CHAPTER 10

Improving Election Forecasting: Allocation of Undecided Respondents, Identification of Likely Voters, and Response Order Effects

Penny S. Visser, Jon A. Krosnick, Jesse F. Marquette, and Michael F. Curtin

> Everything's fine in Pollsville these days. Yes, we had a rocky start in the early days when the polls predicted that Landon would defeat Roosevelt in the election for U.S. president. And yes, we made a little mistake in 1948 when we predicted that Dewey would defeat Truman by 5 percentage points, when in fact Truman won by 5 percentage points instead. But we've solved those problems now. It all came down to bad sampling in those days, and we know how to sample now.

This is what we hear from many quarters these days. On 15 December 1996, for example, Michael Kagay (1996) proclaimed in the *New York Times* that the average error of the seven national polls conducted the weekend before the 1996 presidential election was a mere 2 percentage points. And in a press release issued on 13 February 1997, the National Council on Public Polls (NCPP) announced that the average error of the nine "major" polls done in 1996 forecasting that election was even smaller: 1.7 percentage points. That press release included Figure 10.1, illustrating that forecast errors made by the major polls were not so terrific in the 1930s but have been consistently within sampling error in recent years. And it is not just on this side of the Atlantic—the final NOP poll predicted that Labour would win 47 percent of votes in races for the British House of Commons in 1997, and the party in fact won 44.4

percent, resulting in forecast error of just over 2 percentage points. Everything's *fine* in Pollsville—really!

Well, maybe not. On 22 November 1996, Everett Carll Ladd published an article in the *Chronicle of Higher Education* entitled "The Election Polls: An American Waterloo." In it, he began:

> Election polling had a terrible year in 1996. Indeed, its overall performance was so flawed that the entire enterprise should be reviewed by a blue-ribbon panel of experts—from academe, commercial polling firms, and the news media—who should recommend ways to improve the accuracy of polling. (p. A52)

Ladd is a distinguished professor of political science at the University of Connecticut and director of the Roper Center for Public Opinion Research, certainly not someone biased against survey research or disposed to be overly critical of it. And Ladd was not alone. In the *New York Times* magazine of 15 December 1996, for example, editor Max Frankel said that "in 1996, the opinion polls were disturbingly wide of the mark" (p. 34). This sentiment was echoed by *New York Times* editorialist William Safire a year later: On 7 December 1997, he cited the "grievously misleading" preelection polls of 1996 as evidence that the U.S. Census Bureau should not look to sampling as a means of generating accurate estimates of the population.

One source of concern that Ladd, Frankel, and Safire all mentioned was that the final CBS/*New York Times* poll forecast a Clinton victory over Bob Dole of 18 percentage points, whereas the actual margin of victory was only 8

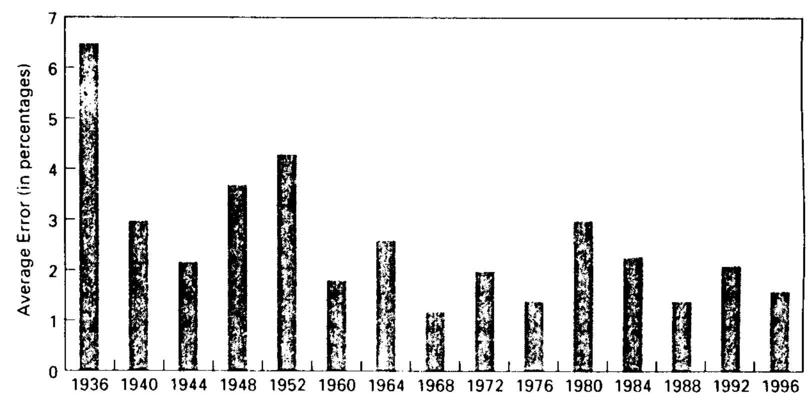

FIGURE 10.1 AVERAGE ERROR IN PRESIDENTIAL ELECTION FORECASTING POLLS

percentage points. Other concerns included the fact that in Britain the polls predicted a Labour victory in 1992, whereas the Conservatives actually won (Jowell et al. 1993). And of course there was the well-publicized 1989 race for governor of Virginia: The polls all predicted that Douglas Wilder would win by a margin of between 4 and 11 percentage points, whereas his actual victory margin was less than 1 percentage point.

How could there be so much error in these cases, when Kagay (1996) and the NCPP say there is almost no error in predictions? One answer involves the selection of polls to describe. Kagay and NCPP described the accuracy of only some polls. Yet more comprehensive analyses of poll accuracy have revealed quite a bit more error. For example, Crespi (1988) found the average error of 430 preelection polls to be 5.7 percentage points. More recently, Gelman and King (1993) reported an average error of 4.5 percentage points in forecasts of presidential races from 1984 through 1992.

These figures, all about 5 percentage points, are *average* errors in the predictions of each candidate's vote share. That means that on average in two-candidate races, one candidate's vote share will be overestimated by 5 percentage points, and the other candidate's share will be underestimated by 5 percentage points. This exaggeration of the difference between the two candidates of 10 percentage points on average is exactly the CBS/*New York Times* overestimation of Bill Clinton's victory margin in 1996. Thus, the poll's "mistake" was not an outlier but, rather, an *average* amount of error. Furthermore, in about half of races, according to Crespi's (1988) and Gelman and King's (1993) analyses, error will be *larger* than 5 percentage points.

Some observers will tell you that these errors are random and are well within sampling error. In fact, though, it appears that the errors in forecasts are systematically biased in one direction. Critics have at times charged that the polls overestimate the popularity of Democratic Party candidates, and some evidence might appear to be consistent with this. For example, all but one of the "major" polls discussed in the NCPP press release overestimated Bill Clinton's victory margin in 1996. And in 1992, all six of the "major" polls examined by the NCPP also overpredicted Clinton's victory margin. But in 1988, four of the five NCPP "major" polls overestimated George Bush's victory, and in 1972, all three "major" polls overpredicted Richard Nixon's victory margin. Instead of a bias in favor of Democratic candidates, then, these results suggest that the bias may be in favor of the leading candidate, no matter what his or her party affiliation. Although this pattern did not hold for all elections in the NCPP analysis, it did hold for many of them. Thus, it appears that the winner's margin of error is typically overpredicted (see also Gruca 1994; Panagakis 1989; Wright 1990, 1992, 1993).

So all is *not* well in Pollsville after all. In fact, we can do better, and we should do better. But how? What can be done to improve a method that has been so carefully fine-tuned over a period of decades? One answer to this question, we argue, lies in a little-recognized phenomenon that has been happening quietly in Columbus, Ohio, for nearly two decades.

The *Columbus Dispatch* Mail Surveys

As we reported in 1996 (Visser et al. 1996), the *Columbus Dispatch* mail surveys have been strikingly accurate in forecasting election outcomes since 1980, with an average error of only 1.6 percentage points. They have been substantially more accurate than telephone polls forecasting the same races conducted by the University of Akron (average error = 5.4 percentage points), the University of Cincinnati (average error = 4.9 percentage points), and the Gallup Organization (average error = 5.2 percentage points), each of which has average error rates comparable to the those reported by Crespi (1988) and Gelman and King (1993).

Interestingly, the *Dispatch*'s unusual accuracy is not a new phenomenon, as documented in Claude Robinson's (1932) book *Straw Votes*. Robinson reviewed the methods and findings of preelection polls attempting to forecast U.S. elections during the first three decades of the twentieth century. The *Columbus Dispatch* newspaper was polling back then, and with unusual accuracy for the time. Whereas some polls made predictions with average errors as large as 20 percentage points, the average error of the *Dispatch* forecasts was 7 percentage points (pp. 68–69). Only the *Chicago Tribune* performed better, with an average error of 6 percentage points. Sampling and questioning methods have changed quite a bit over the century, but the unusual accuracy of the *Dispatch* polls has not.

In our 1996 report, we detailed not only the superior accuracy of the *Dispatch* polls in the 1980s and 1990s but also their invulnerability to usual sources of error in preelection surveys. Telephone survey forecasts are less accurate when a race has received little publicity, when a lot of voters who go to the polls abstain in that race rather than voting in it, when the race is listed low on the ballot, and when voter turnout is lower (Crespi 1988).

Furthermore, we explored a number of possible explanations for the mail polls' accuracy. We found no evidence that weighting of the *Dispatch* samples improved accuracy or that interviewer effects compromised the accuracy of the telephone polls. We did identify four partial explanations, however:

1. The mail surveys have typically involved samples of about 1,600 respondents, in contrast to samples of 500 to 900 for the telephone polls.
   Error was reduced by almost half a percentage point on average by the increased sample size.

2. The mail survey procedure recruited samples of respondents that more closely resembled the actual voting population, perhaps because the act of completing a self-administered questionnaire is in many ways comparable to the act of voting.
3. The mail surveys did not offer respondents an opportunity to say they had not decided how they would vote in a race, whereas the telephone surveys invited and routinely received many such responses from people. Having to allocate those "undecided" voters introduced error into the telephone polls.
4. The mail surveys simply asked people for whom they will vote in each race, whereas the telephone surveys used a variety of other question wordings (e.g., "Which candidate would you like to see win?") and preceded the vote choice questions with other questions on related topics. By mirroring the Election Day ballot precisely, the mail surveys avoided error attributable to question wording and question order effects (for other evidence of such question order effects, see also Hoek and Gendall 1997; Kagay 1992, 101).

