Public Opinion Quarterly, Vol. 70, No. 2, Summer 2006, pp. 197–223

CAN REGISTRATION-BASED SAMPLING IMPROVE THE ACCURACY OF MIDTERM ELECTION FORECASTS?

DONALD P. GREEN
ALAN S. GERBER

Abstract We compare the predictive accuracy of preelection polls using two types of sampling frames, random digit dialing (RDD) and registration-based sampling (RBS). The latter involves stratified random sampling from voter registration lists. In order to assess the accuracy with which RDD and RBS predict election outcomes, we collaborated with the *Washington Post*, Quinnipiac, and CBS News polls, which conducted parallel RDD and RBS surveys in Maryland, New York, Pennsylvania, and South Dakota prior to the November 5, 2002, elections. The results suggest that in the gubernatorial and congressional elections studied, RBS performed as well, if not better, than RDD, both in terms of forecasting accuracy and cost.

Each election year features an implicit competition between polls using two different approaches to sampling. Media, commercial, and academic pollsters rely almost exclusively on random digit dialing (RDD). This sampling method directs calls to randomly generated telephone numbers within certain area codes and exchanges (Groves 1990). Pollsters for political campaigns, on the other hand, tend to direct their calls to random samples of people whose names appear on registration lists. They regard registration-based sampling

DONALD P. GREEN and ALAN S. GERBER are professors of political science at Yale University. Earlier versions of this report were presented at the 2002 annual meeting of the American Association for Public Opinion Research, Nashville, TN, and the Gallup Conference on Improving the Accuracy of Polling, May 2–4, 2002, Washington, DC. The authors are deeply indebted to Christopher Mann, who played a key role in preparing and analyzing the RBS samples used in this study, and Costas Panagopoulos, who helped assemble the results. We are grateful to Rich Morin and Claudia Deane at the *Washington Post*, Kathleen Frankovic and Anthony Salvanto at CBS, and Douglas Schwartz at Quinnipiac University, who conducted the polls reported in this article. We profited from the suggestions of many readers, among them Dale Kulp, Jim Lepkowski, Mark Mellman, and Nick Panagakis. This research was supported by a grant from the Smith Richardson Foundation, which bears no responsibility for the views we express. Address correspondence to Donald P. Green; e-mail: donald.green@yale.edu.

doi:10.1093/poq/nfj022

(RBS) as more economical, particularly in midterm election years, when just two out of five adults cast ballots.

Neither sampling method is foolproof. RDD surveys rely on respondents to provide accurate information about whether they are eligible to vote and likely to do so. RBS polls suffer from incomplete coverage of the population of registered voters because some registered voters have unknown phone numbers, and in some cases lists of registered voters are inaccurate or incomplete. Both sampling methods are susceptible to nonresponse bias. That said, the question remains how these two less-than-perfect sampling approaches compare in practice and whether considerations of cost efficiency weigh in favor of either method.

To assess the accuracy of polls using the two sampling approaches, we present results from head-to-head comparisons of RDD and RBS in four states.[1] The *Washington Post*, CBS News, and the Quinnipiac Poll conducted parallel RDD and RBS surveys during the weeks leading up to the November 2002 elections. Following the approach of Crespi (1988) and others (Lau 1994; Martin, Traugott, and Kennedy 2005; Mitofsky 1998), we use the actual election outcomes from six gubernatorial and senatorial contests as the benchmark for assessing the accuracy of the parallel RDD and RBS polls. We find that the RBS surveys had, on average, slightly smaller forecasting errors than RDD surveys, in part because in the context of these low-turnout elections, the RBS surveys were able to generate large samples of registered voters at a lower cost per interview. These findings suggest that in midterm elections RBS or RBS/RDD dual-frame sampling approaches may be cost-effective alternatives to RDD polls.

This article is structured as follows. We begin by summarizing the basic features of the RBS sampling approach and describing some of the practical issues that arise in the course of conducting an RBS survey. A formal depiction of the biases that may afflict each approach, as well as the potential tradeoff in terms of sampling variance, is presented in an appendix, which may be accessed in the online version of the journal. In brief, this formal analysis explains why, given the many potential sources of bias, the comparison between RDD and RBS polling must be made empirically. The empirical section of the article begins with a description of the four surveys and then assesses the accuracy with which they forecasted gubernatorial and senatorial election outcomes. We find that RBS compares favorably both in terms of forecasting accuracy and cost. Although our limited test is far from definitive, it suggests the need for future investigation of RBS and hybrid sampling frames that use both RBS and RDD.

