollsters involved in the pre-election polling undertook an examination and analysis of the factors that contributed to the miscalling of that election. It is in that spirit, and because of the collective knowledge that will come from this work, that I ask for your cooperation with the request outlined below.

The AAPOR Code of Professional Ethics and Practices as well as the Principles of Disclosure of the National Council of Public Polls (NCPP) and the Code of Standards and Ethics of the Council of American Survey Research Organizations (CASRO) all call for the disclosure of essential information about how research was conducted when the results are widely disseminated to the public. At a minimum, we ask for disclosure of the essential information outlined by these codes. However, you will see that the request does go beyond the disclosure guidelines outlined in these respective codes. This information will be critical for the AAPOR Committee to pursue its evaluation.

The Ad Hoc Committee will focus on addressing empirically-based hypotheses that can be addressed post hoc—for example, whether differences in likely voter screening, turnout models, differential non-response, the allocation of undecideds, weighting procedures, and other sources of measurement error could have contributed to these estimation errors. To address these issues, the request outlined in the attached document is broad-based, ranging from full questionnaires to individual-level data to documentation of procedures. The committee is interested in obtaining information from every firm or research center that collected data prior to these elections and caucuses.

The Roper Center has offered to serve as the archivist for the data associated with the work of the committee because we expect that when the Ad Hoc Committee issues its report, others will be interested in examining how the committee came to the conclusions it did. The Roper Center is keenly sensitive to the risks associated with these data files and the potential for exposure of confidential information. In the short run, access will be limited to the committee (and research assistants working with the committee members). In the long run, the goal is to provide access to the data for other scholars. The Roper Center will work with individual pollsters to determine which files may eventually be made available to the broader community of scholars (for example, after an embargo period). Scholars interested in analyzing these restricted datasets will complete an Application for Restricted Data Use; review of these applications will be completed by a joint committee of the Roper Center and a subgroup of the Ad Hoc Committee.

I look forward to working with you in the weeks to come. The issues we face as a profession are challenging; I hope the work of the committee sheds light on issues that may have been unique to the 2008 pre-election primary polls as well as those issues that will inform and improve the methodology of our industry in the years to come. If you have any questions or would like any additional information about AAPOR or the Ad Hoc Committee, please feel free to contact me.

Regards,

Nancy A. Mathiowetz
President, American Association for Public Opinion Research

Appendix C

AAPOR Special Committee on 2008 Presidential Primary Polls Information Request

The request for information and data has been separated into two parts, (1) information that is part of the AAPOR Standards for Minimal Disclosure and (2) information or data that goes beyond the minimal disclosure requirements. The information you provide could take one or more forms—documentation, tables, and individual-level data. Our mission is to be able to examine data that will empirically inform questions concerning the disjuncture between the pre-election primary polls and the election outcomes. Members of the committee are willing to work with you in order to obtain the information of interest, in whatever form is easiest for you to provide.

AAPOR Standards for Minimal Disclosure

As noted in the AAPOR code of professional ethics and practices, the following items should, at a minimum be disclosed for all public polls:

1. Who sponsored the survey, and who conducted it.
2. The exact wording of questions asked, including the text of any preceding instruction or explanation to the interviewer or respondents that might reasonably be expected to affect the response.

Here the committee would request that you provide the complete questionnaire—hard copy, executable, screen shots—in whatever format is feasible. This should indicate which question or questions are used as the likely voter screening and which question or questions are used for the trial heat. The questionnaire should indicate any randomization of questions or response options

3. A definition of the population under study, and a description of the sampling frame used to identify this population.

The description of the sampling frame should indicate whether or not the frame includes cell phones.

4. A description of the sample design, giving a clear indication of the method by which the respondents were selected by the researcher, or whether the respondents were entirely self-selected.

For those studies that include cell phones in the sample frame, the description should include the sample selection for these numbers.

5. Sample sizes and, where appropriate, eligibility criteria, screening procedures, and response rates computed according to AAPOR Standard Definitions. At a minimum, a summary of disposition of sample cases should be provided so that response rates could be computed.

In addition to sample sizes, demographic information on those screened out would be useful in distinguishing between nonresponse bias and turnout model bias.

6. A discussion of the precision of the findings, including estimates of sampling error, and a description of any weighting or estimating procedures used.

In addition to the standard weighting information, documentation on the weighting procedure used to produce final turnout estimate

7. Which results are based on parts of the sample, rather than on the total sample, and the size of such parts.
8. Method, location, and dates of data collection.

If interviewers are used for the data collection, please indicate if live interviewers or IVR technology was used.

Beyond Minimum Disclosure

To address the various hypotheses concerning the New Hampshire, South Carolina, California, and Wisconsin pre-election polls, the committee is requesting information and data beyond the minimum disclosure outlined above. For several of the items, the request is for the raw data; if the data are not available or can not be made available, the committee would benefit from the provision of the analysis tables as described. In cases where no written documentation is available, one or more of the committee members would be willing to discuss the procedures with you or one of your staff members.

I. Data

1. Individual-level data for all individuals contacted and interviewed, including those who failed the likely-voter screening and including all weights used in the production of the final estimates prior to the election, date of the interview, and interviewer identification number
 - A. In lieu of the individual-level data, demographic information on those who were screened out along with crosstabulations between vote preference and likely/ unlikely voters as well as registered/unregistered voters
 - B. Final estimate of the demographic composition of the turnout
 - C. Share of the voting age population represented by the turnout estimate, within demographic subgroup
 - D. Tabulations that reflect the combination of likelihood of voting and candidate preference, indicating how the trial heat question might differ by different estimates of turnout
 - E. If date of interview is not included on the individual level data file, distribution of candidate preferences (the trial heat question) by date of interview

- F. In lieu of individual-level data, crosstabulations of voter preference by demographic characteristic, within subgroups. We are especially interested in the distributions of candidate preference by age, race, sex, party identification, political ideology, and religiosity
- 2. Reinterview data, if a post-election re-interview was conducted
- 3. To examine hypotheses related to social desirability, it would be beneficial to be able to examine characteristics of interviewers, matched to respondents. Listed in (I.1) above was a request for an interviewer ID as part of the individual file. Ideally, we would like to be able to link individual level records to characteristics of interviewers
- 4. To examine hypotheses related to nonresponse bias, the committee would need to have access to data for the full sample, including call record information, disposition of each sampled number, including attempts at recontacts

II. Documentation

- 1. Interviewer documentation, including instructions not included in the text of the questionnaire (e.g., instructions for probing “Don’t Know” responses).
- 2. Allocation rules for Don’t Knows and Undecideds
- 3. Documentation of the rules, if any, for sample allocation to interviewers
- 4. Documentation of approach to handling “early voting” in the composition of the final sample
- 5. The last press release or releases associate with your poll most proximal to the election..

The Roper Center for Public Opinion Research
University of Connecticut

369 Fairfield Road, U-2164
Storrs, CT 06269-2164

Tel. 860-486-4440
Fax 860-486-6308

Guidelines for Depositing Data
AAPOR Special Committee on 2008 Presidential Primary Polling

Please complete this form to provide the archives with information about the substantive and technical characteristics of your survey data collection. Completing the form ensures that the survey collection will be accurately and thoroughly described for use by future researchers. It is important that the information provided on the form be accurate. Information on the following form allows Roper Center staff to prepare descriptive materials for various searchable databases that the Center maintains. This information also permits the creation of citations researchers may use when working with your surveys. Providing proper citation to the original data producers is vital to maintaining the integrity of the Center archives.

In practical terms, each study should include a data file, exact questions posed in the form of a questionnaire or codebook, and any other supporting documentation (open-ended codes, technical reports, weighting information, summary and final reports, etc.). Please contact the Center if you have questions about what should be included.

Finally, the data deposit form grants permission for the Roper Center for Public Opinion Research to archive and distribute your data collection. Once data are archived with the Center, they will remain a permanent part of the archives and may not be de-accessioned.

Names of files:

1. Descriptive Title of Survey:

2. Name of Principal Investigator/Survey Organization responsible for Data Collection:

3. Name(s) of Sponsors:

4. Sample type (include geographic coverage):

- ❖ Sample Discription _____
- ❖ Size of Sample _____ Target Sample Size: _____
- ❖ Selection Procedure/Screening_____
- ❖ Estimated Sampling Error _____

5. Interview Dates: _____
6. Interview Method: _____
(i.e.: Telephone, In Person/Face to Face, Mail, other, ...)
7. Are these data weighted? If yes, what is the location of the weight variable(s) in the data file?
8. Are there restrictions that are to be placed on the availability of the data? If yes, please specify.
9. Completion rate and the method used to derive this rate:

Contact person and phone number/e-mail by which the data was received:

Name: _____

Title: _____

Organization: _____

E-mail: _____

Telephone: () _____ - _____ Fax: () _____ - _____

This collection of data complete with ***explicit identifying information*** will be made available only for scholarly research purposes as ***approved*** by the committee of the American Association for Public Opinion Research (AAPOR). I give permission for this data collection to be redisseminated by AAPOR and the Roper Center for Public Opinion Research under these restrictions.

I understand that in preparing this data collection ***for public archiving and distribution***, the Roper Center will remove all information directly identifying the survey respondents in these data, and will implement due diligence in preventing information in the collection from being used to disclose the identity of respondents. I give permission for this data collection to be redisseminated by the Roper Center for Public Opinion Research under its standard terms of use.

Signature: _____ Date: _____

Title: _____

Organization: _____ Phone: _____

Address: _____ Email: _____

E-mail: _____ Telephone: () _____ - _____

*On behalf of the wide range of social scientists and other researchers that the Roper Center serves,
thank you for preserving these important data.*

Appendix D

Hypotheses for Sources of Error in 2008 Primary Pre-election Surveys

1. Likely voter screening
 - 1a. Likely voter screening questions used in general elections do not work as well in (the 2008) primary elections.
 - 1b. Likely voter screening questions used in primaries do not work as well in unusually high turnout primaries.
2. Turnout models/Turnout surge
 - 2a. Because of the calendar and the proximity of events, voter interest is stimulated from one event to another.
 - 2b. The higher turnout of African Americans was underestimated on the Democratic side.
 - 2c. The higher turnout of women was underestimated on the Democratic side
3. Inability to capture last-minute deciders or changes in preference
 - 3a. The nature of the contests (in 2008) means that many voters are making up their minds late and deciding to go to the polls (combination of turnout and decisions)
 - 3b. Voters are changing their minds late; turnout estimates are all right but preferences change.
4. Misreporting issues
 - 4a. Voters are misreporting preferences to interviewers (social desirability) because they are unwilling to say they won't vote for an African American or a female candidate
 - 4b. Respondents are misreporting their intention to vote (staying home or going to the polls)
 - 4c. Misreporting is greater for "non-traditional" (non-white, non-male) candidates than for traditional candidates
5. Nonresponse bias
 - 5a. The short time between events and the need to take previous results into account in measurement means that response rates are low, and nonrespondents are different from respondents.
 - 5b. Differential nonresponse by key groups in the electorate affects the distribution of candidate preferences (African Americans, Whites, men, women).
6. Question wording effects
 - 6a. Differences in the way that the "trial heat" question is asked account for differences in polls
 - 6b. Differences in the way that the trial heat question is asked (e.g., explicit or implicit Don't Know alternatives) produce different levels of Don't Know or Undecided responses
 - 6c. The ordering of the candidates in the trial heat question affects the distribution of responses (i.e., primacy or recency effects)
7. Question order effects
 - 7a. The preceding questions affect responses to the "trial heat" question but do not have an impact in the voting booth
8. Allocation of undecideds
 - 8a. Some results are being reported with undecideds included while others are being reported with them excluded
 - 8b. Some results are being reported with the undecideds allocated (by different methods) while others are not

9. Sampling Issues

- 9a. Some groups are being oversampled, others under sampled
- 9b. In states where there are open primaries, Independents or Not identified people are being oversampled or weighted
- 9c. Weighting algorithms are not working as well in the primaries as in a general election.
- 9d. A large proportion of votes are being cast absentee or by other “early” procedures and they are not being captured in the survey or they are mis-weighted proportionately when combined with those who intend to vote in person on Election Day.
- 9e. Differential loss of cell phone only voters by state in certain primaries?