This Chapter

Inspired partly by Crespi (1988), our goal in this chapter is to take a comprehensive and systematic approach to testing various hypotheses about the errors in election forecasts. We report results from our ongoing investigation of the differences between mail and telephone survey methods. In particular, we address a series of new questions:

1. In our 1996 report, we focused only on statewide elections in Ohio. What about the accuracy of mail survey forecasts of the outcomes of referenda and local races? Did the mail surveys forecast these outcomes better than telephone surveys?
2. How do various methods of allocating undecided respondents in the telephone surveys compare in terms of effectiveness? Can we identify an optimal method?
3. How do various methods of screening out telephone respondents who are unlikely to vote compare in terms of effectiveness?
4. Even if the mail and telephone surveys differ in accuracy, do they track similar trends in candidate support over time?
5. Does the order in which candidates' names are presented to respondents affect answers?

We also expanded the scope of our investigation of mail and telephone survey accuracy by looking not just at the forecasts generated immediately before the elections, but at those generated weeks and months beforehand as well. In conducting the research to be described, we tested various hypotheses about the conditions under which forecasts are more and less accurate, in order to develop techniques for improving forecast accuracy, and we ended up with many intriguing results.

In the end, however, the implications of this work are much broader. Elections offer a unique opportunity for assessing the validity of survey data because each Election Day provides a benchmark of truth against which to measure forecast accuracy. Such benchmarks are difficult if not impossible to obtain in many areas of social science research, and when they are obtained, there is sometimes error in their measurement. But we can have strong confidence in the benchmark used here, so we can have confidence in the implications of our findings as well. As a result, the findings we report here shed new light on methods by which the precision of all survey data can be enhanced.

Mail Survey Accuracy in New Contests

Table 10.1 displays the accuracy of the mail survey forecasts for the statewide candidate races that we already examined (in column 1, with the addition of a new figure for the 1996 U.S. presidential race), plus forecasts for new contests that occurred before June 1997: local candidate races (in column 2), statewide referenda (in column 3), and local referenda (in column 4). At the bottom of each column is the average error for races in the column. As is apparent, the statewide candidate races (average error = 1.5 percentage points) were predicted more accurately than were the other contests. The average error for the local candidate races was 2.7 percentage points, as compared to 5.8 percentage points for the statewide referenda and 3.9 percentage points for the local referenda. Some of the errors for referenda were quite substantial, peaking at 12.8 percentage points. We do not yet have evidence to explain why the mail poll forecasts are not as good in these cases, and we look forward to future research exploring this puzzle.

The University of Akron did not conduct surveys on the local candidate races and referenda addressed by the *Dispatch* polls. However, both outfits conducted surveys forecasting some of the same statewide referenda. As Table 10.2 shows, the average error for *Dispatch* forecasts of these referenda was 5.4 percentage points, compared to 7.2 percentage points for the telephone surveys. Thus, it appears that although referenda were more difficult to forecast than candidate races using either method, the mail surveys still outperformed the telephone surveys when doing so. However, in one of the six referenda examined

TABLE 10.1 ACCURACY OF THE FINAL *COLUMBUS DISPATCH* FORECASTS (IN PERCENTAGE POINTS)

	STATEWIDE CANDIDATE RACES		LOCAL CANDIDATE RACES		STATEWIDE REFERENDA		LOCAL REFERENDA	
YEAR	Race	Average Error	Race	Average Error	Issue	Average Error	Issue	Average Error
1980	President	1.4						
1982	Governor	0.6						
	Attorney General	1.1						
	Auditor	0.8						
	Treasurer	0.8						
	Secretary of State	4.0						
	U.S. Senate	0.7						
1984	President	0.7						
1986	Governor	0.5						
	Attorney General	1.2						
	Auditor	1.3						
	Treasurer	1.4						
	Secretary of State	0.5						
	U.S. Senate	3.2						
1988	President	0.7						
	U.S. Senate	0.5						
1989							Amendment	0.6
							COTA tax	2.9
							Park District	4.4
							Children's Services	8.3
1990	Governor	0.2			Gambling	0.7		
	Attorney General	4.5			Housing	3.4		
	Auditor	3.7			Tax increase	6.4		
	Treasurer	1.6						
	Secretary of State	0.9						
1991			Mayor	4.7				
			City Council	0.8				
1992	President	1.5			Convention	11.2		
					Federal term limits	9.4		
					State term limits	7.3		
					State term limits	6.6		
					Safety warnings	8.6		
1993			City Clerk	1.9			Amendment	7.3
			City Council	1.5			Levy	3.2
			City Attorney	4.7			City Charter 1	6.1
							City Charter 2	1.5
1994	Governor	0.6			Death penalty appeals	3.8		
	Attorney General	2.6			Victim rights	8.1		
	Auditor	0.5			Tuition credits	6.8		
	Treasurer	1.5			Soft drink tax	12.8		
	Secretary of State	0.9						
	U.S. Senate	1.8						
1995					Amendment 1	0.9		
					Amendment 2	4.1		
					COTA tax	2.7		
					Mental health tax	4.8		
1996	President	2.7			Gambling	0.2		
1997							Arena tax	0.9
Average		1.5		2.7		5.8		3.9

TABLE 10.2 COMPARING THE *COLUMBUS DISPATCH* AND UNIVERSITY OF AKRON FORECAST ACCURACY ON REFERENDA (IN PERCENTAGES)

YEAR	STATEWIDE REFERENDA	*DISPATCH* AVERAGE ERROR	AKRON AVERAGE ERROR
1990	Gambling	1.6	1.6[a]
1994	Death penalty appeals	3.8	8.3[b]
	Victim rights	8.1	0.9[b]
	Tuition credits	6.8	10.7[b]
	Soft drink tax	12.8	17.2[b]
1996	Gambling	0.2	4.5[a]
Average		5.4	7.2

[a]Using data from respondents who were very or somewhat interested in the election and definitely intended to vote; dropping undecided
[b]Using data from all respondents; dropping undecided

(Victim rights, in 1994), the telephone survey was much more accurate than the mail poll, so the superiority of the mail polls is not universal.

An examination of the direction of the forecast errors uncovers the expected overprediction of the winner's share of the vote. In the thirty-three statewide forecasts of candidate races by the telephone polls, 73 percent overpredicted the winner's share of the vote, and only 27 percent did not. Across the thirty-three races, the margin of victory was overpredicted by 2.65 percentage points. In the six statewide forecasts of referenda by the telephone polls, 67 percent overpredicted the winner's share of the vote, and only 33 percent did not. Across the six referenda, the margin of victory was overpredicted by 1.97 percentage points. Thus, these surveys manifest the same systematic error as national surveys.

The same bias was apparent in the mail surveys, but more weakly in the candidate races. In the thirty-three *Dispatch* forecasts of candidate races, the winner's share of the vote was overpredicted in 64 percent of the cases and was not overpredicted in 36 percent of the cases. On average across the thirty-three races, the margin of victory was overpredicted by 0.67 percentage points. In the twenty-five mail survey forecasts of referenda, the margin of victory was overpredicted in 76 percent of the cases and was not overpredicted in only 24 percent of the cases. On average across twenty-five referenda, the margin of victory was overpredicted by 2.8 percentage points.

Allocating Undecided Respondents

Can the accuracy of the telephone forecasts be improved? As mentioned earlier, our 1996 investigation revealed that one source of inaccuracy in these forecasts involved allocation of respondents who said they had not yet decided how to vote in a particular contest. Whereas the final mail surveys did not solicit "undecided" responses, and respondents volunteered them extremely rarely, the telephone surveys have routinely gotten relatively large numbers of such responses, even less than a week before Election Day. As Table 10.3 shows, the University of Akron telephone surveys have found proportions of "undecideds" ranging from 3.2 percent to 73.4 percent, averaging 19.9 percent in the early September surveys, 36.9 percent in the late September surveys, and 19.0 percent in the early October surveys.

Various methods have been used to handle these "undecided" responses (e.g., Crespi 1988; Daves and Warden 1995; Fenwick et al. 1982), based upon different presumptions about the meanings of these responses. One possibility is

TABLE 10.3 PERCENTAGE OF UNDECIDED RESPONDENTS IN THE UNIVERSITY OF AKRON SURVEYS (IN PERCENTAGES)

		EARLY SEPTEMBER		LATE SEPTEMBER		EARLY OCTOBER	
YEAR	RACE	Percent of Undecided	Percent Who Lean toward One Candidate	Percent of Undecided	Percent Who Lean toward One Candidate	Percent of Undecided	Percent Who Lean toward One Candidate
1988	President	18.1	10.9	20.9	11.3	13.1	6.5
	U.S. Senate	17.4	8.2	17.3	7.3	13.0	6.2
1990	Governor			37.4	27.1	30.0	19.7
	Attorney General			61.2		37.7	
	Auditor			44.8		38.2	
	Treasurer			73.4		47.7	
	Secretary of State			38.9		19.5	
	Gambling amendment			27.5		9.0	
1992	President	24.2	13.5	27.7	18.4	18.2	13.0
1994	Governor			38.5	11.5	13.9	7.7
	Attorney General			33.0	9.2	12.9	5.4
	Auditor			40.5	8.7	16.9	5.6
	Secretary of State			32.5	9.8	13.7	5.4
	U.S. Senate			31.4	9.8	11.6	6.1
	Issue 1			40.2		8.5	
	Issue 2			28.2		3.2	
	Issue 3			57.8		14.9	
	Issue 4			40.8		3.8	
1996	President					30.2	17.2
	Gambling amendment					23.2	
Average		19.9	10.9	36.9	12.6	19.0	9.3

that people who say they are undecided do in fact have no real candidate preference in a race. Therefore, if these individuals were pressed to express a candidate preference, any answer they would give would be a meaningless nonattitude (Converse 1964) and would only reduce forecast accuracy. Based upon this logic, respondents should be encouraged to say they are undecided, and people who give such responses should be presumed not to vote in a race.