1. Unfortunately, we never learned the results from polls we conducted in three states in collaboration with Voter News Service, which went out of business immediately after the polls were conducted.

Registration-Based Sampling

One of the main advantages of sampling from registration lists is that the lists contain information that may be used to forecast voter turnout. Although the quality of information available from public records varies across jurisdictions, the typical registration file contains date of birth, date of registration, party registration, and the number of voters registered at each address. In addition, one may obtain data on past voter turnout. Sometimes these voter histories are furnished by registrars and sometimes by private vendors. The quality of registration records varies, but the point remains that one can produce a powerful statistical prediction of future voting behavior simply by reference to public records—before one speaks with a respondent.

The strong statistical relationship between public information and voter turnout is illustrated by voting patterns in Michigan. Suppose we were interested in forecasting voter turnout in the 2002 general election. The statewide voter history files contain two very useful pieces of information: when the voter registered and whether he or she voted in 1998 and 2000. The data indicate that voters' prior vote history strongly predicts their subsequent voter turnout. For example, turnout in 2002 was 82 percent among those who voted in the prior two elections but only 7 percent among those who were registered for both elections but voted in neither. Those who voted in 2000 but not in 1998 fell in between, voting at rate of 61 percent in the 2002 election.

Intuition suggests that any survey of the voting electorate that is designed to forecast an election outcome should place special emphasis on the opinions of citizens with a high *ex ante* probability of voting. Consider the limiting case, for example, in which members of the population come in two types, those who are certain to vote and those who are certain to abstain. There would be no point in interviewing any members of the second group (except for purposes of gaining demographic information for weighting purposes). By this logic, constructing an optimal sample—that is, a sample that gives the smallest prediction errors when forecasting the actual election outcome—should take notice of prior voting history because it predicts whether someone will vote in the upcoming election.

At this point, it may be helpful to distinguish between two different variants of RBS. Simple random sampling from registration lists, a technique that is common among political pollsters and occasionally by academic researchers (e.g., Chang and Krosnick 2001; Visser et al. 2000), might be dubbed random RBS, or RRBS. This approach makes no use of background information, such as past voting history, that can be helpful in predicting whether a potential respondent will vote in the upcoming election. Ignoring background information makes statistical sense when the electoral environment changes rapidly from one election to the next, undermining one's ability to correctly forecast the relative voting rates of different segments of the electorate. However, when the electoral setting is similar over a succession of elections, this background information can be quite useful; therefore, when we refer to RBS, we have in mind a stratified sampling frame. Each person on the registration list is assigned a probability of voting in the next election, based on factors such as registration date and past voting history. These probabilities enable the researcher to create a profile of those who will actually cast ballots. A random sample is drawn from this expected electorate.

In practice, the procedure boils down to the following steps. First, one breaks the registration list into strata. For purposes of illustration, suppose we divide our registration list into three strata: those who were not registered prior to the last election, those who were registered but did not vote, and those who voted. Second, one estimates the rate at which each stratum is expected to vote (more about how to do this in a moment). Third, one multiplies the voting rates by the number of people in each stratum in order to find the size of the anticipated electorate and what fraction of the electorate each stratum constitutes. These fractions then become the sampling weights for the RBS survey. Imagine that the three strata consist of 100,000 new registrants, 400,000 past abstainers, and 500,000 past voters. Suppose that voting rates among the three strata are 50 percent, 20 percent, and 90 percent. The expected electorate therefore consists of 50,000 new registrants (8.6 percent), 80,000 past abstainers (13.8 percent), and 450,000 past voters (77.6 percent). When drawing a sample from the registration list to be interviewed, 8.6 percent of the sample should be new registrants, 13.8 percent past abstainers, and 77.6 percent past voters (for a discussion of stratified sampling see Thompson 1992, chap. 11).