10. External Factors

- 10a. The rules of voting in the primaries are not adequately captured in the polling methodology (who is eligible to vote in a specific event)
- 10b. The order of the names on the ballot has an independent effect on the outcome (or in relation to the order of the names on in the “trial heat” question)
- 10c. Senator Clinton’s emotional response swayed voters at the last minute
- 10d. President Clinton’s sharp criticisms of Obama swayed African American voters at the last minute

Appendix Table 1. Accuracy of Final 2008 Primary Pre-election Polls by Mode of Data Collection*

	Mode	A [Ob/Cl]	% Obama	% Clinton		Mode	A [1st/2nd]	% McCain	% Romney	% Huck abee
Iowa Democratic Caucuses			38	29	Iowa Republican Caucuses			13	25	34
Zogby	CATI	-0.01	31	24	ABC/W.P.	CATI	-0.05	6	27	35
Selzer & Co.	CATI	-0.02	32	25	Research 2000	CATI	-0.08	8	27	34
PSRAI	CATI	-0.08	35	29	Zogby	CATI	-0.09	10	25	31
Strategic Vision	CATI	-0.10	32	27	Selzer & Co.	CATI	-0.10	13	26	32
ABC/W.P.	CATI	-0.14	33	29	ARG	CATI	-0.12	11	24	29
Research 2000	CATI	-0.24	29	28	Financial Dyn.	CATI	0.14	5	23	36
Financial Dyn.	CATI	-0.27	27	27	LA Times	CATI	0.17	11	23	37
Mason-Dixon	CATI	-0.31	22	23	Rasmussen	IVR	-0.27	14	27	28
ORC	CATI	-0.33	31	33	Strategic Vision	CATI	-0.38	16	30	28
LA Times/Bloom.	CATI	-0.38	26	29	ORC	CATI	-0.41	10	31	28
Rasmussen	IVR	-0.41	27	31	Mason-Dixon	CATI	-0.47	13	27	23
Global Strategy	CATI	-0.48	22	27	PSRAI	CATI	0.52	6	17	39
ARG	CATI	-0.58	25	34						
Average (abs) for CATI polls		0.25			Average (abs) for CATI polls		0.23			
Average (abs) for IVR poll**		0.41			Average (abs) for IVR poll**		0.27			
New Hampshire Democratic Primary		37	39	New Hampshire Republican Primary		37	32	11		
Research 2000	CATI	0.10	34	33	Marist	CATI	-0.02	35	31	13
Mason-Dixon	CATI	0.13	33	31	Gallup	CATI	-0.02	34	30	13
LA Times	CATI	0.13	32	30	UNH	CATI	0.03	31	26	13
RKM	CATI	0.16	34	31	Research 2000	CATI	0.04	35	29	13
Op. Dynamics	CATI	0.20	32	28	Op. Dynamics	CATI	0.09	34	27	11
Suffolk	CATI	0.21	39	34	ARG	CATI	0.11	31	24	14
Marist	CATI	0.26	34	28	Strategic Vision	CATI	0.11	35	27	13
Rasmussen	IVR	0.28	37	30	Rasmussen	IVR	-0.11	32	31	10
CBS News	CATI	0.29	35	28	RKM	CATI	0.13	38	29	9
ARG	CATI	0.32	40	31	Zogby	CATI	0.14	36	27	10
UNH	CATI	0.33	39	30	Mason-Dixon	CATI	0.14	32	24	12
Strategic Vision	CATI	0.34	38	29	Suffolk	CATI	-0.29	26	30	13
Zogby	CATI	0.44	42	29	LA Times	CATI	-0.68	20	34	12
Gallup	CATI	0.45	41	28						
Average (abs) for CATI polls		0.26			Average (abs) for CATI polls		0.15			
Average (abs) for IVR poll**		0.28	28.00		Average (abs) for IVR poll**		0.11			
South Carolina Democratic Primary		55	27	South Carolina Republican Primary		33	15	30		
PPP	IVR	-0.11	44	24	Mason-Dixon	CATI	-0.02	27	15	25
Zogby	CATI	-0.26	41	26	SurveyUSA	IVR	0.04	31	17	27
Rasmussen	IVR	-0.28	43	28	Zogby	CATI	-0.06	27	16	26
SurveyUSA	IVR	-0.35	43	30	Rasmussen	IVR	-0.10	24	18	24
Ebony/Jet	CATI	-0.40	37	27	Clemson	CATI	0.18	29	13	22
Clemson	CATI	-0.41	27	20	Op. Dynamics	CATI	0.20	27	15	20
Mason-Dixon	CATI	-0.48	38	30	PPP	IVR	0.24	28	18	20
ARG	CATI	-0.63	39	36	ARG	CATI	-0.33	26	9	33
Average (abs) for CATI polls		0.43			Average (abs) for CATI polls		0.16			
Average (abs) for IVR polls		0.25			Average (abs) for IVR polls		0.13			
California Democratic Primary		43	52	California Republican Primary		42	35	12		
SurveyUSA	IVR	-0.04	42	52	Suffolk	CATI	0.02	39	32	8
Mason-Dixon	CATI	-0.05	36	45	Mason-Dixon	CATI	0.07	40	31	13
ARG	CATI	0.09	45	49	Gallup	CATI	0.08	35	27	12
Field	CATI	0.12	34	36	Field	CATI	0.11	32	24	13
Gallup	CATI	-0.12	35	47	SurveyUSA	IVR	-0.16	39	38	11
Datamar	IVR	-0.15	31	43	Rasmussen	IVR	-0.18	38	38	10
Rasmussen	IVR	0.20	45	44	ORC	CATI	0.22	39	26	11
Suffolk	CATI	0.20	40	39	ARG	CATI	-0.24	35	37	14
ORC	CATI	-0.25	32	49	Zogby	CATI	-0.37	33	40	12
PPIC	CATI	-0.25	28	43						
Zogby	CATI	0.48	49	36	Average (abs) for CATI polls		0.16			
Average (abs) for CATI polls		0.20			Average (abs) for IVR polls		0.17			
Average (abs) for IVR polls		0.13								
Wisconsin Democratic Primary		58	41	Wisconsin Republican Primary		51	43			
PPP	IVR	-0.07	53	40	ARG	CATI	0.00	51	43	
ARG	CATI	-0.14	52	42	PPP	IVR	0.08	50	39	
Research 2000	CATI	-0.24	47	42	Research 2000	CATI	0.23	48	32	
Strategic Vision	CATI	-0.26	45	41	Strategic Vision	CATI	0.34	45	27	
Rasmussen	IVR	-0.27	47	43						
Average (abs) for CATI polls		0.21			Average (abs) for CATI polls		0.19			
Average (abs) for IVR polls		0.17			Average (abs) for IVR poll**		0.08			

* Only polls collecting data within two weeks of election day are included in this analysis

** There was only one IVR poll conducted in the last two weeks of the Iowa caucuses (Dem and Rep), the New Hampshire primary (Dem and Rep), and the Wisconsin primary (Rep).

Appendix Table 2. Logistic Regression Predicting Vote Preference for Clinton in the 2008 New Hampshire Democratic Primary

	B	Std. Error	Wald Chi-Square	df	Sig.
(Intercept)	-4.38	0.97	20.43	1	0.000
White Interviewer	2.88	0.77	13.99	1	0.000
Respondent Education	0.71	0.23	9.60	1	0.002
Respondent Age	0.04	0.01	8.97	1	0.003
Male Respondent	-1.04	0.50	4.21	1	0.040
Male Interviewer	-0.20	0.12	3.08	1	0.079
Respondent is a Democrat	0.36	0.29	1.46	1	0.226
Survey	-0.35	0.37	0.92	1	0.338
Four Days Until Election	0.09	0.35	0.07	1	0.797
Three Days Until Election	0.14	0.34	0.18	1	0.675
Two Days Until Election	0.23	0.35	0.44	1	0.509
White Interviewer*Respondent Education	-0.48	0.14	11.66	1	0.001
Male Respondent*Respondent Age	0.03	0.01	11.72	1	0.001
Male Respondent*Respondent Education	-0.35	0.11	9.95	1	0.002
Respondent Age*Respondent Education	-0.01	0.00	5.32	1	0.021
White Interviewer*Respondent Age	-0.02	0.01	4.18	1	0.041

Model based on 1,455 New Hampshire poll respondents who intended to vote in the Democratic primary.

The first GLM estimates included all the main effects as well as all two-way interactions. The model shown above is a reduced model that includes all the main effects but only the interaction terms that were significant in the original model. This reduced model was preferred for the report because the full model required estimation of more than 30 parameters, which reduced the degrees of freedom.

A REVIEW AND PROPOSAL FOR A NEW MEASURE OF POLL ACCURACY

ELIZABETH A. MARTIN
MICHAEL W. TRAUGOTT
COURTNEY KENNEDY

Abstract This article proposes a new measure of the *predictive accuracy* (A) of election polls that permits examination of both accuracy and bias, and it applies the new measure to summarize the results of a number of preelection polls. We first briefly review past measures of accuracy, then introduce the new measure. After the new measure is described, the general strategy is to apply it to three presidential elections (1948, 1996, and 2000) and to compare the results derived from it to the results obtained with the Mosteller measures. Then, the new measure is applied to the results of 548 state polls from gubernatorial and senatorial races in the 2002 elections to illustrate its application to a large body of preelection polls conducted in "off-year" races with different outcomes. We believe that this new measure will be useful as a summary measure of accuracy in election forecasts. It is easily computed and summarized, and it can be used as a dependent variable in multivariate statistical analyses of the nature and extent of biases that affect election forecasts and to identify their potential sources. It is comparable across elections with different outcomes and among polls that vary in their treatment or numbers of undecided voters.