However, many of the people who claim to be undecided may in fact have real preferences but may be reluctant to report them. For example, some might not want to do the cognitive work required to think about and choose between candidates, and opting for the "don't know" option is a way to avoid that work (Krosnick 1991). Alternatively, some people might have very little information with which to justify their preferences, so they might lack confidence to feel comfortable expressing them. Other people might be reluctant to express their preferences because of "spiral of silence" processes (Noelle-Neumann 1984); that is, they may believe that their candidate preferences are not shared by most other people, either as the result of conversations with others or by learning about opinion poll results through the news media. Admitting their preferences would identify these people as being out of step with the majority of voters, so they may claim not to favor any candidate. Consistent with this last possibility is evidence that the number of people claiming to be undecided drops significantly when survey respondents can "vote" in a secret "ballot box" instead of telling an interviewer how they will vote, and the apparent popularity of "losing" candidates increases on secret ballots (Perry 1979).

In all these cases, encouraging individuals to express their opinions might yield valid data that would increase forecast accuracy. One way to do so in telephone surveys involves asking respondents who initially say they are undecided whether they lean toward one of the candidates in a race. For most races, the University of Akron did exactly this. As Table 10.3 shows, sizable proportions of people indicated leaning, ranging from 5.4 percent to 27.1 percent. If we examine the races where surveying was done in both early and late September (the 1988 and 1992 races), there is no clear decline in "undecided" responses (averaging 19.9 percent and 22.0 percent in early and late September, respectively), but these responses were notably less common in early October than in late September for races where polling was done at both times (averaging 36.9 percent and 18.1 percent in late September and early October, respectively). If people had insight into their final decisions, making use of these reports might increase the accuracy of forecasts. Table 10.4 reports all possible tests of this hypothesis with the available Akron data from three waves of surveying, in early September, early October, and late October.

When undecided people who declined to indicate a leaning were dropped altogether, taking people's reports of leaning seriously consistently increased forecast

TABLE 10.4 IMPACT OF ALLOCATION STRATEGIES ON FORECAST ERROR IN THE UNIVERSITY OF AKRON SURVEYS ON CANDIDATE RACES (IN PERCENTAGE POINTS)

	EARLY SEPTEMBER				EARLY OCTOBER				LATE OCTOBER			
	UNDECIDED DROPPED		UNDECIDED ALLOCATED EQUALLY		UNDECIDED DROPPED		UNDECIDED ALLOCATED EQUALLY		UNDECIDED DROPPED		UNDECIDED ALLOCATED EQUALLY	
	Leaners Dropped	Leaners to Candidate	Leaners Allocated Equally	Leaners to Candidate	Leaners Dropped	Leaners to Candidate	Leaners Allocated Equally	Leaners to Candidate	Leaners Dropped	Leaners to Candidate	Leaners Allocated Equally	Leaners to Candidate
1988												
President	3.1	3.2	3.6	3.3	1.3	0.8	0.4	0.5	4.3	3.5	3.1	2.9
U.S. Senate	5.6	5.2	3.4	4.1	7.1	1.9	4.7	4.3	5.1	4.5	3.6	3.7
1990												
Governor					7.6	3.1	2.4	1.2	1.5	0.6	0.7	1.1
1992												
President					5.4	7.6	4.5	6.2	2.9	1.1	1.0	1.5
1996												
President									6.0	5.6	4.3	4.0
Average	4.4	4.2	3.5	3.7	5.4	3.4	3.0	3.1	4.0	3.1	2.5	2.6

accuracy. The first two columns of Table 10.4 show that for the two races forecast with early September data, the average error was 4.4 percentage points when the leaning "undecideds" were dropped from forecasts and 4.2 percentage points when the leaning "undecideds" were allocated to the candidates. The same average figures for the early October forecasts were 5.4 and 3.4 percentage points, respectively, again suggesting a gain in accuracy when allocating leaning "undecideds" to the candidates (see columns 5 and 6). Finally, the same pattern is apparent in the late October surveys, with an average error of 4 percentage points when leaning "undecideds" were dropped and 3.1 percentage points when they were allocated to the candidates (see columns 9 and 10). This improvement in accuracy is consistent with the claims that some people who initially say they are undecided do indeed have real preferences and that the people who say they don't lean also don't vote.

Another possible strategy is based upon a different sort of speculation about the behavior of nonleaning undecided respondents. These individuals may be drawn to their precincts in order to cast votes in other races, about which they do have strong preferences. Finding themselves there, they may feel the obligation to vote in all the other races on the ballot and may do so essentially randomly (e.g., Converse 1964). Consequently, we might see an improvement in forecast accuracy if we were to allocate the nonleaning undecided respondents equally to the various candidates instead of dropping them (e.g., Crespi 1988, 22).

To test this idea, we can first compare column 2 of Table 10.4 to column 4, where the nonleaning undecided respondents were allocated equally to the candidates. In fact, this did improve accuracy, as the average error for column 2 is larger (4.2 percentage points) than the average error for column 4 (3.7 percentage points). Similar comparisons for the October surveys reveal the same gain in accuracy: 3.4 versus 3.1 percentage points for early October, and 3.1 versus 2.6 percentage points for late October. All this is consistent with the claims that people's answers to questions about leaning are valid and that people who say they are undecided and don't lean vote randomly.

Before concluding this exercise, we explored one other possibility: that all undecided respondents, those who said they leaned toward a candidate and those who said they did not lean, might vote randomly and should therefore be allocated equally among the candidates. To our surprise, this approach yielded the most accurate forecasts of all (see Table 10.4). The average in column 3 (3.5 percentage points) is the lowest of the early September columns. Likewise, the average error in column 7 (3.0 percentage points) is the lowest of the early October columns. And the average error in column 11 (2.5 percentage points) is the lowest of the late October columns.

We were able to test this approach more thoroughly by including races that were omitted from Table 10.4 because undecided respondents had not been

asked whether they leaned, or because they were given the option of saying they would not vote in the race. The larger set of races shown in Table 10.5 again offers strong support for random allocation of all undecided respondents. The average errors for columns 2, 4, and 6, where such allocation was done (4.7, 4.2, and 3.1 percentage points, respectively), were smaller than the average errors for columns 1, 3, and 5, where the undecided respondents were dropped altogether (5.3, 7.5, and 5.9 percentage points, respectively). Averaging across all the columns, the average error when undecided respondents were dropped was 6.2 percentage points compared to 4.0 percentage points when the undecided respondents were allocated equally to the candidates.

Indeed, this method even appeared to improve the accuracy of forecasts of referenda. As Table 10.6 shows, the average error of forecasts dropping the undecided respondents was 4.6 percentage points compared to 2.6 percentage points when half of the undecided respondents were presumed to vote yes and the other half were presumed to vote no. Because so few referenda could be examined here, additional studies of this sort are clearly needed before we can have strong

TABLE 10.5 IMPACT OF ALLOCATION STRATEGIES ON FORECAST ACCURACY IN THE UNIVERSITY OF AKRON SURVEYS ON CANDIDATE RACES (IN PERCENTAGE POINTS)

	EARLY SEPTEMBER		EARLY OCTOBER		LATE OCTOBER	
	DROP UNDECIDED	ALLOCATE EQUALLY	DROP UNDECIDED	ALLOCATE EQUALLY	DROP UNDECIDED	ALLOCATE EQUALLY
1988						
President	3.1	3.6	1.3	0.4	4.3	3.1
U.S. Senate	5.6	3.4	7.1	4.7	5.1	3.6
1990						
Governor			7.6	2.4	1.5	0.7
Attorney General			9.7	3.7	7.5	4.7
Auditor			19.8	9.7	15.5	8.5
Secretary of State			2.6	2.8	1.2	1.6
Treasurer			6.1	5.3	8.8	0.1
1992						
President	7.2	7.2	5.4	4.5	2.9	1.0
1996						
President					6.0	4.3
Average	5.3	4.7	7.5	4.2	5.9	3.1

TABLE 10.6 IMPACT OF ALLOCATION STRATEGIES ON FORECAST ACCURACY IN THE UNIVERSITY OF AKRON SURVEYS ON REFERENDA (IN PERCENTAGE POINTS)

	EARLY OCTOBER		LATE OCTOBER	
	Undecided Dropped	Undecided Allocated Equally	Undecided Dropped	Undecided Allocated Equally
1990 STATE POLL				
Gambling amendment	0.3	3.1	0.9	0.3
1996 STATE POLL				
Gambling amendment			12.6	4.3
Average	0.3	3.1	6.8	2.3

confidence in this pattern; but the pattern is certainly consistent with the evidence regarding candidate races.

It is tempting to infer from this that people who said they leaned toward one candidate or another did not really do so, voting randomly instead. However, there is another possible explanation for these results. It may be that people who leaned toward one candidate did, more often than not, vote for that candidate. But because of spiral of silence pressures, some of the respondents who claimed to be nonleaning undecideds or who claimed to support the most popular candidate might in fact have voted for an unpopular candidate. Correcting for this bias can be accomplished by mixing in some 50/50 forecasts along with all the other survey data, and this is what our last allocation strategy did. Although this may be the best approach to forecasting, its success does not necessarily mean that people who claimed to lean toward one candidate in fact voted randomly.