Of course in practice, voting rates within each stratum are not known beforehand. It is therefore handy to know how voting history has predicted subsequent turnout in recent elections. Before attempting to construct strata proportions for the 2002 election, we examined the relationship between turnout in the 1998 elections and the two preceding elections. We found, for example, that Pennsylvanians who voted in the 1994 and 1996 elections comprised 60 percent of the 1998 electorate. In the RBS survey for 2002, therefore, 60 percent of the names selected for our sample consisted of voters who had voted in the two previous federal elections. The state's actual voter turnout figures for 2002 showed that 61 percent of those who cast ballots had voted in the two previous federal elections. We were less fortunate in the case of New York, where we anticipated that 48 percent of the electorate would consist of those who had voted in the two previous federal elections. Due to differences in the competitiveness of the 1998 and 2002 elections, this forecast proved inaccurate, and 62 percent of the 2002 electorate consisted of these regular voters. Clearly, mistaken strata weights are a potential source of error, and future exploration of RBS must consider more robust and sophisticated stratification and weighting schemes. The empirical question addressed below is whether, given the pros and cons of the limited stratification scheme used here, stratified RBS outperforms random RBS in terms of forecasting accuracy.

Like any sampling scheme, stratified RBS can be made more complex. One could, for example, weight observations according to background characteristics such as age, party registration, gender, or the political characteristics of the voting precincts. A more sophisticated weighting scheme would take account of the fact that the candidate preferences of voters with ardent partisan preferences are easier to predict than those with moderate views. Knowing that a voter is a regular participant in Republican primaries leaves relatively little uncertainty about how he or she will vote. For this reason, the pollster need not interview as many regular primary voters as voters in other equally sized centrist groups in the population. The statistical rule of thumb is this: allocate more interviews to large subgroups of voters, unless you know ahead of time how they are likely to vote.[2]

An interesting practical question is what to do with stated vote intention. The issue is whether screening RBS surveys for self-described "likely voters" improves the forecasting accuracy of RBS polls. As we note in the online appendix, there is a potential trade-off between bias and variance: admitting nonvoters into the sample may introduce bias, but excluding nonvoters will increase sampling variability. Thus, our empirical assessment below compares the forecasting accuracy of RBS results based on the inclusion or exclusion of unlikely voters.

Registration-Based Sampling in Practice

Compared to the extensive literature on random digit dialing, relatively little has been written on the practical details of RBS. In this section we briefly describe how one obtains a registration list and how one conducts a list-based survey.

OBTAINING REGISTRATION LISTS

Registration-based sampling starts with a registration list. These lists are, with a few exceptions, available from one or two sources. The first source is the registrar of a local jurisdiction, such as a city or county. In a few cases these lists are officially off-limits to anyone who is not connected with a political party. Parties seem to be quite porous, however, and these forbidden lists are sometimes commercially available from database vendors. These vendors charge more for these data than local registrars, but the advantage of purchasing

2. Consider, for example, states with party primaries. In these states a large proportion of those who cast ballots during the 2002 general election were known to be primary voters of one party or the other. This group requires a bit of attention due to its enormous size. On the other hand, a pollster will quickly learn after a few interviews that primary voters disproportionately favor their party's nominee in the general election. Given that so much is known about these respondents in advance of the interview, polling resources are better allocated to other segments in the electorate about which there is more uncertainty.

registration lists from vendors is that they have often taken the trouble to append to the data set information about the previous elections in which a person has voted. Some vendors have even begun to automate the process of generating stratified random samples of registration lists.

Unfortunately, the decentralized way in which registration lists are compiled and maintained means that the quality of voter registration lists varies both across states and within them. Although the federally mandated standardization of voter registration lists is expected to improve list quality and comparability, pollsters who wish to use RBS must investigate the feasibility of this approach for the regions they seek to study. Fortunately, pollsters are typically most interested in close races, which are precisely the races that most interest political parties and campaigns. Parties and campaigns are the list vendors' big customers, and vendors make special efforts to keep up-to-date registration lists for tight races. Nevertheless, vendors, like pollsters, are sometimes caught off guard by a race that suddenly heats up. As we note below, this type of surprise occurred in Maryland, when the governor's race suddenly became competitive. As a result, the registration list that we used did not have the vote history for the 2000 election. We were forced, therefore, to stratify based on 1994 and 1996 vote history rather than on the 1998 and 2000 vote history.[3]