ELIZABETH A. MARTIN is the senior survey methodologist for the U.S. Bureau of the Census. MICHAEL W. TRAUGOTT is professor of communication studies and senior research scientist at the Institute for Social Research, University of Michigan. COURTNEY KENNEDY works at the Pew Research Center for the People and the Press and is a graduate student in the Joint Program in Survey Methodology. A previous version of this article was prepared for presentation at the annual meeting of the American Association for Public Opinion Research, Nashville, TN, May 16-19, 2003. This article reports the results of research and analysis undertaken collaboratively by Census Bureau staff and researchers from the University of Michigan. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau or the University of Michigan. Special thanks are due to Bob Fay for assistance on the derivation of the variance. We also thank Pat Cantwell, Diane Colasanto, Bob Fay, Jon Krosnick, Warren Mitofsky, Colm O'Muircheartaigh, Paul Siegel, Eric Slud, and anonymous reviewers for helpful comments on an earlier draft. Address correspondence to Elizabeth Martin; e-mail: elizabeth.ann.martin@census.gov.

Introduction

In the aftermath of the polling industry's disaster in the 1948 election, a distinguished group of social scientists and statisticians quickly mounted an intensive review of the election polling procedures and results to evaluate what had gone wrong. In just five weeks, they produced a remarkably thorough assessment, published by the Social Science Research Council (Mosteller et al. 1949). Mosteller's own chapter, "Measuring the Error," was a lucid review of the major poll estimates, and his work established the measures that have been used ever since to evaluate the accuracy of election polls.

The authors of the Social Science Research Council (SSRC) report considered their work to be preliminary, and they expressed the hope and expectation that "definitive and more leisurely studies" of the problems they identified would be conducted. Since then, there have been discussions of the merits of Mosteller's error measures (see, for example, Mitofsky 1998); but curiously, there has been relatively little follow-up work by statisticians to improve or evaluate them. In the succeeding period, there have been occasional controversies about the accuracy of preelection polls, including the underestimation of Ronald Reagan's victory over Jimmy Carter in 1980 and the overestimation of Bill Clinton's victory over Bob Dole in 1996. From time to time, there have also been more general calls for a review of the accuracy of election polls. For example, in 1984 the Panel on Survey Measurement of Subjective Phenomena recommended the establishment of "a panel or committee to evaluate the performance and methodology of election polls" and noted that "a regular review of the accuracy of such forecasts could be of use both to the survey industry and to the public" (Turner and Martin 1984, p. 314). Since the advent of the modern polling period, the role of preelection polls in forming the image of the entire industry has grown because, unlike most surveys, preelection forecasts may be judged against an external criterion of validity—the actual outcome of an election. Thus, the performance of preelection polls in forecasting elections may shape public perceptions of the accuracy of surveys more generally.

Across this same period, political strategists and social critics from all domains of the political spectrum have challenged the accuracy of polls and the role they play in contemporary society. Polls in recent elections have been charged with partisan bias (by, for example, Huffington 1996, 1998, and 2001; Ladd 1996; Rutenberg 2004). These criticisms and claims of bias should be addressed empirically in order to evaluate them systematically. This argues for a regular, independent review of the polls, in the good years as well as the bad. We also think it is time to take a fresh look at measures of election poll accuracy. In the 50 years since Mosteller's work, there have been advances in statistical theory and estimation of error that might yield alternative and perhaps better measures of election poll accuracy.

Especially in the last two presidential election cycles, there has been increased attention paid to the volatility of some of the preelection polls, as well as to

differences in the estimates that they produced during the campaign. Erickson, Panagopoulos, and Wlezien (2004) provide one explanation for this volatility, suggesting that the "likely voter" screen used by Gallup in its 2000 preelection polls was too sensitive to short-term fluctuations in voter enthusiasm early in the campaign, producing day-to-day volatility in preferences. In the 2004 campaign the American Association for Public Opinion Research (AAPOR) posted on its Web site a primer on sources of variation in published election polls (Zukin 2004). This concern suggests the need for the development of a statistical tool to assess the behavior of polls within the campaign, as well as at the very end.

This article proposes a new measure of the predictive accuracy of election polls that permits examination of both accuracy and bias. We illustrate and assess the measure by applying it to a number of preelection estimates, including the results of 548 state polls from gubernatorial and senatorial races in the 2002 elections. As we show in a series of examples, this new measure replicates assessments of past poll performance in historically important elections where such performance has been questioned. It is less vulnerable to alternative treatments of undecided voters than the traditional Mosteller measures. It can be used to make comparisons across elections with different outcomes and to analyze variability in the results achieved by different polling organizations within and between elections.

A Review of Past Measures of Poll Accuracy

Many researchers have written about preelection surveys and their methodology, including discussions of how such polls should be interpreted. However, the systematic evaluation of polling accuracy begins with the report by Mosteller et al. (1949) after the debacle of 1948. A multidisciplinary team of academic and government survey researchers, called the Committee on the Analysis of Preelection Polls and Forecasts, was assembled under the auspices of the SSRC to review and assess the polling methods, as well as the results. The committee received the full cooperation of the major public data collectors of the time, and the results of its work were widely distributed within the industry. Mosteller proposed eight measures that might be used to evaluate the accuracy of election forecasts, six of which were based on the estimated proportion of the vote that a (leading) candidate received or the difference in the estimated margin between the leaders.¹

1. An additional measure involved a chi-square statistical test; it was dismissed because of the burden of calculations 50 years ago, before the advent of modern computing. Another was based on projections of electoral votes, a common practice in the 1940s. While that practice disappeared with the advent of telephone surveys and broad national samples that did not represent individual states, changes in the cost and technology of polling and the use of sophisticated statistical modeling techniques brought this practice back in the 2000 presidential election campaign (Traugott 2001). This may become more prevalent in the future.

The Mosteller team acknowledged a number of problems that were associated with producing estimates of election outcomes from preelection polls. It is always easier when there are only two candidates in a race rather than three or more. It also gets more complicated conceptually when one considers the "total error" in a survey, rather than the difference between the outcome and the estimate for a single candidate. In presidential elections the number of significant third-party candidacies (those receiving more than 5 percent of the vote) has remained important, although third-party candidates are less likely to appear in other statewide offices. The timing of the estimation or projection relative to Election Day can also present problems because campaigns do matter, and last-minute shifts can occur in the electorate.

Mitofsky (1998) noted a lack of consensus in the industry about the best measure for gauging poll accuracy and compared results for four of Mosteller's original methods. He decided that the best choice was between Mosteller's measure 3 (average deviations for each party or candidate) and measure 5 (the difference in the differences between the leading candidates in the polls and the actual results). Measure 3 captures "the error by averaging the deviations in percentage points between predicted and observed results for each party (without regard to sign)," and Measure 5 uses "the difference of the oriented differences between predicted and actual results for the two major candidates" (Mosteller et al. 1949, p. 55; one might quibble with Mosteller's lack of formulas to define the measures). Measure 3 corresponds to the error on the candidates, and measure 5 to the error on the margin between the two leading candidates. When there are just two candidates, measure 3 is half of measure 5, if there are no undecided voters. Mitofsky (1998) favored measure 5 because it is comparable for both two-candidate and multiple-candidate elections. It also evaluates the statistic most often reported by the media, the margin between the top two candidates. He criticized measure 3 because it is not comparable between elections with different numbers of meaningful candidates. Measures 3 and 5 have been used in subsequent evaluations of election poll accuracy, such as Traugott's (2001) evaluation of poll performance in the 2000 campaign and the National Council on Public Poll's (2002) review of the 2002 election polls.

Crespi (1988) conducted the major study of the accuracy of preelection polling. He assembled a set of 430 final preelection polls that had been publicly available or disseminated after 1979. Almost three-quarters of them were for races other than president, and more than 400 of them were for subnational geographical units, mostly states and municipalities. Crespi recalculated the percentage for each candidate after excluding the undecideds, and then he considered three different measures of accuracy (the deviation from the election results): the difference in the outcome for the winning candidate; mean percentage difference in the outcome for the top three candidates; and the largest difference between the poll result and the actual outcome for any of the top three candidates. The three measures were highly correlated (between .81 and .93), and Crespi chose to pursue his study with the first measure because it

was simplest to calculate. In some analyses Crespi used the actual value of the difference for the top vote getter and the poll; in some cases he used an ordinal variable with three categories that distinguished the relative accuracy of the poll. Using an ordinal measure of the length of time the interviews were conducted before the election, Crespi found that accuracy increased in final polls that were taken closer to an election ($r = .21$).

Lau (1994) analyzed the performance of 56 national "trial heat" polls in 1992 that were conducted during the last month of the campaign. His primary dependent variable was the difference between a specific poll result and the "average of all available poll results (weighted by sample size) except the poll whose accuracy was being judged." Lau found the most significant predictors of accuracy were the number of days a poll was in the field, conducting interviews only on weekdays (negative), and conducting a tracking poll. To his surprise, he found no relationship between sample size and accuracy.

While most of the previous assessments of polling accuracy focused on national polls in presidential election years, several examined state polls. Rademacher and Smith (2001) looked at 79 state-level estimates of presidential races in 2000. Their analysis paralleled a National Council on Public Polls (NCP) analysis of the national polls, using the same measure of "candidate error"—taking one-half of the absolute difference between the top two candidates in the poll and the difference between their electoral results. This approximates Mosteller's measure 5, although it ignores the relative standing between the two candidates. In addition, the measurement in absolute terms eliminates the possibility of investigating systematic errors in the estimates. As a result, both Rademacher and Smith and the NCPP report also looked at whether the polls predicted the correct winner. Using these dual criteria, Rademacher and Smith found the state polls did not compare favorably with the national polls. The "candidate error" was about 70 percent greater (averaging 1.9 percentage points compared to 1.1 for the national polls), and in about one in five cases the error was greater than sampling error would suggest. In 15 percent of the cases, the polls suggested the wrong candidate would win, although many of these estimates were in states that turned out to be very close.

DeSart and Holbrook (2003) replicated Lau's (1994) analysis using statewide presidential "trial heat" polls conducted during the 1992, 1996, and 2000 elections. Using the absolute value of measure 5, they found that larger sample sizes and use of likely voter screens produced more accurate predictions of election outcomes. Like Lau, they found no overall effect of the number of days to election on accuracy, but the relationship was complex and varied in different election years.