In 1994, the Akron surveys offered respondents the option to say they would not vote in each particular race, and the proportions of people saying so ranged from 10 percent to 28 percent (see Table 10.7). These respondents were asked whether they leaned toward one candidate or another, and relatively small but nontrivial numbers said they did. Therefore, although it might be best to presume that these respondents will indeed skip the race, it might be preferable to incorporate answers to the leaning question in forecasts if they have some validity and if those respondents might in fact vote. We decided to test the impact of such a move, and Table 10.8 displays the results of these tests. Specifically, we tried allocating leaners to candidates, dropping leaners, dropping nonleaning undecideds and people who said they would skip the race, and allocating equally the nonleaning undecideds and people who said they would skip the race.

The implication of these tests is apparent in the lower-right-hand corner of

TABLE 10.7 PROPORTION OF RESPONDENTS IN THE UNIVERSITY OF AKRON SURVEYS WHO SAID THEY WOULD SKIP EACH RACE (IN PERCENTAGES)

		LATE SEPTEMBER		EARLY OCTOBER	
YEAR	RACE	Total Number of "Skip Race"	Number of "Skip Race" Who Lean toward One Candidate	Total Number of "Skip Race"	Number of "Skip Race" Who Lean toward One Candidate
1994	Governor			14	5
	Attorney General	18		14	3
	Auditor	26		28	4
	Secretary of State	19		23	4
	U.S. Senate			10	2

the table, where the smallest average error appears. This was obtained by allocating all people who said they leaned toward a candidate to that candidate and allocating equally all respondents who were nonleaning undecideds or were nonleaning people who said they would not vote in the race. This reinforces the notion that answers to leaning questions do have some validity, so people who say "don't know," "undecided," or "skip the race" should be pressed to express candidate preferences. The worst inaccuracy of all in Table 10.8 occurred when all people who said "undecided" or "skip the race" were dropped entirely from the predictions, which is the approach taken these days by many survey research groups.

Adding Random Responses to the Mail Survey Forecasts

Although the mail surveys conducted just before Election Day did not offer respondents an explicit option to say they were "undecided," the earlier mail surveys did, and some respondents regularly expressed indecision (see Table 10.9). If we focus just on those races where polling was done at all four time points (the 1990 races for governor, attorney general, secretary of state, and the gambling amendment referendum), it is clear that "undecided" rates were highest in January (averaging 18.2 percent, 12.0 percent, 12.3 percent, and 10.2 percent for January, June, September, and October, respectively). When examining all races where polling was done in June, September, and October, we noted that "undecided" responses declined from June (14.2 percent) and September (14.4 percent) to October (12.4 percent). When examining all races where polling was done in September and October, we noted that "undecided" responses were more common in September than in October (averaging 15.9 percent and 12.8 percent, respectively). Thus, there appears to be a decline in indecision over the

TABLE 10.8 IMPACT OF ALLOCATION STRATEGIES ON FORECAST ACCURACY IN THE 1994 UNIVERSITY OF AKRON SURVEYS ON CANDIDATE RACES (IN PERCENTAGE POINTS)

	DROP NONLEANING "SKIP THE RACE"						ALLOCATE NONLEANING "SKIP THE RACE" EQUALLY					
	DROP NONLEANING UNDECIDED			ALLOCATE NONLEANING UNDECIDED EQUALLY			DROP NONLEANING UNDECIDED			ALLOCATE NONLEANING UNDECIDED EQUALLY		
RACE	Drop All Leaners	Undecided Leaners to Cand.	Undecided and "Skip" Leaners to Cand.	Allocate Leaners Equally	Undecided Leaners to Cand.	Undecided and "Skip" Leaners to Cand.	Drop All Leaners	Undecided Leaners to Cand.	Undecided and "Skip" Leaners to Cand.	Allocate Leaners Equally	Undecided Leaners to Cand.	Undecided and "Skip" Leaners to Cand.
Governor	7.4	6.8	6.1	4.7	4.9	4.7	4.8	4.6	5.0	5.0	4.6	3.9
Attorney General	8.4	8.3	7.2	7.3	7.4	6.5	7.2	7.2	6.5	6.5	6.5	6.0
Auditor	14.4	14.2	13.5	12.9	12.9	12.5	12.3	12.2	12.1	11.7	11.6	11.5
Secretary of State	8.5	8.4	7.2	4.4	4.7	3.9	2.3	2.3	2.3	.0	.2	.3
U.S. Senate	1.8	1.8	2.0	1.4	1.1	1.4	1.3	1.4	1.4	2.7	2.3	2.4
Average	8.1	7.9	7.2	6.1	6.2	5.8	5.6	5.5	5.5	5.2	5.0	4.8

course of the election campaigns. The proportion of undecideds was often large, reaching a maximum of 50.3 percent, so handling of these responses can, in principle, have a substantial impact on forecast accuracy.

As Table 10.10 displays, the accuracy of forecasts was indeed improved when undecided respondents were allocated equally to the candidates. Specifically, the average error across all surveys was 3.9 percentage points when the undecided responses were allocated equally to the candidates, compared to 4.6 percentage points when the undecided responses were dropped instead of being allocated. This gain in accuracy was also apparent for the referenda shown in Table 10.11.

TABLE 10.9 UNDECIDED RESPONDENTS IN THE *COLUMBUS DISPATCH* SURVEYS (IN PERCENTAGE POINTS)

YEAR	RACE	JANUARY	JUNE	EARLY SEPTEMBER	EARLY OCTOBER
1988	President			9.0	
	U.S. Senate			8.0	
1990	Governor	15.3	8.4	7.4	6.9
	Attorney General	28.6	16.4	18.3	16.8
	Auditor		14.1	13.9	12.0
	Treasurer		23.2	23.6	21.4
	Secretary of State	17.5	10.6	11.0	9.1
	Gambling amendment	11.2	12.7	12.4	8.0
	Housing amendment				23.5
	Tax amendment				7.7
1991	Mayor			22.4	16.3
1992	President			8.8	
	Issue 1			10.6	
	Issue 2			7.5	
	Issue 3			8.0	
	Issue 4			8.0	
	Issue 5			13.1	
1993	City Clerk			50.3	
	City Attorney			39.8	
	Amendment			16.9	
	Levy			13.2	
1994	Governor			11.6	7.7
	Attorney General			17.8	12.0
	Auditor			26.2	20.2
	Treasurer			27.0	22.1
	Secretary of State			14.1	12.3
	U.S. Senate			8.6	5.6
	Issue 1				19.8
	Issue 2				5.3
	Issue 3				26.6
	Issue 4				4.2
1996	Gambling amendment			8.6	8.2
Average		18.2	14.2	16.0	13.3

TABLE 10.10 IMPACT OF ALLOCATION STRATEGIES ON *COLUMBUS DISPATCH* FORECAST ACCURACY FOR CANDIDATE RACES (IN PERCENTAGE POINTS)

	JANUARY		MAY		EARLY SEPTEMBER		EARLY OCTOBER	
	Drop Undecided	Allocate Equally	Drop Undecided	Allocate Equally	Drop Undecided	Allocate Equally	Drop Undecided	Allocate Equally
1988 STATE POLL								
President					2.8	3.0	3.2	3.5
U.S. Senate					1.5	2.0	2.1	1.5
1990 STATE POLL								
Governor	0.9	0.2	3.8	3.0	4.0	3.3	1.4	0.9
Attorney General	10.9	7.8	1.3	1.0	10.3	8.3	5.7	4.8
Auditor			13.2	10.9	4.0	3.1	6.5	5.3
Secretary of State	4.3	4.1	3.3	3.3	0.8	1.0	0.4	0.7
Treasurer			2.4	4.1	1.9	3.7	3.5	4.8
1991 CITY POLL								
Mayor					6.4	4.5	5.0	3.9
City Council								
1992 STATE POLL								
President					10.3	8.7		
1993 CITY POLL								
Clerk					10.9	8.9		
Attorney					9.5	2.1		
City Council								
1994 STATE POLL								
Governor					8.5	8.0	0.8	1.8
Attorney General					1.9	1.8	2.7	0.5
Auditor					0.3	2.5	1.1	0.9
Secretary of State					6.6	3.6	5.2	2.7
Treasurer					1.4	5.1	1.1	3.7
U.S. Senate					5.7	4.2	5.6	4.5
Average	5.4	4.0	4.8	4.5	5.1	4.3	3.2	2.8

The average error when undecided respondents were dropped was 9.1 percentage points, compared to 8.7 percentage points when half the undecided respondents were presumed to vote yes and the other half were presumed to vote no.

We went one step further, testing whether the final mail survey forecasts could be improved by adding more random responses. As Table 10.12 shows, the average discrepancy between the actual election outcome and an even distribution of votes for all candidates was 11.3 percentage points, whereas the average discrepancy between the final *Dispatch* forecasts of these elections and an even distribution of votes was 12.0 percentage points. Thus, the *Dispatch* forecasts were farther from evenly distributed than were the actual election outcomes, which is consistent with our earlier demonstration that the forecast errors overpredicted the winner's share of the votes more often than not. Therefore, adding a bit more random responses to the *Dispatch* data would enhance the accuracy of the final forecasts further. As the last column of Table 10.12 shows, this strategy would have increased forecast accuracy for 76 percent of the candidate races and decreased accuracy for only 24 percent of them.