Matching addresses and phone numbers is the main source of slippage in the RBS sampling process. The availability of listed numbers varies widely by region. Iowa is very good; New York City is very bad. In some cases the matching rate can be improved by sending registration lists out to multiple vendors, who draw their phone numbers from different sources. The augmented statewide registration lists for Maryland, New York, Pennsylvania, and South Dakota contained phone numbers for approximately two-thirds of each sample.[4] This figure, however, includes phone numbers that proved to be outdated once interviewing began. In South Dakota, for example, 10 percent of the numbers called were found to be incorrect. Failure to obtain numbers from a sizable portion of the electorate is an important concern, particularly if voters with unknown phone numbers have distinctive political views. The Maryland RBS poll shows indications of underrepresentation of African-Americans (Deane and Morin 2003), and the Pennsylvania and New York RBS polls show some underrepresentation of urban dwellers (Schwartz and Richards 2003). The empirical test presented below is designed to assess whether these demographic imbalances lead to biased election forecasts.

Nothing prevents the pollster from falling back on RDD in cases where the phone match is particularly poor, and indeed the pioneering work of Mitofsky

3. In contrast to many forecasting exercises, where one uses similar elections as predictors, here we are interested in using as much recent information as possible to stratify people into groups with different voting probabilities. That is why we sought to use 2000 and 1998 rather than 1998 alone.

4. The initial phone match rates were 65 percent for Maryland, 69 percent for New York, 66 percent for Pennsylvania, and 70 percent for South Dakota.

et al. (2005) provides an example of this kind of dual-frame design. Prior to the 2004 election, these authors augmented a random RBS survey (N = 622) of Oregon voters using interviews with an RDD sample (N = 280) whose phone numbers were not part of the RBS sampling frame. Their dual-frame poll generated forecasts with somewhat lower mean-squared error than the RBS poll alone.

TELEPHONE INTERVIEW PROTOCOL

Once the RBS sample of respondents has been selected, the mechanics of the survey are quite straightforward. Since the identity of the respondent is known to the interviewer, the caller may ask for the respondent by name. In contrast to an RDD poll, it is not necessary to ask to speak to the adult with the next birthday or to measure the number of phone lines in the household. In addition, an RBS survey can dispense with the measurement of background variables such as age, location, or registration status, as these are known in advance. In the South Dakota poll we nevertheless asked respondents' age in order to gauge the quality of the registration information. Ages were available for 90 percent of the voter file. Among the survey respondents who disclosed their age to the interviewer, 95 percent of the reported ages were within one year of their "official" age according to the voter file.

A useful illustration of how response rates may vary by sampling methodology comes from a *Washington Post* survey of the Maryland governor's race conducted approximately nine days before the 2002 general election. RDD and RBS polls were conducted by a single calling house over an identical time period. The surveys were the same, aside from the fact that the RBS poll asked for respondents by name and dispensed with the within-household respondent-selection procedures used in the RDD survey. The RDD poll attempted 14,051 distinct phone numbers and completed interviews with 1,738 persons. Among the 960 who claimed to be registered to vote, 725 claimed to be "certain" to vote in the upcoming election. The RBS poll attempted to reach 3,752 registered voters and completed interviews with 838, of whom 657 claimed to be certain to vote.[5]

Although knowing the names of potential respondents alleviates some of the clumsiness of RDD enumeration procedures, it may have some drawbacks as well. It is possible that respondents to RBS surveys need extra assurances about confidentiality. Mann (2003) examined RBS and RDD response patterns to see whether RBS respondents showed more reluctance to disclose

5. Using American Association for Public Opinion Research *Standard Definitions* version 4, formula 1 (AAPOR 2000), the response rates for the RBS surveys are 42 percent for Maryland, 26 percent for Pennsylvania, 24 percent for New York, and 29 percent for South Dakota. The corresponding rates for the RDD surveys are 27 percent for Maryland, 18 percent for Pennsylvania, 18 percent for New York, and 21 percent for South Dakota. These figures were calculated by the respective polling firms, except for the RBS polls in Pennsylvania and New York, which were calculated by the authors.

vote intentions. He found differences in response rates to be statistically insignificant, consistent with earlier findings (Traugott, Groves, and Lepkowski 1987). Respondents' unwillingness to disclose vote intentions does not appear to impair the accuracy of RBS polls.