To summarize this prior work, we note that most of the measures of accuracy have focused on the relationship between estimates for a single candidate or party's vote and the outcome of the election or on the difference between the two leading candidates. These assessments have predominantly been conducted on the basis of the absolute value of these measures, which eliminates

the possibility of evaluating the direction of any bias. In addition, many of the assessments have focused narrowly on the "horserace of the horserace polls" by looking at how well individual polling organizations predicted election outcomes, usually by ranking their performance on one or more of the Mosteller measures. This approach begs the question of how important house differences actually are as a source of variations in accuracy and neglects other substantive or methodological factors that may influence accuracy. We address these issues in proposing a new measure of polling accuracy and assessing its utility.

Measurement Issues

An ideal measure for assessing poll accuracy would have several properties, including:

- The measure would be comparable over different elections and different polls (Mitofsky [1998] gave this as the overriding consideration);
- The measure could be used in two-candidate as well as multiple-candidate elections;
- The measure would permit aggregate analyses of a large number of polls in order to examine correlates and predictors of accuracy;
- The measure would make it possible to assess both accuracy, in the sense of closeness to an election outcome, as well as bias, the direction of errors; and
- The measure would be unaffected by artifactual variations among polls, for example, in the size of the undecided category.

Next, we discuss three measurement issues that must be addressed in developing a measure of poll accuracy.

TREATMENT OF UNDECIDED VOTERS

As Mitofsky (1998) noted, handling the undecided vote in the polls is a significant problem that was not addressed by the SSRC report. Most of the methods defined by Mosteller are affected by the size of the undecided category and by decisions about whether and how to allocate them (and, for measure 3, the size of any third-party or other parties' candidate's share). Some polls allocate undecided voters and some do not, and measures that rely on percentage point differences (or differences of differences) will not be comparable between such polls.

It is useful to distinguish several influences on undecided responses in a preelection poll. One is genuine uncertainty on the part of people who have not yet made up their minds about whom to vote for. Their subsequent decisions may be influenced by campaign activities and real-world events that cannot be predicted at the time of a poll. Different statistical methods are often used to impute candidate preferences for respondents who are undecided by

using their answers to other poll questions to draw inferences about their likely candidate preferences, and such methods may be reasonably predictive of how these individuals vote (see, for example, Fenwick et al. 1982; Lam and Stasny 2001). Any global, simplifying assumption about their emerging preferences may be erroneous because it ignores the effect of a campaign on how voters decide to vote.

A second influence is methodological. The wording, format, and order of the questions in a survey may influence the fraction of people who report they are undecided. For example, asking substantive questions prior to the question on candidate preference reduces the undecided fraction (McDermott and Frankovic 2003), as does asking for candidate preference by means of a secret ballot (Perry 1979). Different procedures and questions used by different polling organizations can result in differences among polls in the size of the undecided category. These artifactual variations reduce the comparability of most of the Mosteller measures across polls.

The issue of how to treat undecided voters in preelection polls is related to the broader question of how to treat "don't know" or "no opinion" responses in surveys more generally. Experimental research shows that the presence or absence of an explicit "no opinion" or middle category affects the percentage that provide each substantive response but not the relative proportions (Kalton, Collins, and Brook 1978; Presser and Schuman 1980; Schuman and Presser 1978). That is, encouraging or discouraging "don't know" responses appears to influence the number of "don't know" responses but not the substantive distribution when the "don't knows" are excluded.

If the preferences of undecided respondents are not imputed, there are several alternatives for allocating them after the fact, as described by Mitofsky (1998), such as:

1. Allocate the undecided in proportion to the votes for the candidates in a poll;
2. Allocate the undecided evenly between the two major parties;
3. Allocate all the undecided to the challenger, if there is an incumbent; or
4. Drop the undecided and recalculate the candidates' shares.

The NCPP review of the 2002 polls did not adjust its error calculations for the number of undecided, "in order to avoid an arbitrary decision about how to allocate" them (NCPP 2002). However, this decision implied that their error measure was not comparable across polls that imputed preferences for undecided voters and those that did not. The ranking of the accuracy of the preelection polls in 1996 was different when the undecided were allocated (proportionally) than when they were not allocated.

The methodological literature seems to support the assumption of a proportional distribution of undecided voters, especially to the extent that variations in the undecided fraction result from methodological causes. In addition, Mitofsky (1998) found more consistency among different measures when the

undecided were allocated proportionally than when they were not allocated. These two considerations suggest the use of alternative 1, or dropping the undecided, as does alternative 4, which implicitly assumes a proportional distribution.

A desirable property of a measure of poll accuracy is that it be unaffected by variability in the size of the undecided category and by decisions about whether to allocate undecided respondents. However, it would be useful to have more empirical research on the extent to which the relative distribution of candidate preferences is constant in the presence of fluctuations in the size of the undecided category.²

BIAS

The Mosteller measures focus on accuracy in the sense of comparing estimates to the actual election outcome. Another issue is how to measure bias, or the extent to which polls systematically over- or underestimate a given party's share of the vote. This has been a contentious element of the statistics used to measure accuracy because sampling error is assumed to be symmetrical around a poll-based estimate. Measure 3 does not permit analysis of bias because it is the average of (unsigned) deviations between predicted and observed results for a candidate. Mitofsky (1998) used measure 5 to address Ladd's (1996) claim that the election polls have frequently overestimated the Democrats' share of the vote by counting the number of polls that overstated Democratic or Republican strength. He concluded that on this point Ladd was correct since more than twice as many polls overstated the Democratic share as understated it. But most evaluations of poll accuracy examine absolute errors (for example, NCPP 2002; Traugott 2001) and hence do not examine bias. Bias is also a difficult issue to deal with because one component of the difference between a poll estimate and a candidate's actual standing in the election can be attributed to sampling error, a random statistical element of the study design. Any measure of bias must take into account that differences in estimates may arise by chance alone.

THIRD-PARTY CANDIDATES

Assessing poll accuracy is more difficult when there are multiple candidates in a race. One problem is how and whether to combine information about the closeness of a poll's prediction for each of multiple candidates to provide a single measure of accuracy, if this is desired. A second problem is that a small error in absolute terms is much larger in relation to a minority party's share of the vote than it is in relation to a majority party's share. An error of 2 percentage

2. Research on the allocation of "undecided" voters has usually examined one polling organization at a time and the difference in estimates produced by different allocation schemes (Daves and Warden 1995; Fenwick et al. 1982; Lam and Stasny 2001; Visser et al. 2000), with almost no comparisons of differences in procedures across polling organizations and their consequences.

points (say) is large for a poll forecasting the outcome for a candidate who received just 4 percent of the vote but fairly small for a candidate who received 50 percent of the vote. In the former case the poll is off by half, and in the latter it is off by 4 percent of the true proportion. Error measures that rely on percentage point differences do not adequately capture the magnitude of an error in relation to a candidate's share of the vote, especially when candidates receive widely disparate shares in a lopsided election. Measures that average the percentage point deviation between the poll and the actual vote for each candidate (such as Mosteller's measure 3) were judged by the SSRC committee and by Mitofsky to underestimate the error in multicandidate races, with the problem increasing as the number of candidates grows.

In the measure we develop and apply here, we ignore third-party candidates to focus on the two major parties. However, the measure can be extended to third-party candidates, and we suggest how this might be done in the discussion.

A New Measure of Predictive Accuracy

To examine both the accuracy and bias of preelection polls, we introduce a measure based upon odds ratios rather than percentage point differences.³ The measure has the desirable statistical property of being unaffected by fluctuations in the size of the undecided category and by the size and number of third-party candidacies. The measure of predictive accuracy that we propose is based on the ratio of the following odds (the two major American political parties are used for purposes of illustration):

1. The odds on a *Republican choice in a given poll*, defined as r_i / d_i , where r_i is the proportion of respondents favoring the Republican candidate and d_i is the proportion favoring the Democratic candidate in poll i , and where n_i is the total number of respondents who favor either the Democrat or the Republican in the poll.

The odds measure has a clear interpretation: odds greater than one imply a Republican lead in poll i , odds less than one imply a Democratic lead, and odds equal to one imply a tie.

A poll conducted for the 2002 Alabama governor's race provides an illustration. A total of 900 people were interviewed, with 39 percent favoring the Democratic candidate, 45 percent the Republican, and 16 percent undecided. We ignore the undecided, and form the odds $r_i / d_i = .5357 / .4643 = 1.154$. Note that the effective sample size n_i is reduced to 756, not 900, for this measure. Note also that the same value of the ratio is obtained using numbers or proportions, regardless of whether the undecideds are included or excluded from the denominator ($405 / 351 = .5357 / .4643 = .45 / .39 = 1.154$).

3. While it may be less familiar than the margin, the odds is the natural unit for expressing expectations in the original model for "horserace" polls—the races that involve betting and real horses.

2. The odds on a *Republican choice in an actual election*, defined as R_{jk}/D_{jk} , where R_{jk} is the number of voters who favor the Republican candidate and D_{jk} is the number of voters who favor the Democratic candidate in an election for the j th office (governor, senator) in the k th state.

In the 2002 Alabama governor's race, for example, the Republican won a cliff-hanger with 50.1 percent of the vote, or 672,225 votes to 669,105 for his Democratic opponent. Thus, the election odds is 1.005—very close to a tie, but slightly greater than one, indicating a Republican victory.

From the two odds, we calculate the odds ratio by dividing the poll odds by the election odds: odds ratio $_{ijk} = (r_{ijk} / d_{ijk}) / (R_{jk} / D_{jk})$ for poll i , office j , in state k . In our example, this would be $1.154 / 1.005$, or 1.148.

The odds ratio also has a clear conceptual interpretation: an odds ratio of exactly one implies the poll and the election odds are in perfect agreement, with exactly the same *relative* distribution of voter preferences between the Republican and Democratic candidates. The farther from one an odds ratio is, then the worse the poll performed at predicting relative preferences in the election. An odds ratio less than one implies that a poll favored the Democrat compared to the actual election result, while an odds ratio greater than one implies a poll favored the Republican compared to the election result. The odds ratio of 1.148 in our example shows that the poll overestimated the Republican share of the vote. Some departures from one are to be expected due to sampling error, of course. Departures that exceed sampling error can be regarded as the bias characterizing a poll.