To assess how much more randomness would be optimal, we executed an

TABLE 10.11 IMPACT OF ALLOCATION STRATEGIES ON *COLUMBUS DISPATCH* FORECAST ACCURACY FOR REFERENDA (IN PERCENTAGE POINTS)

	JANUARY		MAY		EARLY SEPTEMBER		EARLY OCTOBER	
	Drop Undecided	Allocate Equally	Drop Undecided	Allocate Equally	Drop Undecided	Allocate Equally	Drop Undecided	Allocate Equally
1990 STATE POLL								
Gambling amendment	3.4	4.4	9.6	9.9	15.6	15.1	2.3	3.1
Housing							3.8	2.3
Tax							7.7	4.0
1992 STATE POLL								
Issue 1					22.3	18.7		
Issue 2					11.8	9.7		
Issue 3					10.6	8.3		
Issue 4					10.4	8.1		
Issue 5					58.1	54.0		
1993 CITY POLL								
Amendment					12.9	8.4		
Levy					1.3	0.7		
1994 STATE POLL								
Issue 1							6.9	3.3
Issue 2							9.4	7.1
Issue 3							0.8	4.7
Issue 4							17.5	17.1
1996 STATE POLL								
Gambling amendment					5.0	5.6	6.1	6.6
Average	3.4	4.4	9.6	9.9	16.4	14.3	6.8	6.0

TABLE 10.12 IMPACT OF ADDING 50/50 RESPONSES ON *COLUMBUS DISPATCH* FORECAST ACCURACY IN THE STATEWIDE CANDIDATE RACES (IN PERCENTAGE POINTS)

YEAR	RACE	AVERAGE DISCREPANCY BETWEEN ACTUAL ELECTION RESULTS AND AN EVEN DISTRIBUTION OF VOTES	AVERAGE DISCREPANCY BETWEEN DISPATCH POLL RESULTS AND AN EVEN DISTRIBUTION	ACCURACY IMPROVED BY ADDING 50/50 RESPONSES?
1980	President	17.7	16.1	no
1982	Governor	10.0	11.5	yes
	Attorney General	20.6	20.9	yes
	Auditor	2.2	3.0	yes
	Treasurer	18.2	18.9	yes
	Secretary of State	19.3	19.6	yes
	U.S. Senate	23.9	24.0	yes
1984	President	9.4	10.0	yes
1986	Governor	10.6	10.1	no
	Attorney General	9.8	11.0	yes
	Auditor	16.5	17.8	yes
	Treasurer	4.9	6.3	yes
	Secretary of State	9.7	10.2	yes
	U.S. Senate	12.5	9.3	no
1988	President	5.5	4.9	no
	U.S. Senate	7.0	7.8	yes
1990	Governor	5.7	5.9	yes
	Attorney General	0.0	4.5	yes
	Auditor	2.8	6.5	yes
	Treasurer	9.5	11.1	yes
	Secretary of State	3.0	3.9	yes
1992	President	7.9	9.4	yes
1994	Governor	25.6	26.2	yes
	Attorney General	1.4	1.2	no
	Auditor	8.5	9.0	yes
	Treasurer	19.6	19.2	no
	Secretary of State	14.8	15.7	yes
	U.S. Senate	17.3	17.0	no
1996	President	14.8	17.4	yes
	Average	11.3	12.0	

iterative procedure using the twenty statewide and local candidate races gauged by final *Dispatch* surveys between 1988 and 1996. The average error in forecasts of these races was 1.75 percentage points. In the first step of iteration, we added 1 percentage point to the distribution of votes predicted for each candidate and then repercentaged to obtain a new set of forecasts. This yielded an average error of 1.67 percentage points. Then, we added another percentage point to the number of votes predicted for each candidate and repercentaged. This yielded an average error of 1.65 percentage points. When we went another step and added an additional percentage point to each candidate's share, the average error rose to 1.7 percentage points. Thus, the optimal amount of additional random voting to add was between 2 percent and 3 percent.

This improvement in accuracy could occur for a number of possible reasons. First, a few people who do ultimately vote randomly may not respond to the *Dispatch* surveys. Alternatively, a few *Dispatch* respondents may be reluctant to admit supporting unpopular candidates and instead claim to support more popular ones. Because the *Dispatch* survey is so clearly confidential and anonymous, we suspect the former of these two possible explanations is more likely than the latter.

Eliminating Respondents Unlikely to Vote

One reason for concern about telephone survey forecasts of election outcomes is the fact that some adults who are interviewed will not in fact vote. Therefore, researchers must identify and eliminate these respondents from forecasts. Various approaches to this task have been taken (see Crespi 1988, 79; Freedman and Goldstein 1996; Petrocik 1991; Traugott and Tucker 1984; Voss, Gelman, and King 1995), and the University of Akron's approach has involved three filters: reported registration status, reported likelihood of voting, and reported interest in the election. Specifically, forecasts have been based upon only those respondents who said they were registered to vote, said they would definitely vote in the election, and said they were very interested or somewhat interested in the election.

It certainly seems wise to ignore data from telephone survey respondents who are unlikely to vote. But it is not obvious that basing the filtering process on self-reports is best. Surveys typically generate rates of predicted turnout that significantly exceed actual rates (Clausen 1968; Traugott and Katosh 1979). In the case of the 1996 University of Akron survey, the proportion of respondents who reported being registered to vote matched the actual proportion quite well (83 percent of respondents said they were registered to vote, whereas the actual proportion of Ohio adults registered to vote that year was 82 percent). Self-reported intention to vote, however, vastly overestimated actual turnout. Fully 97.6 percent of respondents who said they were registered said they would definitely (85.5 percent) or probably (12.1 percent) vote in the election, whereas in fact 68 percent of registered Ohio adults actually participated in the 1996 election.

In order to present themselves to interviewers as responsible citizens, people may be biased toward overreporting likely turnout. Yet these reports may have been quite accurate, and the overestimation of turnout rates may have occurred because members of the survey sample who were not interviewed were especially unlikely to vote (Clausen 1968). Had these latter individuals been interviewed, the total sample's turnout estimates may have been quite a bit more accurate. Therefore, eliminating some telephone survey respondents may actually have reduced forecast accuracy instead of improving it. We set out to test this possibility.

In nearly all University of Akron surveys, respondents who said they were not registered to vote were not asked their candidate preferences at all. But in 1996 all respondents were asked about their preferences. As Table 10.13 shows, forecasts from all respondents yielded an average error of 4.6 percentage points, compared to 2.8 percentage points when respondents who said they were not registered to vote were dropped. Although this test is based only upon a single race, this result does suggest that accuracy can be improved by focusing only on respondents who said they were registered.

Using the final telephone surveys, we next explored whether further elimination of respondents improved accuracy. Specifically, in Table 10.14, we compared the accuracy of forecasts from all respondents who said they were registered (in columns 1–4) to those from respondents who said they were registered, would definitely vote, and were very or somewhat interested in the election (in columns 5–8). This latter group represented 76 percent of all telephone survey respondents in 1996, a much greater percentage than the 56 percent of Ohio adult residents who voted that year. We also examined forecasts from a more restricted set of respondents: those who said they were registered, would definitely vote, and were very interested in the election (in columns 9–12). This group represented merely 35 percent of all telephone survey respondents in 1996, a much smaller proportion than the 56 percent of adults who voted that year. Thus, we tightened up the interest filter well beyond what the University of Akron has typically done and well beyond what would be representative of the state.

If we focus on the third column in each set of four (which our earlier analyses suggested optimized accuracy by allocating all undecided respondents equally among the candidates), we see a clear improvement in forecast accuracy by using a tighter "likely voter" filter. The average error for all respondents who said they

TABLE 10.13 IMPACT OF VOTER REGISTRATION SCREEN ON UNIVERSITY OF AKRON FORECAST ACCURACY (IN PERCENTAGE POINTS)

CANDIDATES	ALL RESPONDENTS	REGISTERED TO VOTE	ACTUAL RESULTS
Clinton	49.7	49.0	47.4
Dole	34.9	37.2	41.0
Perot	15.3	13.8	10.7
Average	4.6	2.8	

Note: Undecided respondents have been allocated equally to the candidates.

TABLE 10.14 IMPACT OF "LIKELY VOTER" SCREENS AND ALLOCATION STRATEGIES ON FORECAST ACCURACY IN THE UNIVERSITY OF AKRON CANDIDATE RACES (IN PERCENTAGE POINTS)

	REGISTERED				REGISTERED, DEFINITELY WILL VOTE, VERY OR SOMEWHAT INTERESTED				REGISTERED, DEFINITELY WILL VOTE, VERY INTERESTED			
	UNDECIDED DROPPED		UNDECIDED ALLOCATED EQUALLY		UNDECIDED DROPPED		UNDECIDED ALLOCATED EQUALLY		UNDECIDED DROPPED		UNDECIDED ALLOCATED EQUALLY	
	Leaners Dropped	Leaners to Candidate	Leaners Allocated Equally	Leaners to Candidate	Leaners Dropped	Leaners to Candidate	Leaners Allocated Equally	Leaners to Candidate	Leaners Dropped	Leaners to Candidate	Leaners Allocated Equally	Leaners to Candidate
1988 STATE POLL												
President	4.3	3.5	3.1	2.9	3.3	2.8	2.3	2.7	3.8	3.7	3.2	3.4
U.S. Senate	5.1	4.5	3.6	3.7	2.7	1.8	1.6	1.3	1.5	0.5	0.4	0.1
1990 STATE POLL												
Governor	1.5	0.6	0.7	1.1	0.6	0.7	0.8	1.2	1.5	1.8	2.1	2.1
1992 STATE POLL												
President	2.9	1.1	1.0	1.5	5.2	2.3	1.0	1.9	2.5	1.6	1.3	1.3
1996 STATE POLL												
President	6.0	5.6	4.3	4.0	2.2	4.6	2.0	3.0	0.6	2.8	0.6	2.8
Average	4.0	3.1	2.5	2.7	2.8	2.4	1.5	2.0	2.0	2.1	1.5	1.9

were registered was 2.5 percentage points (column 3), compared to 1.5 percentage points for people who said they were registered, would definitely vote, and were very or somewhat interested (column 7), and 1.5 percentage points for people who said they were registered, would definitely vote, and were very interested (column 11). As we saw earlier, it is apparent throughout Table 10.14 that allocating all undecided respondents equally to the candidates improves accuracy over the other three methods for handling the undecideds.