In sum, RBS presents pollsters with a number of trade-offs. The potential benefits include (1) useful background information from registration lists, (2) simplification of the interview protocol, and (3) higher rates of completed interviews with likely voters, particularly in low-salience elections. Potential drawbacks include (1) incomplete coverage across and within states, (2) lack of phone number information, particularly for certain segments of the registered population, and (3) respondents' unwillingness to disclose their vote intentions when the survey is not anonymous.

The first two drawbacks appear to warrant special attention. The online appendix takes a closer look at how gaps in the RBS sampling frame may contribute to bias. The basic conclusion from a formal analysis is that both RBS and RDD are susceptible to several different sources of bias. One can devise scenarios according to which either emerges as the superior approach, depending on the severity of coverage gaps, nonresponse problems, and preference heterogeneity among different segments of the population. For this reason, the trade-offs between different sampling frames can only be assessed empirically.

Parallel RBS and RDD Polls in 2002

In order to compare the forecasting accuracy and cost of RBS and RDD surveys, we collaborated with three polling organizations—CBS News, the *Washington Post*, and Quinnipiac University—that conducted statewide polls in South Dakota, Maryland, Pennsylvania, and New York. These states all have traditional electoral systems with registration deadlines and in-person balloting only on Election Day. Each of the four states featured a gubernatorial race in 2002.

All of the polls were conducted within the last month of the campaign. Although the National Council on Public Polls (NCPP) excludes polls conducted more than 18 days prior to Election Day from its roster of surveys used for preelection forecasting, a close examination of forecasting accuracy reveals that polls conducted in the last two weeks are not markedly more accurate than polls conducted during the two weeks before that. For example, Bloom's (2003) analysis of 232 polls in 2002 shows that the forecasting errors in the last two weeks were only slightly smaller than the errors associated with polls conducted in early October.[6] This finding is corroborated by our own

6. Martin, Traugott, and Kennedy (2005, table 6) show in addition that the slight pro-Democratic bias of the 548 polls they examined from the 2002 election did not diminish significantly as Election Day approached. Crespi's multivariate analysis of polls conducted in the 1980s showed a somewhat stronger effect of proximity to Election Day (1988, p. 167); this result may reflect sampling variability or genuine differences in the period studied.

examination of 70 gubernatorial and senatorial polls conducted during 1998 and the 159 NCPP gubernatorial and senatorial polls in 2002 (O'Neill, Mitofsky, and Taylor 2002). Although accuracy increases as the election draws closer, a regression of candidate error on the closing date of the survey shows that the level of candidate error is reduced by approximately .04 percentage points per day. Thus, a poll completed 24 days prior to the election can be expected to have less than a 1 percentage point higher rate of candidate error than a poll conducted 4 days before Election Day.

CBS News conducted the South Dakota poll October 9–11. The RDD and RBS surveys were both conducted using a CATI system to automate calling and coding of responses. Up to nine attempts were made to contact respondents. Although this poll was conducted a few weeks in advance of the election, the timing was the same for both the RBS and RDD polls. Thus, both polls have the same expected forecasting accuracy, which makes the comparison unbiased and potentially informative. As it turns out, both polls performed well in terms of forecasting accuracy. Had we dropped the South Dakota comparison on the grounds that these polls occurred too early, the results presented below would favor RBS even more strongly.

From October 20 to 24, the *Washington Post* conducted simultaneous surveys using RDD and RBS samples in Maryland. The surveys were conducted by TNS Intersearch, the firm that normally conducts polling for the *Washington Post*. Both surveys were conducted simultaneously using a CATI system to automate calling and coding of responses. Up to eight attempts were made to contact respondents.

The Quinnipiac University Polling Institute (see Schwartz and Richards 2003) conducted simultaneous surveys using RDD and RBS samples in Pennsylvania October 21–27. From October 28 to November 3 the institute also conducted RDD and RBS surveys in New York State. The pairs of surveys in each state were conducted using the facilities and staff at Quinnipiac University. The RDD surveys were conducted using a computerized CATI system to automate the process of calling and completing the questionnaire for live interviewers. Due to capacity limits, the RBS surveys were conducted using paper questionnaires filled out by the interviewers, who also manually dialed the phone numbers. On weeknights, calls were attempted from 5:30 P.M. to 9:30 P.M. for the RDD survey and from 7:00 P.M. to 9:30 P.M. for the RBS survey. On weekends, calls were attempted during the same periods for the RDD and RBS surveys: 10:00 A.M. to 3:00 P.M. on Saturdays and 4:00 P.M. to 9:00 P.M. on Sundays. The administration of the paper RBS survey was kept as similar as possible to the computerized RDD survey.