We transform the odds ratio by taking its natural log to make it symmetric and to simplify the calculation of the variance.⁴

4. We are grateful to Bob Fay for deriving the variance formula for predictive accuracy A , as follows:

Let r_i and d_i be random variables, with

r_i = the proportion of people preferring the Republican candidate, and

d_i = the proportion preferring the Democratic candidate,

in poll i with sample size n_i , with $d_i + r_i = 1$. Let p = probability of preferring the Republican, and $q = 1 - p$ be the probability of preferring the Democrat. The covariance matrix of the vector (r_i, d_i) is

$$\text{Cov}\begin{pmatrix} r_i \\ d_i \end{pmatrix} = \frac{1}{n_i} \begin{pmatrix} pq & -pq \\ -pq & pq \end{pmatrix}, \text{ so that}$$

$$\text{Relcov}\begin{pmatrix} r_i \\ d_i \end{pmatrix} = \frac{1}{n_i} \begin{pmatrix} \frac{q}{p} & -1 \\ -1 & \frac{p}{q} \end{pmatrix}$$

Because $E((r_i/d_i)/(p/q)) \equiv 1$, $\text{Var}(\log((r_i/d_i)/(p/q))) \equiv \text{Var}((r_i/d_i)/(p/q)) = \text{Relvar}(r_i/d_i) = \text{Relvar}(r_i) + \text{Relvar}(d_i) - 2 \text{Relcov}(r_i, d_i)$

$$= \frac{1}{n_i} \begin{pmatrix} \frac{q}{p} + \frac{p}{q} + 2 \\ p - q \end{pmatrix} = \frac{1}{n_i pq} (q^2 + p^2 + 2pq) = \frac{1}{n_i pq}$$

Thus, we define our measure of predictive accuracy A as:

$$A_{jk} = \log[(r_{jk}/d_{jk})/(R_{jk}/D_{jk})] \quad [1]$$

$$\text{Variance}(A_{jk}) = 1/n_j r_{jk} d_{jk} \quad [2]$$

The statistic A may take on values of zero, or positive or negative values, and has the following properties:

- A is zero when the odds ratio defined above is one, reflecting perfect agreement between a poll and election result;
- A significantly negative value of A indicates a poll is biased in a Democratic direction (that is, its distribution was too Democratic compared to the election outcome);
- A significantly positive value indicates a Republican bias;
- Negative magnitudes are comparable to positive (unlike the odds ratio); and
- A is a logarithm to the base e , and its scale values represent exponents. Again illustrating with the poll conducted for the 2002 Alabama governor's race: with odds of 1.154 on a Republican choice in the poll and 1.005 in the election itself, the odds ratio is $1.154 / 1.005 = 1.148$ and the log of the odds ratio, A , is .138. A positive value of A indicates the poll overstated the preference for the Republican candidate, compared to the election outcome, by a factor of $e^{.138}$, or 1.148.

Was this poll biased? To assess the poll's bias, we construct a confidence interval around zero, the expected value in the absence of bias. The variance of A is $1/nrd$, or $1/(756 \times .5357 \times .4643) = .005$, and its standard error is .073, so a 95 percent confidence interval includes $0 \pm .143$. Since the value of A for this poll is within the confidence interval, we conclude it is not significantly biased. A significantly positive value of A would have indicated Republican bias, while a significantly negative value would indicate Democratic bias.

This measure has several advantages compared to the traditional measures that rely on percentage point differences to measure a discrepancy. First, the odds ratio and log of the odds ratio are amenable to multivariate analysis and modeling using log linear methods. That is to say, they can become dependent variables in equations where the explanatory factors can be either methodological attributes of the preelection polls, such as timing or sample selection procedures, or contextual factors that distinguish the elections, such as type of race, state attributes, or incumbency. Thus, this measure facilitates analyses of the factors that predict preelection poll accuracy.

Second, the measure is less vulnerable to decisions about allocation of the undecided, as we show below. Indeed, the odds is the natural way of representing what seems to be fairly well established in public opinion measurement,

which is that the *relative* proportions in substantive categories are unaffected by changes in the size of the “no opinion” category.

Third, the measure is standardized for the actual election result, providing a measure of bias that is comparable over elections with different outcomes. The magnitude of a poll’s bias is defined relative to an election outcome. This makes it possible to do a meta-analysis of the nature and causes of bias affecting an entire corpus of polls conducted for different races or different years. The measure can also be used to compare the performance of individual polling firms or polls across a number of elections or races.

It is important to note the particular sense in which we interpret A_{jk} as a measure of accuracy: A_{jk} measures the accuracy of a poll as a *predictor of an election result*. A poll result might not accurately reflect voters’ relative preferences between the Republican and Democratic candidates for several reasons, including sampling error and flaws in its design. For example, a particular poll might not employ a sample designed to reflect the participating electorate (often referred to as “likely voters”) on Election Day.⁵ However, it is important to note that a poll might reflect a different distribution of Republican versus Democratic preferences for perfectly valid reasons that have nothing to do with problems or errors in the poll. While the typical assessment of poll performance is based on the final preelection estimates, other interesting research questions can assess the relative accuracy of estimates across the campaign. It is entirely possible, even likely, that public preferences shift during the course of a campaign. Thus, a poll that perfectly measured preferences at the time of the poll might still be a “biased” predictor of an eventual election outcome due to changes in the electorate and its preferences, not a flaw of the poll. We refer to A as a measure of *predictive accuracy* to emphasize the sense in which we interpret a poll as “accurate.”

Applying and Evaluating the New Measure: Illustrative Analyses

We illustrate the measure’s features by applying it to several well-known presidential elections and the state polls conducted for the 2002 gubernatorial and senatorial races and then comparing its results to the traditional measures. Our illustrative analyses draw on published election poll data (Mitofsky 1998; Mosteller et al. 1949; Traugott 2001), as well as the complete corpus of state polls conducted in 2002 for statewide races. In our first example, we use A to confirm historical assessments of several presidential elections. In a second example, we show how A and Mosteller’s measure S are affected by the fraction

5. Mosteller acknowledges this difficulty as well (Mosteller et al. 1949, p. 54). For research on estimating likely voters, see Chang and Krosnick 2002; Freedman and Goldstein 1996; Lau 1994; Lavrakas et al. 1997; Monson 1998; Perry 1973; Traugott and Tucker 1984; Visser et al. 2000; and Voss, Gelman, and King 1995.

of voters who are undecided and their allocation. A third example compares the two measures as applied to previously analyzed polls from the 2000 presidential election. In a fourth example, we show how A can be applied to characterize and analyze all of the published statewide polls that were conducted in 2002.

EXAMPLE 1: THREE PRESIDENTIAL ELECTIONS

We first illustrate our new measure of predictive accuracy by applying it to characterize the well-studied 1948, 1996, and 2000 presidential elections. In each case, we take the mean value of A over the final preelection polls conducted for that election, treating each poll as a single (unweighted) observation to calculate the standard error of \bar{A} .⁶

Data presented in table 1 show that the mean value of A for the 1948 election is significantly positive (it is more than three times its standard error), consistent with the familiar fact that the election polls that year showed a spectacular Republican bias. As shown in table 1, the 1948 polls found that 54.6 percent preferred the Republican candidate Thomas Dewey, but only 47.7 percent of the electorate voted for him. The mean value of A for the 1996 presidential election is significantly negative, showing a Democratic bias, as Ladd (1996) charged and Mitofsky (1998) affirmed using a cruder measure of bias. Finally, the mean value of A for the 2000 presidential election is significantly positive, consistent with the fact that 14 of the 19 preelection polls analyzed indicated that George W. Bush would win the popular vote. Because A is standardized for the election outcome, we may directly compare the magnitude of bias that characterized the polls in these three election years. We note that the 1948 polls were much more biased than the polls in either 1996 or 2000, as reflected by a significantly larger value of A . The overall bias characterizing the polls was about the same size in the 1996 and 2000 presidential elections, but in the opposite direction, with the 1996 polls showing a Democratic (negative) bias and the 2000 polls showing a Republican (positive) bias. The value of A is consistent with and confirms the generally accepted evaluations of the performance of the polls in these three historically important elections.

EXAMPLE 2. POLLS THAT DIFFER IN LEVEL AND TREATMENT OF UNDECIDED VOTERS

A second example allows us to show how the measures are affected by different levels of, and global assumptions about, undecided voters. The left-hand part of table 2 presents results for four hypothetical polls that vary in the fraction of voters who are undecided. In this hypothetical example,

6. Standard errors for \bar{A} were calculated using a jackknife replication method using VPLX (Fay 1998) and treating each election year's polls as simple random samples.

Table I. Mean Predictive Accuracy in Polls for Three Presidential Elections

Election	% Republican in election	Mean % Republican in polls	N of polls	\bar{A}	Standard error of \bar{A}
1948	47.7%	54.6%	3	.2783	.0781
1996	45.2%	43.2%	9	-.0838	.0221
2000	49.7%	51.3%	19	.0630	.0121

SOURCES.—Mosteller et al. 1949, p. 17; Mitofsky 1998, table 1; Traugott 2001, table 1.

NOTE.—Percentages are recalculated excluding votes for any third-party candidates.

each poll shows a breakdown of 60 percent for the Republican candidate and 40 percent for the Democrat when the undecided are excluded. We assume the election outcome also was a 60-40 victory for the Republican. That is, the election and all four polls show the same relative proportions for r_i and d_i .

Table 2 shows calculations of measure 5 (signed), the error on the margin, under three alternative, neutral treatments of undecided voters: including them in the base when calculating the margin (column 1), dropping them entirely or allocating them in proportion to r_i and d_i (column 2), or allocating them equally to each candidate (column 3). These different allocations are used to show their effects on the measures and not to advocate one over another. Different polling firms make their own determinations as to whether and how to impute the preferences of the undecided voters they interview. The poll margin and measure 5 are calculated for each treatment.

With no allocation (column 1), the poll margin changes as the size of the undecided category changes. The poll margin is 10 for poll 1 (with 50 percent undecided) and increases to 20 for poll 4 (with no undecided voters). Because the poll margin changes, so does the value of measure 5, even though the relative Republican and Democratic proportions are constant. On the other hand, when the undecided are dropped or proportionally allocated, the margin for all three polls is calculated as 20, the same as the election margin, and therefore the error on the margin is zero, as shown in column 2. Allocating the undecided equally to each candidate (column 3) preserves the poll margin and so produces an equivalent error on the margin as no allocation (column 1).

When measure 5 is calculated by dropping or proportionally allocating the undecided, we are led to conclude that all four polls are equally, and perfectly, accurate. On the other hand, when the undecided are split equally or included in the base, we are led to conclude that the polls with larger fractions of undecided are less accurate (they have higher absolute values of measure 5). Thus, different assumptions about the undecided lead to different conclusions about poll accuracy using measure 5. As Mitofsky (1998) noted, the rankings of poll

Table 2. Calculations of Poll Accuracy, Using Measure 5, Assuming Three Alternative Treatments of Undecided Voters

Poll	Rep. r	Dem. d	Und. u	Percentage Preferring				Treatment of Undecided Voters			
				Column 1: No allocation				Column 2: Proportional allocation (or undecided dropped)			
				Poll margin = $r - d$	Measure 5	Poll margin	Measure 5	Poll margin	Measure 5	Poll margin	Measure 5
1	30	20	50	10	-10	20	0	10	-10		
2	54	36	10	18	-2	20	0	18	-2		
3	57	38	5	19	-1	20	0	19	-1		
4	60	40	0	20	0	20	0	20	0		

Note.—Measure 5 = $(r_i - d_i) - (R - D)$. Calculations assume an election with margin $R - D = 60 - 40 = 20$.

accuracy according to measure 5 are altered when the undecided are allocated proportionally or not allocated.