Therefore, Table 10.15 uses this optimal allocation method and expands the pool of races to include ones where undecided respondents were not asked whether they leaned toward a candidate. Here, it is even clearer that the tightest filtering works best. When undecided respondents were allocated equally to the candidates (shown in columns 2, 4, and 6), average error was 3.1 percentage points for all registered voters; 2.8 percentage points for people who said they were registered, would definitely vote, and were very or somewhat interested in the election; and

TABLE 10.15 IMPACT OF "LIKELY VOTER" SCREENS AND ALLOCATION STRATEGIES ON FORECAST ACCURACY IN THE UNIVERSITY OF AKRON CANDIDATE RACES (IN PERCENTAGE POINTS)

	REGISTERED		REGISTERED, DEFINITELY WILL VOTE, VERY OR SOMEWHAT INTERESTED		REGISTERED, DEFINITELY WILL VOTE, VERY INTERESTED	
	Undecided Dropped	Undecided Allocated Equally	Undecided Dropped	Undecided Allocated Equally	Undecided Dropped	Undecided Allocated Equally
1988 STATE POLL						
President	4.3	3.1	3.3	2.3	3.8	3.2
U.S. Senate	5.1	3.6	2.7	1.6	1.5	0.4
1990 STATE POLL						
Governor	1.5	0.7	0.6	0.8	1.5	2.1
Attorney General	7.5	4.7	7.9	5.2	2.9	2.1
Auditor	15.5	8.5	14.5	8.6	11.0	7.1
Secretary of State	12.0	1.6	2.6	2.7	2.0	2.1
Treasurer	8.8	0.1	8.5	1.0	5.9	0.7
1992 STATE POLL						
President	2.9	1.0	5.2	1.0	2.5	1.3
1996 STATE POLL						
President	6.0	4.3	2.2	2.0	0.6	0.6
Average	5.9	3.1	5.3	2.8	3.5	2.2

TABLE 10.16 IMPACT OF "LIKELY VOTER" SCREENS AND ALLOCATION STRATEGIES ON FORECAST ACCURACY IN THE UNIVERSITY OF AKRON REFERENDA (IN PERCENTAGE POINTS)

	REGISTERED		REGISTERED, DEFINITELY WILL VOTE, VERY OR SOMEWHAT INTERESTED		REGISTERED, DEFINITELY WILL VOTE, VERY INTERESTED	
	Undecided Dropped	Undecided Allocated Equally	Undecided Dropped	Undecided Allocated Equally	Undecided Dropped	Undecided Allocated Equally
1990 STATE POLL						
Gambling amendment	0.9	0.3	1.6	0.5	1.6	0.7
1996 STATE POLL						
Gambling amendment	12.6	12.4	11.5	11.5	4.5	4.5
Average	6.8	6.4	6.6	6.0	3.1	2.6

2.2 percentage points for people who said they were registered, would definitely vote, and were very interested. Allocating the undecided voters equally to the candidates again clearly improved prediction accuracy over dropping them.

The same improvement in accuracy with tighter filtering is apparent in Table 10.16, which displays results for the referenda. When undecided respondents were allocated equally between yes and no votes, average error was 6.4 percentage points for people who said they were registered; 6.0 percentage points for people who said they were registered, would definitely vote, and were very or somewhat interested in the election; and 2.6 percentage points for people who said they were registered, would definitely vote, and were very interested in the election. Therefore, it seems that the University of Akron forecasts would have been more accurate if a tighter "likely voter" screen had been used.

Trends over Time

Forecast error declines as the months of an election campaign pass. Table 10.17 focuses on races that were forecast in all waves and shows that average error was largest using the first survey's data and progressively smaller in later months. In the *Columbus Dispatch* surveys, average error in early September was 3.8 percentage points, compared to 2.3 percentage points in early October and 1.5 percentage points in late October. In the telephone surveys, average error was 4.7 percentage points in early September, 3.2 percentage points in early October, and 2.6 percentage points in late October.[1] These results replicate the findings of

TABLE 10.17 *COLUMBUS DISPATCH* AND UNIVERSITY OF AKRON FORECAST ERROR (IN PERCENTAGE POINTS)

	COLUMBUS DISPATCH			UNIVERSITY OF AKRON		
	Early September	Early October	Late October	Early September	Early October	Late October
1988						
President	3.0	3.5	0.6	3.6	0.4	3.1
U.S. Senate	2.0	1.5	0.5	3.4	4.7	3.6
1990						
Governor	3.3	0.9	0.2			
Attorney General	8.3	4.8	4.5			
Secretary of State	1.0	0.7	0.9			
1991						
Mayor	4.5	3.9	4.7			
1992						
President				7.2	4.5	1.0
1994						
Governor	8.0	1.8	0.6			
Attorney General	1.8	0.5	2.6			
Auditor	2.5	0.9	0.5			
Secretary of State	3.6	2.7	0.9			
U.S. Senate	4.2	4.5	0.6			
Average	3.8	2.3	1.5	4.7	3.2	2.6

Campbell and Wink (1990) and Crespi (1988), who also showed that forecasts become more accurate as the date of the survey approaches Election Day.

This trend in accuracy could have occurred simply because people's senses of their candidate preferences were fuzzy early in a campaign. Those preferences may not have truly changed in a systematic way over a campaign, but people may have gotten better at reporting those preferences precisely. If so, the greater error in early waves would presumably be random. At the same time, the trend in accuracy may have occurred because people's candidate preferences genuinely changed during the course of the campaign. That is, campaign events may have led people to shift their loyalties in one direction or the other. It is therefore of interest to see whether the two survey methods tracked comparable trends in preferences over time. If the shifts paralleled each other, that would suggest that real change occurred. But if the shifts were largely independent of each other, that would suggest random error, perhaps attributable to vague internal cues.

To explore this, we assessed the measured changes in candidate preferences between early September and late October, and they appear in Table 10.18. These percentages are changes in the predicted proportion of votes for the Republican candidate.[2] A positive percentage means the candidate gained predicted votes between early September and late October. A negative number means the candidate lost predicted votes between early September and late October.

Although some of these trends are quite small, others are relatively large and could make the difference in whether a candidate is predicted to win or lose. Interestingly, the telephone surveys tracked larger changes (average absolute value = 4.3 percentage points) than the mail surveys (average absolute value = 2.9 percentage points). This may reflect greater measurement error in the telephone surveys, but it may also reflect a greater ability of the telephone surveys to detect real trends.

The correlation between columns 1 and 2 of Table 10.18 is only .33, suggesting that the changes documented by the two methods are quite different from one another. That is, the two columns of numbers share only about 10 percent of their variance. For the majority of the races (nine of fourteen), the two surveys tracked trends in the same direction, whereas for the other five races, the two surveys tracked trends in opposite directions. But in some cases where the two surveys agreed on the direction of a trend (governor in 1990 and attorney general in 1994), the Akron surveys showed a relatively large trend whereas the mail surveys showed essentially no trend at all. So it is difficult to view the "correspondence"

TABLE 10.18 TRENDS IN SUPPORT FOR THE REPUBLICAN CANDIDATE FROM EARLY SEPTEMBER TO LATE OCTOBER (IN PERCENTAGE POINTS)

		TREND	
YEAR	RACE	University of Akron	*Columbus Dispatch*
1988	President	6.6	2.4
	U.S. Senate	0.2	1.5
1990	Governor	–3.1	–0.7
	Attorney General	1.0	–0.3
	Auditor	1.2	1.6
	Treasurer	–5.4	–6.4
	Secretary of State	1.2	1.6
1992	U.S. Senate	13.3	1.4
1994	Governor	12.7	6.3
	Attorney General	–2.2	–0.8
	Auditor	–1.0	3.0
	Treasurer	–3.5	5.7
	Secretary of State	6.0	–2.7
	U.S. Senate	2.5	–5.8
Average Absolute Value		4.3	2.9

in these cases as especially meaningful. We are therefore inclined to view these results as suggesting surprisingly little agreement between the two survey methods in terms of trends.

Although this could mean that one of the survey methods is more accurate at tracking trends than the other, there is no way to test this possibility because there is no "gold standard" with which to know what real trends in candidate popularity occurred for these races. But because the early September mail surveys forecast the final election outcomes more accurately than the telephone surveys, the mail surveys evidence indicating relatively little change in candidate popularity over time may be most on target, and the changes documented by the telephone surveys may be mostly the result of random error.