The RDD and RBS surveys in each state used nearly identical questionnaires. In each state the surveys differed in their introduction: the RBS questionnaire requested the voter by name, while the RDD questionnaire asked for the individual over age 18 with the next birthday (a pseudo-random selection technique frequently used in RDD polls). In each survey the vote preference question asked whether voters were certain to support the candidate or might change their mind. If voters declared no preference in the initial question, they were asked if they leaned toward a candidate; in keeping with the practice of each polling organization, the results reported below include the preferences of leaners. In Maryland the questionnaires had another major difference: the RDD questionnaire followed the vote preference section with questions about the favorability ratings of the candidates and about a number of issues in the campaign before concluding with a battery of demographic questions common to both questionnaires; the RBS questionnaire omitted the favorability and issue questions. In New York and Pennsylvania both the RBS and RDD questionnaires followed the vote preference questions with questions on the favorability ratings of the candidates and concluded with a battery of demographic questions. Since we focus our attention on vote preference, the fact that the questionnaires diverge after the vote preference questions were asked is inconsequential.

The RDD samples in each state were provided by Survey Sampling Inc. (SSI), which generated numbers randomly for the area codes within the state. SSI attempted to purge random numbers that were not in service from the potential sample. The RBS samples in Maryland, New York, and Pennsylvania were drawn randomly from a list of registered voters in each state maintained by Voter Contact Services (VCS). During the late summer, VCS gathered current lists of registered voters from each county of the three states. VCS completed updating its list of registered voters in New York and Pennsylvania by August 21, 2002. Because a few Pennsylvania counties (4.2 percent of the registered voters) did not collect voter history in the late 1990s, we used a simplified stratification system for those counties, using only three strata rather than our usual five. Due to delays in redistricting and the late primary election, VCS was not able to gather the 2000 voter history prior to the election in Maryland, forcing us to use elections four and six years prior rather than two and four years prior, as we had done in New York and Pennsylvania. In Maryland the registration list was updated on September 5, 2002. The voter records in Maryland, Pennsylvania, and New York were matched to phone records that were current as of October 2002. South Dakota registration data were updated as of September 23 and were obtained directly from the state. The registration information was matched to a database of telephone numbers by InfoUSA.

Prior to drawing the RBS sample, the entire list of registered voters was stratified into mutually exclusive groups based on past voting history. After estimating what share of the anticipated electorate each stratum represented, we drew a stratified sample of 40,000 registered voters in each state. From this stratified random sample, a subsample of 10,000 voters with phone numbers was selected in each state. Half of the RBS samples in Maryland, New York, and Pennsylvania were sent mail in advance explaining the purpose of the survey and encouraging their participation. As Mann (2005) reports, the letter increased survey response rates somewhat but did not lead to a consistent improvement in forecasting accuracy. We therefore pool all of the RBS responses for a given state for purposes of the analysis presented below.

Comparing Forecasting Accuracy

The central question of this study is whether the use of RBS increases forecasting accuracy. It could be argued that the reliance on listed telephone numbers and within-state variation in list quality introduce bias. On the other hand, it could be that RBS outperforms RDD. RBS generates more usable interviews, and its weighted sampling frame takes advantage of hard data on respondents' prior voting behavior. The relative magnitude of these biases therefore remains an empirical question.

Table 1 presents the head-to-head comparison of RBS and RDD for four gubernatorial, one senatorial, and one statewide House election.[7] The leftmost columns of table 1 list the type of office, as well as candidates' names, party, and incumbency. These races varied in terms of competitiveness. The South Dakota senate race was very close, whereas the New York gubernatorial contest was a rout. As it turned out, these were the only two contests featuring an incumbent.