Table 3 calculates A for the same four hypothetical polls. In contrast to the poll margin, the poll odds (the numerator in measure A) and therefore measure A are constant across the four polls under both the no allocation and proportional allocation conditions (columns 1 and 2). Keeping the undecided, dropping them, or allocating them proportionally does not alter the relative proportions r_i and d_i , and therefore the poll odds are constant, and so is measure A .

Splitting the undecided between the candidates alters the relative proportions favoring each, and so affects the poll odds and A in column 3.

A disadvantage of an accuracy measure based on the margin (such as measure 5) is its greater vulnerability to variations among polls in the size of the undecided category and to different assumptions about the preferences of undecided voters. The poll margin (and therefore measure 5) is not constant when the undecided category is included in or subtracted from the base:

$$\text{Poll Margin} = \frac{r-d}{(r+d)} \neq \frac{r-d}{(r+d+u)}.$$

In contrast, the calculation of the odds on voting Republican (rather than Democratic) is not affected by the size or exclusion of the undecided category. The skeptical reader can satisfy him or herself that the poll odds are the same for all four hypothetical polls regardless of whether the calculation includes undecided voters. The calculation of the poll odds is unaffected by what is included in the base when calculating the percentages favoring the Republican and the Democrat, because the base cancels out in both the numerator and the denominator of the odds:

$$\text{Poll Odds} = \frac{\cancel{r/(r+d)}}{\cancel{d/(r+d)}} = \frac{\cancel{r/(r+d+u)}}{\cancel{d/(r+d+u)}} = \frac{r}{d}.$$

This is an important positive feature of our measure because it simplifies decisions about how to treat undecided voters, simplifies calculations, and preserves the comparability of the measure among polls that differ in size and treatment of the undecided category. This comparability feature is important not only for comparing different polls in the same race but also for assessing the performance of polls conducted by a single or multiple polling firms across races or elections.

Table 3. Calculations of Poll Accuracy, Using Measure A , Assuming Three Alternative Treatments of Undecided Voters

Poll	Percentage Preferring		Treatment of Undecided Voters						
			Column 1: No allocation			Column 2: Proportional allocation (or undecided dropped)			
	Rep. r	Dem. d	Und. u	Poll odds = r/d	A	Poll odds	A	Poll odds	
1	30	20	50	1.5	0	1.5	0	1.22	-.20
2	54	36	10	1.5	0	1.5	0	1.44	-.04
3	57	38	5	1.5	0	1.5	0	1.47	-.02
4	60	40	0	1.5	0	1.5	0	1.5	0

NOTE.— $A = \log \frac{r/d}{R/D}$. Calculations assume an election with odds $R/D = 60/40 = 1.5$.

Table 4. A Comparison of the Accuracy of the 2000 Preelection Presidential Polls Using Mosteller's Measures 3 and 5 and the Predictive Accuracy Measure (A)

Polling Firm	Poll Odds	A	Rank on A	Rank on Measure 3	Rank on Measure 5
Fox	1.0000	0.0104	1	1	1
Harris (Internet)	1.0000	0.0104	1	2	1
Harris (Phone)	1.0000	0.0104	1	4	1
CBS	0.9778	-0.0121	4	2	1
Reuters/MSNBC/Zogby	0.9583	-0.0322	5	5	5
IBD/CSM/TIPP	1.0413	0.0509	6	5	5
Pew	1.0426	0.0521	7	5	7
CNN/USA Today/Gallup	1.0435	0.0529	8	5	7
ICR	1.0455	0.0548	9	14	7
Knowledge Networks	1.0455	0.0548	9	5	7
<i>Newsweek</i>	1.0465	0.0558	11	13	7
ABC	1.0667	0.0749	12	5	12
<i>Washington Post</i>	1.0667	0.0749	12	5	12
NBC/Wall Street Journal	1.0682	0.0763	14	5	12
Voter.com	1.1111	0.1157	15	13	15
Marist College	1.1136	0.1180	16	14	15
CBS/New York Times	1.1190	0.1229	17	14	15
Rasmussen	1.1329	0.1351	18	14	15
Hotline	1.1750	0.1717	19	19	19

SOURCE.—Traugott 2001, table 2.

EXAMPLE 3. 2000 PRESIDENTIAL POLLS

We now compare the results of our new measure with measures proposed by Mosteller by reconstructing Traugott's (2001) assessment of poll accuracy in the 2000 election, revised to include our measure of predictive accuracy. Table 4 presents rankings of the accuracy of the 19 preelection polls by three measures—predictive accuracy (A), Mosteller's measure 3, and Mosteller's measure 5.⁷ Polls are listed in order of their rank on measure 5 (and alphabetically where ranks are equivalent). A ranking of "1" indicates the most accurate poll, while the poll ranked "19" is least accurate. For our measure of predictive accuracy, we rank polls according to how close the absolute value of A is to zero.

Recalling that Al Gore won the popular vote by 48.4 percent to 47.9 percent for George W. Bush and 2.7 percent for Ralph Nader, the odds on voting

7. Recall that measure 3 is the average absolute difference in the candidate estimates, and measure 5 is the difference between the margin in the election for the top two candidates and their margin in the polls, which in Traugott (2001) is the absolute difference of the differences.

Republican in the presidential election were .9897. However, most of the poll odds are greater than one, indicating that most polls favored Bush, as shown in table 4. (Two polls produced odds of less than one, and three produced odds of exactly one.) The rankings produced by the three different error measures are roughly consistent. Because the scales for the three measures are composed quite differently, we correlate the rankings rather than the exact values of the measures. The correlation between rankings on the two Mosteller measures is .77, while the correlation between rankings on measure 3 and A is .81, and the correlation between the rankings on measure 5 and on A is .97. The data used to construct the Mosteller measures include an allocation of the "undecided" portion of the sample for measure 3. Thus, when used to rank individual polls in an election in which most were "biased" in the same direction, A provides consistent information with the traditional measures, especially measure 5.

EXAMPLE 4: META-ANALYSIS OF 2002 STATE PREELECTION POLLS

In presidential elections the number of distinct preelection polls is quite limited, and each of the estimates is assessed against the same outcome. The advantages of the proposed measure of predictive accuracy emerge more clearly when we apply it to statewide elections. For the hundreds of polls conducted in statewide races in a given election year, it would not be feasible to examine rankings of poll accuracy, as Mitofsky (1998) and Traugott (2001) did. In addition, the measures used in those assessments are not comparable across elections with different numbers of candidates and outcomes.

In this fourth example, we illustrate how the proposed measure can be applied to summarize and analyze a much larger number of polls—548 state-level poll results for the offices of governor and senator in the 2002 election.⁸ The proposed new measure allows us to characterize an entire body of polls and examine potential sources of bias.

As a group, the 2002 state election polls were accurate in forecasting the winners of state elections, if they could support any forecast at all. Of the 548 polls, 504 reported sufficient information to calculate a margin of error.⁹ Of

8. We gathered all the state 2002 polls that we could locate, including polls from <http://www.nationaljournal.com>, The Hotline, <http://www.harrisinteractive.com>, the 2002 NCPP report, an ABC News file, and <http://www.depolticalreport.com/2002/polls02short.htm>. We included partisan as well as nonpartisan polls. All polls were fielded on or after Labor Day. The variables analyzed in this article were coded based on publicly reported information, which was not available for all polls. This compilation of polls is different from the 159 polls analyzed by NCPP (2002), which excluded partisan polls and polls that were released too far in advance of Election Day.

9. We calculated a margin of error for each poll by applying the standard assumption of simple random sampling and dropping third-party and undecided voters from the calculation. We calculated a 95 percent confidence interval around the percentage preferring the Republican (versus Democratic) candidate in order to assess whether any projection of a winner was supportable from the poll. If a confidence interval included 50 percent, no projection could be supported.

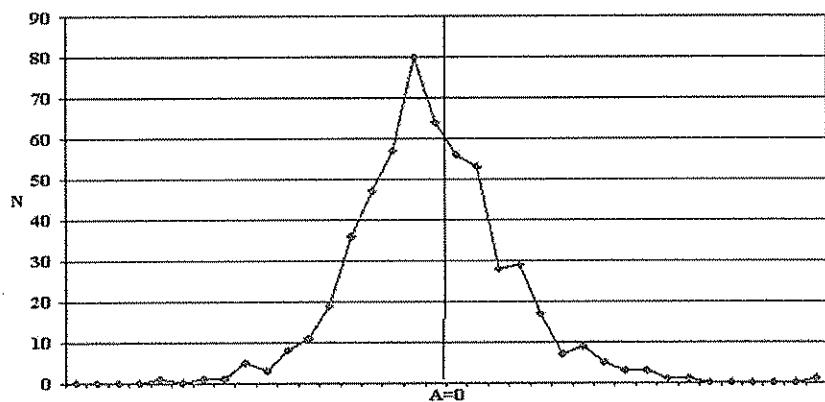


Figure 1. Frequency distribution of A_{yk} for preelection polls for statewide offices in 2002.

these, 57 percent (or 286 polls) could statistically support a projection of a winner or a leader, while 43 percent (218 polls) could not. Of the polls that could project a winner, the projection was correct 95 percent of the time (54 percent of all polls correctly projected the winner, and 3 percent were in error). When they could support projections, the polls were highly accurate in all elections, except in close elections won by Republicans. Polls ($N = 8$) in such elections were more often in error than correct. The fact that a large number of the polls could not statistically support any projection may indicate that their sample sizes were too small for the purpose of election forecasting.

In the meta-analysis below, we treat each poll as a single (unweighted) observation.¹⁰ Figure 1 shows the frequency distribution of the A_{yk} for the 548 state polls. Clearly, they are not centered on zero, as would be expected in the absence of overall bias. The mean value of A_{yk} is $-.0330$, with standard error .0077. In other words, there is a statistically significant Democratic bias over the polls as a whole. The bias is quite small, but it is potentially important in close races: if the races were all perfectly tied, on average the polls would have estimated a Republican share of 49.18 percent

10. We might obtain different results if we weighted by sample size, although the differences would probably not be too great since the variability in the number of interviews is not too large (ranging from about 300 to 1,500 among our 548 polls, with a mean of 484, excluding the undecided). Standard errors are calculated with jackknife replication methods using VPLX (Fay 1998).

rather than 50 percent.¹¹ In other words, there was a bias of -.72 percentage points.

The measure allows us to examine methodological and contextual factors that may be correlated with accuracy. For example, we may use A to compare the performance of individual polling organizations, a traditional focus of assessments of poll accuracy. Our measure permits comparisons of the accuracy of estimates produced by different survey organizations across races that differ in their outcomes. The full set of polls we analyzed was conducted by a large number of companies and groups, only five of which were nonpartisan organizations that conducted at least 10 polls and polled in multiple (three or more) states. These firms have some claim on broad regional and, in most cases, national coverage. Table 5 shows the performance of these five organizations and all other nonpartisan polls combined. Partisan Republican or Democratic polls are shown separately.