Response Order Effects

The University of Akron telephone surveys have routinely rotated the order of candidate names across respondents. Krosnick's (1991) theory of satisficing anticipates an effect of name order on responses. Specifically, when response alternatives are read aloud to respondents, as they were in these surveys, recency effects are expected, advantaging names presented last in a list. To test this prediction, we analyzed data from telephone surveys asking about the two-candidate races listed in Table 10.19. The proportions of people saying they were undecided were not significantly different in the groups of people who received different name orders, so these respondents were dropped from these analyses. As the third column of numbers in Table 10.19 shows, each candidate received more support when listed last than when listed first, by an average of 3.1 percentage points. A meta-analysis of these tests revealed that this recency effect was statistically reliable (d = .08, z = 3.06, p < .003).

This finding has at least two practical implications. Ohio rotates candidate name order in elections, thereby eliminating any impact of name order on election outcomes. Therefore, it has been wise for the University of Akron to rotate name order to eliminate any order-induced bias in forecasts. But many states do not rotate candidate name order (e.g., Nevada, Illinois, Georgia, Massachusetts, and Colorado). Instead, for example, an incumbent running for reelection is always listed first in Massachusetts. In New Hampshire, the candidate of the party that won the last election for an office is listed first; and in Georgia, Connecticut, and Maryland, the first candidate listed on the ballot for each office is that of the party that won the most recent election for governor of the state. In states such as these, one might imagine, forecasting surveys should present the candidate names in the same order as they will appear on the ballot, so as to have the same order effect present in the survey as will appear in the election.

TABLE 10.19 IMPACT OF CANDIDATE NAME ORDER ON UNIVERSITY OF AKRON ELECTION FORECASTS

		CANDIDATE NAME ORDER				
YEAR	RACE	Democratic Candidate First	Republican Candidate First	Difference	X^2	p
1986	Governor					
	Democratic Candidate	61.3%	63.7%			
	Republican Candidate	38.7%	36.3%	2.4	0.22	0.64
	N	199	168			
	U.S. Senate					
	Democratic Candidate	70.9%	73.7%			
	Republican Candidate	29.1%	26.3%	2.8	0.35	0.55
	N	206	167			
1988	President					
	Democratic Candidate	38.0%	42.8%			
	Republican Candidate	62.0%	57.2%	4.8	2.36	0.12
	N	465	511			
	U.S. Senate					
	Democratic Candidate	60.5%	64.0%			
	Republican Candidate	39.5%	36.0%	3.5	1.29	0.26
	N	461	516			
1990	Governor					
	Democratic Candidate	42.5%	43.1%			
	Republican Candidate	57.5%	56.9%	0.6	0.02	0.89
	N	275	276			
1992	U.S. Senate					
	Democratic Candidate	51.4%	56.0%			
	Republican Candidate	48.6%	44.0%	4.6	7.77	0.01
	N	439	420			
Average				3.1		

However, recent research by Miller and Krosnick (1998) suggests that this may be unwise if the preelection surveys are done by telephone. Krosnick's (1991) theory of response order effects anticipates primacy effects (advantaging candidates presented first) when a list of alternatives is presented visually, as are candidate names on election ballots and voting machines. Indeed, Miller and Krosnick (1998) found that in the 1992 Ohio elections, reliable name order effects appeared in 48 percent of 118 races. These significant name order effects nearly always advantaged the candidate listed first on the ballot, who gained an average of 2.5 percent more votes. Such primacy effects were largest in races where candidate party affiliation was not listed on the ballot, in races that had been minimally publicized, and when no incumbent was running for reelection.

Therefore, in states that do not rotate candidate name order across precincts on Election Day, telephone surveys may find it more difficult than mail surveys to forecast outcomes. Because the mail surveys present the names visually, the primacy effects that appear on Election Day should appear in the preelection surveys as well. But telephone surveys can only (1) generate results containing recency effects, or (2) counterbalance name order across respondents to generate results unbiased by order. Both of these would be less than optimal. Therefore, the best approach might be to rotate name order in the telephone surveys, estimate the size of the recency effect for each race, and use those effect sizes to introduce primacy advantages of the same magnitudes into the forecasts.[3] We look forward to future research exploring the effectiveness of this approach.

One final step we took in this investigation was to explore whether the *Dispatch* surveys overstated support for candidates listed first, which would follow from Krosnick's (1991) response order theory, given that names were presented visually. Candidate names in these surveys were presented to respondents only in alphabetical order. Therefore, their forecasts might have systematically overpredicted the vote share of the alphabetically first candidate, because primacy effects present in the survey are canceled out of the real election outcome by the name rotation across precincts.

To test this, we assessed the directions of the forecast errors made by the final mail surveys. As the first column of Table 10.20 shows, the first-listed candidate was sometimes predicted to receive more votes than he or she actually received (indicated by positive numbers) and was sometimes predicted to receive fewer votes than he or she actually received (indicated by negative numbers). The primacy effect prediction anticipates mostly positive numbers in this column, but this is not the case. Only 38 percent of the errors listed are positive, compared to 59 percent that are negative. Further, although the primacy hypothesis anticipates mostly negative errors for the candidate listed last, only 46 percent of the numbers in the second column of Table 10.20 are negative, and 49 percent are positive. Thus, there is clearly no indication that errors were in line with primacy effects, and indeed, there is a trend toward recency effects.

According to the evidence reported by Miller and Krosnick (1998), response order effects would not be expected in most of these races because partisan affiliations of the candidates were listed on the questionnaires. Specifically, only four of the races in Table 10.20 were nonpartisan: the race for chief justice and the two races for justice in 1986, and the city attorney race in 1993. Averaging across these four races, there was no greater error for candidates listed first alphabetically (mean forecast error for the candidate listed first = –.10). This is certainly a surprise and may suggest that no steps need to be taken to correct for primacy bias in the mail survey forecasts.

TABLE 10.20 DIRECTION OF *COLUMBUS DISPATCH* FORECAST ERROR FOR CANDIDATES LISTED FIRST AND LAST (IN PERCENTAGE POINTS)

YEAR	RACE	FORECAST ERROR FOR CANDIDATE LISTED FIRST	FORECAST ERROR FOR CANDIDATE LISTED LAST
1980	President	2.1	–5.5
1982	Governor	–1.0	0.0
	Attorney General	1.7	–0.5
	Auditor	–0.8	0.8
	Treasurer	–0.9	–0.1
	U.S. Senate	–0.1	–1.1
	Secretary of State	–0.6	–0.4
1984	President	–0.1	1.1
1986	Governor	–0.5	0.5
	Attorney General	1.2	–1.2
	Auditor	1.3	–1.3
	Secretary of State	0.5	–0.5
	Treasurer	–1.4	1.4
	U.S. Senate	3.2	–3.2
	Chief Justice	0.4	–0.4
	Justice	–0.6	0.6
	Justice	4.5	–4.5
1988	President	–0.1	1.0
	U.S. Senate	–0.5	0.5
1990	Governor	–0.2	0.2
	Attorney General	–4.5	4.5
	Auditor	3.7	–3.7
	Treasurer	–1.6	1.6
	Secretary of State	–1.6	1.6
1991	Mayor	–4.7	4.7
	City Council	–1.0	0.7
1992	President	1.4	–2.2
1993	City Clerk	–1.9	1.9
	City Council	–1.3	–0.2
	City Attorney	–4.7	4.7
1994	Governor	0.0	0.9
	Attorney General	2.6	–2.6
	Auditor	0.5	–0.5
	Treasurer	1.7	–2.2
	Secretary of State	–0.9	0.9
	U.S. Senate	0.3	0.6
1996	President	–0.1	0.0

Discussion

Our findings can be summarized as follows. First, mail surveys have been more accurate than telephone surveys in forecasting the outcomes of all sorts of contests, although local candidate races and referenda are more difficult than statewide candidate races for both methods to predict. Both methods manifested a bias toward overpredicting the winner's margin of victory, although the mail surveys manifested this less than the telephone surveys did.

One way to overcome this bias is to allocate undecided respondents randomly to the various candidates or referendum response options. Although taking seriously undecided respondents' reports of which candidate they lean toward does increase accuracy, it can be increased even more if these respondents are assumed to vote randomly. In fact, adding 2 percent random responses to each candidate's predicted take in the mail surveys also improved their accuracy by reducing the predicted margin of victory of the winner.

Thus, in contrast to the implication of Converse's (1964) observation that random responses plague surveys and introduce error, we found here that forecast accuracy could actually be increased by adding *more* random responses. Further, our experimental evidence that the handling of "undecided" responses influences accuracy challenges correlational evidence suggesting that "undecided" responses do not influence accuracy (Lau 1994). Our evidence might also appear at first to contradict Daves and Warden's (1995) conclusion that equal accuracy is obtained regardless of whether one drops undecideds or apportions them equally. In fact, a close look at their results reveals a pattern in line with ours: Dropping the undecideds yielded an average error of 2.33 percentage points across seven polls in their Table 7.1, compared to average error of 2.01 percentage points when undecideds were allocated equally. Our finding in this regard is in line with similar evidence reported by Erikson and Sigelman (1995) showing that allocating undecided voters equally improves the accuracy of forecasts of congressional elections.