Four columns of table 1 present the RBS results under different weighting and sampling criteria. The first of these columns presents the results using all RBS respondents. Note that the term "stratified" in this context means that the sample was prestratified as described above, but no further weighting was done after the samples were created and given to the survey organizations. Recall that these stratified RBS samples were initially created using weights for voting propensities and were designed, therefore, to be representative of the voting population.

The second of the RBS columns weights the RBS data to take into account the fact that different sampling strata had different survey response rates. Those who completed interviews were weighted so that the initial strata proportions were preserved. This column helps assess whether preserving the initial strata proportions leads to any improvement in forecasting accuracy. The third column weights the RBS data so that the results reflect what would have been obtained from a random sample of the registration list, which is the RRBS sampling frame that is frequently used by political campaigns.[8] The

7. The New York poll solicited opinions about a lopsided attorney general race and a more competitive, but obscure, comptroller election. We have omitted these contests from our calculations here so that the forecasting results would be more directly comparable to the benchmarks that Bloom (2003) and O'Neill, Mitofsky, and Taylor (2002) provide for gubernatorial and senatorial races.

8. Weights were determined according to the sampling fractions used to draw the RBS sample from the original list of registered voters. To obtain an RRBS sample, for instance, we weighted each RBS stratum so that all observations had the same likelihood of being drawn from the original voter file.

Table 1. Comparison of Actual and Projected Election Results

					Registration-Based Sampling					Random Digit Dialing	
State/Office	Candidate	Party	Incumbent	Actual Vote[a]	Stratified But Not Reweighted	Weighted to Original Strata Proportions	Weighted to All Registered Voters	Using "Definitely Vote" Screen	Among Actual 2002 Voters	Registered Voters	Using Likely Voter Screen[b]
MD/Governor	Townsend	D		47.9%	44.6%	45.9%	33.0%	44.6%	44.7%	45.6%	49.0%
	Ehrlich	R		51.4%	48.0%	45.3%	42.6%	48.4%	48.1%	49.4%	49.4%
PA/Governor	Rendell	D		53.3%	48.6%	50.1%	55.3%	50.2%	49.8%	51.4%	54.0%
	Fisher	R		44.5%	40.3%	39.7%	36.7%	40.5%	40.5%	32.9%	35.0%
NY/Governor[c]	McCall	D		33.2%	27.6%	27.2%	26.2%	28.8%	27.3%	27.8%	29.0%
	Pataki	R	x	49.3%	42.4%	41.8%	40.9%	42.2%	43.0%	42.1%	45.0%
	Golisano	I		14.5%	18.5%	20.0%	21.7%	18.8%	17.9%	17.0%	14.0%
SD/Senate	Johnson	D	x	49.6%	39.5%	39.3%	42.0%	41.3%	39.7%	43.8%	44.8%
	Thune	R		49.5%	39.5%	40.1%	35.9%	40.0%	39.7%	38.6%	41.1%
SD/Governor	Abbott	D		41.9%	31.5%	31.1%	30.2%	31.3%	31.0%	37.4%	36.3%
	Rounds	R		56.8%	49.3%	50.4%	44.2%	51.1%	50.1%	44.3%	48.4%
SD/House	Herseth	D		45.6%	38.6%	38.5%	36.7%	40.0%	39.4%	43.4%	43.1%
	Janklow	R		53.5%	44.3%	43.4%	48.2%	43.2%	42.7%	41.8%	44.5%

Table 1. *(Continued)*

					Registration-Based Sampling					Random Digit Dialing	
State/Office	Candidate	Party	Incumbent	Actual Vote[a]	Stratified But Not Reweighted	Weighted to Original Strata Proportions	Weighted to All Registered Voters	Using "Definitely Vote" Screen	Among Actual 2002 Voters	Registered Voters	Using Likely Voter Screen[b]
Sample Sizes											
MD					738	738	738	657	669	960	725
PA					745	745	745	661	620	1,214	636
NY					735	735	735	656	575	1,018	624
SD					438	438	438	380	393	438	353

[a] The actual vote reflects the percentage of the vote received by the major candidates. It may not sum to 100 percent because minor party candidates were excluded. The results were downloaded from the state elections office Web site in each state.