Most of the most active nonpartisan organizations performed well in the 2002 state elections, by our measure. Mason-Dixon, Quinnipiac, and Zogby had mean predictive accuracy not significantly different from zero, indicating no bias. Research 2000 and SurveyUSA showed significant Democratic biases. There were few significant differences among the firms. Both Research 2000 and SurveyUSA were significantly more Democratic than Quinnipiac, and SurveyUSA was significantly more Democratic than Zogby. However, these firms were not polling in the same races, and that should be kept in mind when interpreting differences in predictive accuracy. With one exception, organizational differences (or "house effects") were smaller than biases due to partisanship. Possibly, past assessments have overemphasized differences among individual polling organizations in the accuracy of their estimates.

11. The estimated Republican share \hat{r}_{ijk} is calculated as $\hat{r}_{ijk} = \frac{100 \times e^{A_{ijk}}}{1 + e^{A_{ijk}}}$. The difference between the Republican share so calculated and 50 percent is a percentage point difference measure that is in effect standardized by assuming a tied election. The derivation is as follows:

$$\text{From equation 1, we have } A_{ijk} = \log \frac{r_{ijk}/d_{ijk}}{R_{jk}/D_{jk}}.$$

$$\text{Exponentiating, we have } e^{A_{ijk}} = e^{\log \frac{r_{ijk}/d_{ijk}}{R_{jk}/D_{jk}}} = \frac{r_{ijk}/d_{ijk}}{R_{jk}/D_{jk}}.$$

In a tied election, $R_{jk} = D_{jk}$, and $R_{jk}/D_{jk} = 1$. Thus $e^{A_{ijk}} = r_{ijk}/d_{ijk}$. Assuming $r_{ijk} + d_{ijk} = 100$, then

$$e^{A_{ijk}} = \frac{r_{ijk}}{100 - r_{ijk}} \text{ and } (100 - r_{ijk})e^{A_{ijk}} = r_{ijk}.$$

$$\text{Solving for } \hat{r}_{ijk} \text{ we have } \hat{r}_{ijk} = \frac{100 \times e^{A_{ijk}}}{1 + e^{A_{ijk}}}.$$

This formula may be used to convert the parameter A_{ijk} into a percentage point difference measure that is likely to be more intuitive to many analysts. However, the analyst must keep in mind that \hat{r}_{ijk} is artificial because it is standardized by assuming a tied election.

Table 5. Mean Predictive Accuracy (A) of Results Obtained by Nonpartisan, Democratic, and Republican Polling Organizations

Polling Organization	Number of Polls	Mean Predictive Accuracy (A)	Standard Error
Mason-Dixon	47	-.0251	.0203
Quinnipiac	19	.0371	.0387
Research 2000	35	-.0583	.0203
SurveyUSA	28	-.0774	.0213
Zogby	51	-.0039	.0251
Other Nonpartisan Polls	289	-.0324	.0115
All Nonpartisan Polls	469	-.0304	.0082
Democratic Polls	41	-.1576	.0241
Republican Polls	38	.0699	.0270

Both Democratic and Republican polls were significantly biased in favor of candidates of their own party, relative to election results, and both were significantly more biased than the nonpartisan polls as a group. Using the formula provided in footnote 11, we calculate that, in perfectly tied races, the average bias would have been -4.03 percentage points for Democratic polls, +1.75 percentage points for Republican polls, and -.76 percentage points for the nonpartisan polls.¹²

Our proposed measure also allows us to examine methodological factors that may influence poll accuracy. For example, different polls frame preference questions differently (see, for example, McDermott and Frankovic 2003) and may rely on different methods for identifying people who are likely to vote in an election. Such differences in methods may contribute to "house effects," such as those shown in table 5. To the extent that information about methods is available, we can apply the measure to examine whether accuracy is correlated with methodological differences among polls. For example, one factor often assumed to influence the accuracy of a poll is its timing. Thus, the NCPP included only final polls in its review and dropped polls in which the interviewing was completed before October 20, 2002. Our measure can be applied to test empirically the common assumption that polls taken close to an election are more accurate than those taken far in advance. Table 6 shows mean predictive accuracy by the number of weeks in advance of the 2002 election that a state poll was taken, separately for neutral, Democratic, and Republican polls.

12. These biases in partisan polls may have many sources. For example, partisan pollsters might release polls selectively, so that Democratic pollsters only release their polls publicly if they favor Democratic candidates, and similarly for Republican pollsters. In addition, the partisan and neutral polls occurred in different types of races. For example, partisan polls were concentrated in certain states (for example, Louisiana, New Hampshire, South Dakota, Tennessee, and Texas), in governors' more than Senate races, and early rather than late in the campaign. Some of the differences in predictive accuracy shown in table 5 may reflect differences in the particular campaigns pollsters focused on.

Table 6. Mean Predictive Accuracy (A) of State Polls, by Number of Weeks Before an Election the Poll Was Taken

Partisanship of Poll Auspices	5–10 Weeks Before Election	During Fourth Week Before Election	During Third Week Before Election	During Second Week Before Election	During Final Week Before Election
Nonpartisan Polls	.0115 (.0176)	-.0332 (.0361)	-.0889 (.0266)	-.0350 (.0192)	-.0453 (.0104)
Democratic (or Dual)	-.1321 (.0392)	-.2373 (.0805)	-.2087 (.0323)	-.1259 (.0454)	-.0873 (.0267)
Republican Polls	.0762 (.0331)	.0731 (.1096)	.0755 (.0820)	.0360 (.0617)	.0625 (.0288)

NOTE.—Standard errors are shown in parentheses.

This longitudinal analysis yields modest evidence that the polls became better predictors (trended toward $A = 0$) of the final vote distribution closer to the 2002 election. In terms of both absolute and signed values, however, the improvement was not statistically significant. Indeed, nonpartisan polls taken more than a month in advance were significantly more accurate than those in the final week ($t = 2.78$). A significant Democratic bias emerged in neutral polls the last three weeks of the campaign, as shown by negative values of A that are twice their standard errors. This result casts some light on why the final election results were surprising to many poll watchers. Neutral polls were slightly but significantly overstating Democratic strength in the last month of the campaign, as shown in figure 2.

The bias of Democratic polls was statistically significant throughout the period, but its magnitude declined during the last month, as indicated by values of A that trended closer to zero. In the fourth week before the election, A was $-.2373$, then dropped significantly to $-.0873$ in the final week before the election. The bias of the Republican polls was significantly different from zero in polls taken 5 to 10 weeks before the election and in the final week.

The results for the neutral polls do not support Crespi's (1988) finding of a slight positive correlation ($r = .21$) between accuracy and timeliness of final preelection polls, but they do support Lau's (1994) finding from the 1992 campaign of no significant relationship between timing and his measure of accuracy. DeSart and Holbrook (2003) found that the relationship between a poll's timing and its accuracy varied in different elections.

These results challenge the conventional wisdom about the effects of poll timing on accuracy and suggest that the common assumption that only the final poll has predictive value should be further examined empirically in other elections and other years. The reasons for the patterns shown in figure 2 are

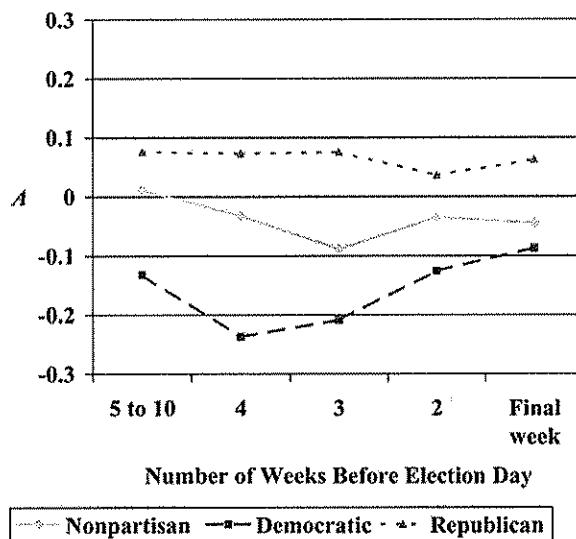


Figure 2. Mean predictive accuracy by number of weeks a poll was taken in advance of election 2002.

beyond the scope of this paper.¹³ Perhaps 2002 was a lackluster election cycle, with no overall “campaign effects” that moved the electorate away from preferences expressed 5 to 10 weeks before the election. DeSart and Holbrook (2003) find that the predictive ability of the polls is significantly shaped by campaign events, with the relationship between the number of days to election and poll accuracy varying greatly in the 1992, 1996, and 2000 presidential elections.¹⁴ Other causal factors correlated with poll timing may have influenced poll accuracy. A more complete multivariate analysis of the simultaneous effects of different variables on A might shed light on the causes and correlates of poll accuracy in 2002.

Discussion

We have developed and illustrated the application of a new measure A of predictive accuracy that makes it possible to quantify how well preelection polls perform, as well as the extent and direction of any biases they produce. We

13. While one explanation of bias in the final preelection polls might have been their inability to pick up the Republican mobilization efforts through the 72-Hour Task Forces, this cannot be the complete explanation because the biases appeared early in the campaign, before their work had started. See Traugott 2003.

14. Such campaign effects might have characterized particular races in 2002 and might be revealed by a state-level analysis.

believe that this new measure may prove useful as a summary measure of accuracy in election forecasts. It is easily computed and summarized, and it can be analyzed using multivariate statistical methods. It is comparable across elections with different outcomes. It does not require allocations of undecided voters, and in fact a major advantage is that it is not affected by variations of the size of the undecided category.

The measure has advantages for a number of potential different applications:

- A and its standard error can be calculated for an individual poll and used to assess whether that poll was significantly biased.
- A can be aggregated across polls, for example, all of the polls conducted in the 1996 presidential election, to characterize their predictive accuracy as a group.
- Because A is standardized for election outcome, it is comparable across elections with different outcomes, which makes it possible to directly compare the accuracy of polls in different elections, as illustrated in example 1.
- Because A is standardized for election outcome and because it relies only on the proportions favoring the Republican and Democratic candidates, it is comparable across different races with different numbers of candidates, making it well adapted for use in assessing the accuracy of a large number of polls representing diverse races and outcomes.
- A can be used to analyze causes and correlates of predictive accuracy in meta-analyses of a large number of polls, as illustrated in example 4. (Analysis relying on multivariate techniques would better control for the influence of multiple causal factors than the illustrative, simple analyses presented in example 4.)
- A is less vulnerable than measures based on the margin to fluctuations in the size of the undecided category and to the effects of alternative assumptions that involve dropping undecided voters, proportionally allocating them, or not allocating at all, as illustrated in example 2. Indeed, it does not require allocation of undecided voters.
- A can be applied as a substantive measure to analyze trends and "campaign effects." Changes in voters' preferences over the course of a campaign will mean (for example) that early polls are less predictive than later ones and will result in systematic changes in values of A over time. A could be analyzed to examine how campaign events, and time, move voters toward their final electoral choice.