It is interesting to note that results from 1997 preelection surveys in Amsterdam, Holland, further reinforce our conclusions (Neijens et al. 1997). In March 1997 the city of Amsterdam held a referendum regarding the building of a new housing project. Polling was done each day from 25 February through 18 March, the day before the election. The first day's data included 45 percent of respondents against the referendum, 30 percent in favor, and 25 percent saying they were undecided. If the latter respondents are simply dropped, the predicted final vote is 60 percent against and 40 percent in favor. If instead the "undecided" respondents are divided equally, the predicted final vote is 58 percent against and 42 percent in favor, precisely the actual election outcome.

As we have suggested, there are at least two possible reasons that allocating undecided voters equally to the candidates improved forecast accuracy. It may be that a small proportion of voters enter the voting booth without a clear preference for one candidate over another in some races. Instead of refraining from casting a vote in these races, they may randomly select a candidate from the list presented to them. Thus, allocating undecided respondents equally to the candidates may capture the randomness with which some voters make their vote choices on Election Day, thereby improving forecast accuracy.

Alternatively, the observed improvement in forecast accuracy may be attributable simply to the fact that dividing undecided voters equally among the candidates was the allocation strategy that gave the greatest advantage to the underdog. That is, preelection surveys may systematically underestimate the support for the candidate perceived to be behind (because of the spiral of silence), and any allocation strategy that gives more of the undecided voters to the trailing candidate may improve accuracy.

Our results in this regard challenge Crespi's (1988) recommendation to drop undecided respondents when calculating election forecast on the assumption that people without a candidate preference will not participate in the election. Our findings suggest, instead, that dropping undecided respondents is the least effective of all the allocation strategies we examined.

It is important to note that our tests of the various allocation and likely voter selection strategies were conducted using unweighted data from the *Dispatch* and Akron surveys. The decision to do so was based on a few considerations. First, preelection survey nonresponse appears to be highest among people least likely to vote (Greenwald et al. 1988). Underrepresentation of members of some social or demographic groups in survey samples is therefore likely to help, rather than hurt, forecast accuracy. Weighting data would therefore counteract this tendency for nonvoters to also be nonresponders. Furthermore, weighting more heavily the responses from members of underrepresented groups is based on the tenuous assumption that members of the group who did not participate in the survey will vote identically to members of the group who did participate in the survey. For both of these reasons, weighting preelection survey data to reflect the demographic composition of the population seems likely to detract from, rather than enhance, the degree to which the survey sample accurately represents the subset of the population that will participate in the election.

Many survey organizations do routinely weight their data to match the demographic characteristics of the entire population, and it is not clear that the optimal allocation and selection strategies that we identified here would be as effective at improving forecast accuracy with weighted data. It seems quite likely, in fact, that the optimal allocation and selection strategies may be quite different for weighted and unweighted data. We therefore caution investigators not to presume that the techniques we developed will work equally well with weighted data, and we encourage experimentation with unweighted data.

Our results have implications for how to make a decision that questionnaire designers routinely confront: whether to offer respondents an explicit "don't know" response option in attitude and belief questions. Many studies document that offering a "don't know" option reduces the likelihood that respondents will offer substantive answers to such questions (e.g., Bishop, Tuchfarber, and Oldendick 1986; Schuman and Presser 1981). Some observers have presumed this to be evidence of improved data quality because people without real opinions are thought to be admitting this, rather than concocting meaningless answers because they feel pressure to appear opinionated.

In contrast, Krosnick's (1991) theory of satisficing suggests instead that these "don't know" filters may do more damage than good by discouraging people with real opinions from reporting them. According to this perspective, some respondents employ low-effort response strategies to avoid the cognitive effort required to provide optimal responses. One such low-effort strategy is choosing the "don't know" response option when it is offered. If this option is not offered, respondents may instead choose to expend the effort needed to generate a valid substantive response. Consequently, data quality would not be compromised by omitting a "don't know" response option. Indeed, a number of previous investigations have supported this notion (e.g., Gilljam and Granberg 1993; Krosnick and Berent 1993; McClendon and Alwin 1993).

Our results can be viewed as consistent with this latter line of thinking. We found that a substantial proportion of respondents claimed to be undecided in the University of Akron surveys, which explicitly offered this as a response option. However, when probed, many of these respondents said that they did lean toward one candidate. When these responses were treated as valid preferences, forecast accuracy was improved. Thus, encouraging respondents to generate substantive responses by eliminating the "undecided" response option led to the collection of data that improved survey quality, a conclusion in line with Crespi's (1988) on this point.

We also found that filtering out respondents who are unlikely to vote (based upon their self-reports of registration status, likelihood of voting, and interest in the campaign) improves forecast accuracy. Surprisingly, though, very strong filtering worked better than weaker filtering. In fact, our evidence indicated that the telephone survey forecasts were most accurate when they were based on data from about only one third of the respondents. Across the nine statewide candidate races from 1988 through 1996, the tight "likely voter" selection screen and the allocation strategies that we devised generated election forecasts with an average error of 2.2 percentage points (see Table 10.15).

This finding is really quite remarkable. On the basis of sampling error alone, we would have expected an average error of 5.04 percentage points for estimates based on samples of these sizes. Therefore, the selection and allocation strategies outlined here consistently produced telephone survey forecasts that contained less than half the inaccuracy that would be expected simply on the basis of sampling error. Of course, other sources of error (e.g., mistakes by respondents and interviewers) are likely to have created imprecision as well, making this small error appear to be even more remarkable.

These results suggest that part of the reason the University of Akron's pre-

election surveys were less accurate than the *Dispatch* mail surveys has been that the telephone surveys employed relatively weak "likely voter" screens. Our evidence that these forecasts were most accurate when only about one third of respondents contributed data is consistent with the fact that *Dispatch* forecasts have also been based upon very small proportions of potential respondents.

In his examination of the features associated with election forecast accuracy, Crespi (1988) found that survey organizations that relied on a single question about voting intention to identify likely voters generated less accurate election forecasts than did organizations that used several questions. This is consistent with our finding with the University of Akron data that more stringent "likely voter" screens based on more questions yielded more accurate forecasts than did less stringent screens based on fewer questions.

However, our findings in this regard contradict another of Crespi's (1988) recommendations. He suggested that the identification of likely voters should be based in part on estimated turnout for a given election, which provides researchers with a rough guide for approximating the proportion of a survey sample that should be considered likely voters. Our evidence suggests that the most accurate forecasts are generated from much smaller proportions of the sample than would be expected on the basis of turnout rates. Projected turnout, therefore, appears not to be particularly useful for calibrating the stringency of a survey organization's "likely voter" screen.

The two survey methods paint only weakly correlated portraits of change in candidate popularity over the course of campaigns, and the telephone surveys registered larger changes than did the mail surveys. Based upon the limited information available on this matter, we are inclined to believe that the mail surveys are probably more accurate in tracking trends and that real trends are probably quite small in most elections. This conclusion is consistent with Gelman and King's (1993) argument that much of the variation over time in telephone survey forecasts is unrelated to actual election outcomes.

Finally, we found that telephone surveys are likely to be biased by recency effects, which can be eliminated by rotating candidate name order. However, primacy effects are likely to appear in some races in states that do not rotate candidate name order on Election Day, which poses a challenge for telephone surveys. We look forward to the testing and development of techniques to manage this problem.

We look forward, as well, to replications of the findings reported here with data from other states and from the nation as a whole. Most of the conclusions reached here were supported by data from several election years and in many cases across the two survey modes, strengthening our confidence in the generalizability of the observed results. But stronger tests of generalizability require data collected from other populations by other survey organizations.

Conclusions

Screening and allocation are currently done differently by different survey organizations, and relatively little collective wisdom in published research has established an empirically validated set of guidelines that have been shown to be optimal. We hope that the approach taken in this chapter will be paralleled by the future work of more researchers systematically exploring the impact of specific survey procedures, subjecting common intuitions and conventional wisdom to rigorous empirical scrutiny and contributing to the establishment of a proven set of professional standards for survey research.

Such systematic explorations may well affirm many of the intuitions that underlie current survey practices. However, empirical scrutiny may also call into question elements of conventional wisdom. The tests we have reported in this chapter shed new light on procedures for optimizing the accuracy of preelection surveys, and future research taking a similar approach will undoubtedly do the same. We look forward to such research, and to the systematic development of empirically validated guidelines for conducting quality survey research.

Notes

The authors thank Quin Monson for his help. This chapter was prepared partly while the second author was a Fellow at the Center for Advanced Study in the Behavioral Sciences. Support was provided by grants SBR-9503822 and SBR-9022192 from the National Science Foundation. The authors wish to thank Michael Tichy and Robert Hudson for their help in preparing the manuscript. Correspondence regarding this chapter should be addressed to Penny S. Visser or Jon A. Krosnick, Department of Psychology, Ohio State University, 1885 Neil Avenue, Columbus, Ohio 43210 (e-mail: visser.102@osu.edu or krosnick@osu.edu).

1. These percentages are based upon the most accurate allocation and screening methods (*Dispatch*: allocating undecideds equally; Akron: allocating undecideds and race-skippers equally and including only people who were registered, said they would definitely vote, and said they were very interested in the election).
2. These percentages are again based upon the most accurate allocation and screening methods (*Dispatch*: allocating undecideds equally; Akron: allocating undecideds and race-skippers equally and including only people who were registered, said they would definitely vote, and said they were very interested in the election).
3. Although it is not necessarily the case that primacy and recency effects will be of the same magnitude, this approach provides at least a starting point for addressing the problem of opposing order effects for visual versus oral presentation of response options.

PART IV

Controversial Issues in Election Polling