Public Opinion Quarterly, Vol. 69, No. 3, Fall 2005, pp. 342–369

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.

doi:10.1093/poq/nfi044
Published by Oxford University Press on behalf of the American Association for Public Opinion Research 2005.

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 (NCPP) 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 understate 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 jth office (governor, senator) in the kth 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: $\text{odds ratio}_{ijk} = (r_{ijk} / d_{ijk}) \,/\, (R_{jk} / D_{jk})$ 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

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

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

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

$$= \frac{1}{n_i}\left(\frac{q}{p} + \frac{p}{q} + 2\right) = \frac{1}{n_i pq}(q^2 + p^2 + 2pq) = \tfrac{1}{n_i pq}.$$

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

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

$$\text{Variance}(A_{ijk}) = 1/n_i r_{ijk} d_{ijk} \qquad [2]$$

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 × .5357 × .4643) = .005, and its standard error is .073, so a 95 percent confidence interval includes 0 ± .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_{ijk} as a measure of accuracy: A_{ijk} 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 5 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 $\overline{A}$.[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 $\overline{A}$ were calculated using a jackknife replication method using VPLX (Fay 1998) and treating each election year's polls as simple random samples.

Table 1. Mean Predictive Accuracy in Polls for Three Presidential Elections

Election	% Republican in election	Mean % Republican in polls	*N* of polls	$\overline{A}$	Standard error of $\overline{A}$
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

	Percentage Preferring			Treatment of Undecided Voters					
				Column 1: No allocation		Column 2: Proportional allocation (or undecided dropped)		Column 3: Equal allocation	
Poll	Rep. r	Dem. d	Und. u	Poll margin = $r - d$	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:

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

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

Percentage Preferring				Treatment of Undecided Voters					
				Column 1: No allocation		Column 2: Proportional allocation (or undecided dropped)		Column 3: Equal allocation	
Poll	Rep. *r*	Dem. *d*	Und. *u*	Poll odds = *r* / *d*	*A*	Poll odds	*A*	Poll odds	*A*
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.—$A = \log \frac{r_i / d_i}{R/D}$. 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
Newsweek	1.0465	0.0558	11	13	7
ABC	1.0667	0.0749	12	5	12
Washington Post	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.dcpoliticalreport.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.

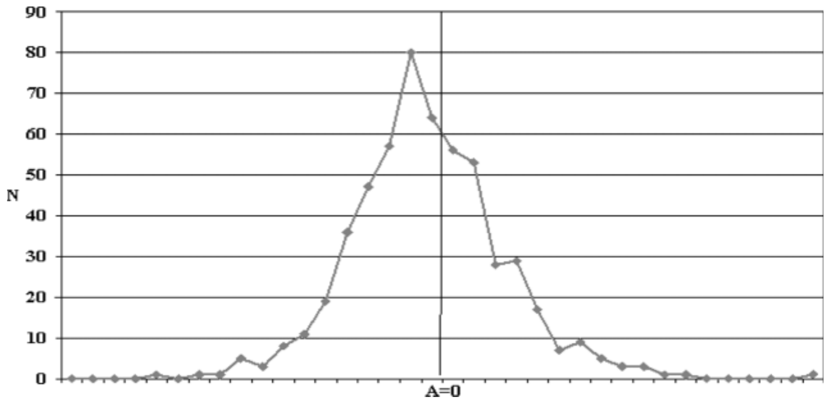

Figure 1. Frequency distribution of A_{ijk} for preelection polls for statewide offices in 2002.

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_{ijk} 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_{ijk} 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:

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

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

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

Solving for $\hat{r}_{ijk}$ 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	.0115	−.0332	−.0889	−.0350	−.0453
Polls	(.0176)	(.0361)	(.0266)	(.0192)	(.0104)
Democratic	−.1321	−.2373	−.2087	−.1259	−.0873
(or Dual)	(.0392)	(.0805)	(.0323)	(.0454)	(.0267)
Republican	.0762	.0731	.0755	.0360	.0625
Polls	(.0331)	(.1096)	(.0820)	(.0617)	(.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

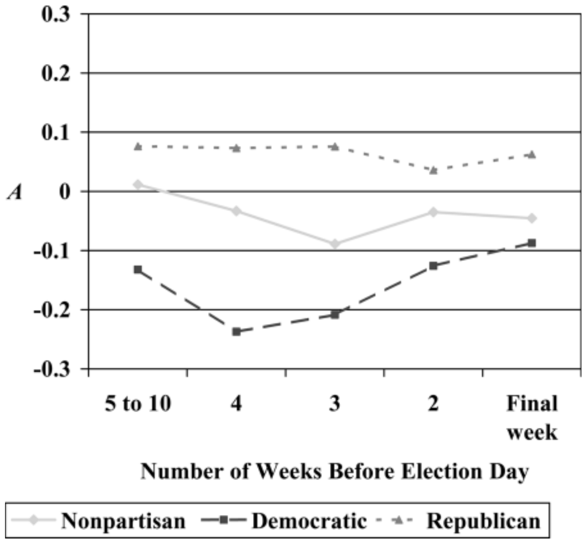

Figure 2. Mean predictive accuracy by number of weeks a poll was taken in advance of election 2002.

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).