The proposed new measure does have several limitations and disadvantages as well:

- It cannot be calculated until the final election outcome is known, so it can only be used to evaluate poll accuracy or variability after the fact.

- Its numerical values are not as readily interpretable as measures based on percentage point differences. We have provided a formula (in footnote 11) that converts values of A into percentage point differences, standardized by assuming a tied election. Transforming values of A into percentage point differences may aid their interpretation.
- A only evaluates the accuracy of a poll's forecast of the split between the two major party candidates in a partisan election.

In this article introducing the new measure, we have not considered the accuracy of poll predictions for third-party candidates, nonpartisan referenda, or primary elections. However, we believe this or related measures can be used to evaluate predictive accuracy in all of those situations.

The measure is readily adapted to nonpartisan referenda or primary elections where the interest is in the accuracy of predicting voter preferences between two candidates or alternatives. The variables r and d (for "Republican" and "Democrat") would be dropped from equations 1 and 2, to be replaced by more neutral variable labels (for example, c_a and c_b for primary candidates a and b). Otherwise the calculations would be identical. It would be an arbitrary decision whether to calculate A based on (for example) the odds c_a / c_b or c_b / c_a ; it would not matter, as long as the decision was made consistently throughout an analysis, and the poll odds and the primary election or referenda odds were formed consistently.

Extending measure A to third-party candidacies would require more developmental work, but we can suggest how it might proceed. Suppose in addition to Republican and Democratic candidates, we have a candidate representing a third party, the Greens. Then we have r_i , d_i , and g_i as the proportions preferring the three candidates in poll i . Then the odds $\frac{g_i}{r_i + d_i}$ represents the odds a person prefers the Green candidate to either the Democrat or the Republican. We can define a modified version of A , labeled A' , to measure how close a poll came to the Green Party vote, expressed as a fraction of the total vote cast for either major party candidate:

$$A' = \log \frac{g_i / (r_i + d_i)}{G / (R + D)}$$

Measure A' would capture the accuracy of prediction for the third-party candidate only and would supplement (and be statistically independent of) the accuracy of prediction A of voters' preferences between the two major party candidates.

We hope that others will assess our measure through empirical analyses of poll data. We hope it proves to be a useful and robust tool for understanding sources of poll variability and for identifying biases that need to be addressed. We believe it can help inform methodological research that leads to improvements in preelection poll accuracy and perhaps even leads to a better understanding of the effects of campaigns on the democratic electoral process.

References

- Chang, LinChiat, and Jon A. Krosnick. 2002. "Improving Preelection Surveys." Paper presented at the American Statistical Association Joint Statistical Meetings, Atlanta, GA.
- Crespi, Irving. 1988. *Pre-Election Polling: Sources of Accuracy and Error*. New York: Russell Sage Foundation.
- Daves, Rob P., and S. P. Warden. 1995. "Methods of Allocating Undecided Respondents to Candidate Choices in Pre-Election Polls." In *Presidential Polls and the News Media*, ed. Paul J. Lavrakas, Michael W. Traugott, and Peter V. Miller, pp. 101–19. Boulder, CO: Westview Press.
- DeSart, Jay, and Thomas Holbrook. 2003. "Campaigns, Polls, and the States: Assessing the Accuracy of Statewide Presidential Trial-Heat Polls." *Political Science Quarterly* 56(4):431–39.
- Erickson, Robert S., Costas Panagopoulos, and Christopher Wlezien. 2004. "Likely (and Unlikely) Voters and the Assessment of Campaign Dynamics." *Public Opinion Quarterly* 68(4):588–601.
- Fay, Robert E. 1998. *VPLX Program Documentation, Vol. 1*. Washington, DC: U.S. Census Bureau. Program and documentation available at <http://www.census.gov/sdms/www/vwelcome.html> (accessed July 5, 2005).
- Fenwick, Ian, Frederick Wiseman, John F. Becker, and James R. Heiman. 1982. "Classifying Undecided Voters in Pre-Election Polls." *Public Opinion Quarterly* 46:383–91.
- Freedman, Paul, and Ken Goldstein. 1996. "Building a Probable Electorate from Pre-Election Polls: A Two-Stage Approach." *Public Opinion Quarterly* 60(4):574–87.
- Huffington, Arianna. 1996. "A Modest Proposal." Available at <http://www.ariannaonline.com/columns/column.php?id=670> (accessed May 7, 2003).
- . 1998. "Investigating the Pollsters." Available at <http://www.ariannaonline.com/columns/column.php?id=484> (accessed May 7, 2003).
- . 2001. "Some Things Never Change: The Unbearable Ludicrousness of Polling." Available at <http://www.ariannaonline.com/columns/column.php?id=150> (accessed May 7, 2003).
- Kalton, Graham, Martin Collins, and Lindsay Brook. 1978. "Experiments in Wording Opinion Questions." *Applied Statistician* 27(2):149–61.
- Ladd, Everett C. 1996. "The Election Polls: An American Waterloo." *Chronicle of Higher Education*, November 22, p. A52.
- Lam, Eric, and Elizabeth A. Stasny. 2001. "Handling Undecided Voters: Using Missing Data Methods to Improve Election Forecasting." Working Papers Series, Department of Statistics, Ohio State University.
- Lau, Richard R. 1994. "An Analysis of 'Trial Heat' Polls during the 1992 Presidential Election." *Public Opinion Quarterly* 62:2–20.
- Lavrakas, Paul J., S. Mockabee, Q. Monson, and T. Thomson. 1997. "Predicting Voter Turnout and Vote Outcome in Ballot Elections: The Results of Four Preelection Surveys." Paper presented at the annual meeting of the American Association for Public Opinion Research, St. Louis, MO.
- McDermott, Monika L., and Kathleen A. Frankovic. 2003. "Review: Survey Methods and Horserace Polling." *Public Opinion Quarterly* 67:244–64.
- Mitofsky, Warren. 1998. "The Polls—Review: Was 1996 a Worse Year for Polls than 1948?" *Public Opinion Quarterly* 62:230–49.
- Monson, Joseph Q. 1998. "Predicting Voter Turnout in Local Elections: A Voter Validation Study." Paper presented at the annual meeting of the Midwest Association for Public Opinion Research, Chicago.
- Mosteller, Frederick, Herbert Hyman, Philip J. McCarthy, Eli S. Marks, and David B. Truman. 1949. *The Pre-Election Polls of 1948: Report to the Committee on Analysis of Pre-Election Polls and Forecasts*. New York: Social Science Research Council.
- National Council on Public Polls (NCPP) Polling Review Board. 2002. *Analysis of the 2002 Election Polls*. Press release issued Dec. 19, <http://www.ncpp.org/2002SenGovPoll/2002ElectionPolls.html> (accessed May 7, 2003). A database of the polls evaluated in the review can be found at <http://www.ncpp.org/2002SenGovPoll/2002SenGovPolls.html> (accessed July 5, 2005).
- Perry, Paul. 1973. "A Comparison of the Voting Preferences of Likely Voters and Likely Non-Voters." *Public Opinion Quarterly* 37(1):99–109.
- . 1979. "Certain Problems in Election Survey Methodology." *Public Opinion Quarterly* 43(3):312–25.

- Presser, Stanley, and Howard Schuman. 1980. "The Measurement of a Middle Position in Attitude Surveys." *Public Opinion Quarterly* 44:70–85.
- Rademacher, E. W., and A. E. Smith. 2001. "Poll Call." *Public Perspective* 12(2):36.
- Rutenberg, Jim. 2004. "A Request to Partisans: Don't Shoot the Pollster." *New York Times*, September 29, p. A17.
- Schuman, Howard, and Stanley Presser. 1978. "The Assessment of 'No Opinion' in Attitude Surveys." In *Sociological Methodology 1979*, ed. K. F. Schuessler, pp. 241–75. San Francisco: Jossey-Bass.
- Traugott, Michael W. 2001. "Assessing Poll Performance in the 2000 Campaign." *Public Opinion Quarterly* 65:389–419.
- . 2003. "Can We Trust the Polls?" *Brookings Review* 21(3):8–11.
- Traugott, Michael W., and Clyde Tucker. 1984. "Strategies for Predicting Whether a Citizen Will Vote and Estimation of Electoral Outcomes." *Public Opinion Quarterly* 48:330–43.
- Turner Charles F., and Elizabeth Martin, eds. 1984. *Surveying Subjective Phenomena*. Vol. 1. New York: Sage.
- Visser, Penny S., Jon A. Krosnick, Jesse F. Marquette, and Michael F. Curtin. 2000. "Improving Election Forecasting: Allocation of Undecided Respondents, Identification of Likely Voters, and Response Order Effects." In *Election Polls, the News Media, and Democracy*, ed. Paul Lavrakas and Michael Traugott, pp. 224–60. New York: Chatham House.
- Voss, D. Stephen, Andrew Gelman, and Gary King. 1995. "The Polls—Review. Preelection Survey Methodology: Details from Eight Polling Organizations, 1988 and 1992." *Public Opinion Quarterly* 59:98–132.
- Zukin, Cliff. "Sources of Variation in Published Election Polling: A Primer." Available online at <http://www.aapor.org/pdfs/varsource.pdf> (accessed July 5, 2005).

Members of the Ad Hoc Committee:

Glen Bolger, Public Opinion Strategies

Darren W. Davis, University of Notre Dame

Charles Franklin, University of Wisconsin - Madison

Robert Groves, Institute for Social Research, University of Michigan

Paul J. Lavrakas, Methodological Research Consultant

Mark S. Mellman, The Mellman Group

Philip Meyer, University of North Carolina

Kristen Olson, University of Nebraska-Lincoln

J. Ann Selzer, Selzer & Company

Michael W. Traugott (Chair),
Institute for Social Research,
University of Michigan

Christopher Wlezien, Temple University

"An Evaluation of the Methodology of the 2008 Pre-Election Primary Polls"

This report was prepared by the American Association for Public Opinion Research (AAPOR), Ad Hoc Committee on the 2008 Presidential Primary Polling.

Revised Version, April 2009.

Copyright is held by the AAPOR. Those wishing to quote from the report may do so if they properly acknowledge the source and do not alter the original work.



AMERICAN ASSOCIATION FOR
PUBLIC OPINION RESEARCH

American Association for Public Opinion Research
P.O. Box 14263
Lenexa, KS 66285-4263
Tele: (913) 895-4601
Fax: (913) 895-4652