[b] This column reflects postsurvey weighting of the data for vote likelihood and other factors as done by the survey organizations for public release. The likely voter screen varied across polling organizations. Likely voters in Maryland were screened based on the question, "I'd like you to rate the chances that you will vote in Maryland's election for governor on November 5: Are you absolutely certain to vote, will you probably vote, are the chances 50-50, or less than that?" The voter screen used in Pennsylvania and New York was based on whether the respondent had thought about the election, knew where to vote, expressed interest in the upcoming election, and responded "definitely" to the question, "How likely is it that you will vote in the election for governor on November 5th—would you say you will definitely vote, probably vote, possibly not vote, or definitely not vote?" The CBS News poll asked, "How likely is it that you will vote in the 2002 election for Congress next month—would you say you will definitely vote, probably vote, probably not vote, or definitely not vote, or have you already voted early by mail using an absentee ballot?"

[c] Vote totals for candidates appearing under more than one party in New York have been combined. The party listed in the second column is the major party affiliation.

fourth column restricts the RBS sample to likely voters.[9] The column helps us assess whether narrowing the sample to likely voters leads to improved accuracy due to a reduction in bias or diminished accuracy due to increased sampling variability.

Finally, we restricted the RBS sample to those respondents who actually voted, according to voter turnout records. Because the identification of actual voters can only be done after the fact, isolating this subgroup is of no practical value to pollsters who are seeking to anticipate election outcomes, but it serves an important methodological function. By comparing results for actual voters to other RBS weighting approaches, we can assess the value-added of an optimal turnout forecast. In addition, by comparing actual voters in the survey to actual voting outcomes, we can gauge each survey's sampling error without any turnout-related error.

For RDD surveys two sets of results are presented. The first column presents the survey results tabulated for all registered voters. The second column restricts attention to "likely voters" as defined by each polling organization. The results for likely voters are also weighted by various demographic factors such as age and gender. For Maryland, New York, and Pennsylvania these numbers are the final results reported for public release by each organization; the South Dakota results were calculated from the raw survey data in consultation with CBS.

The predictive accuracy of these polling results is gauged in table 2 using the calculation of "candidate error" used by Mitofsky (1998) and recommended by the National Council on Public Polls. The benchmark for this assessment is calculated by subtracting the Democratic candidate's vote percentage from the Republican candidate's vote percentage (making no adjustment for third-party candidates). The accuracy of the polling result is obtained by subtracting the Democratic candidate's reported vote share from the Republican candidate's (making no adjustment for undecided or other responses). Candidate error is defined as the absolute value of difference between these two quantities.

Table 2 presents the candidate error associated with each poll and weighting scheme. Looking first at the RDD results, we find, not surprisingly, that the results based on likely voters were on average more accurate than the results based on registered voters. The implication of this finding is that the loss of observations due to the likely voter screen is more than offset by the

9. In Maryland and South Dakota, likely voters were determined based on a single question; in Pennsylvania and New York, an index based on several questions was used instead. Because the procedure by which the Quinnipiac Poll screens for likely voters is proprietary and unknown, we created a simplified screen that was similar across all four RBS surveys. Each survey asked a single question (with slightly varying wording) about the respondent's certainty of voting. Our simplified screen selected only respondents who chose the highest degree of certainty on the scales in the respective polls (e.g., "absolutely certain to vote," "always vote," etc.). Experimenting with alternative screening procedures did not change the basic pattern of results. The RDD results report the likely voter screens used by the polling organizations themselves.

Table 2. Forecasting Error by Type of Sample, Using the National Council on Public Polls (NCPP) Formula for Assessing Candidate Error

	Registration-Based Sampling					Random Digit Dialing	
	Stratified But Not Reweighted	Weighted to Original Strata Proportions	Weighted to All Registered Voters	Using "Definitely Vote" Screen	Among Actual 2002 Voters	Registered Voters	Using Likely Voter Screen
MD/Governor	0.0%	2.0%	3.0%	0.1%	0.0%	0.1%	1.5%
PA/Governor	0.3%	0.7%	4.8%	0.4%	0.2%	4.8%	5.0%
NY/Governor[a]	0.6%	0.7%	0.7%	1.3%	0.2%	0.9%	0.0%
SD/Senate	0.1%	0.5%	3.0%	0.6%	0.1%	2.5%	1.8%
SD/Governor	1.5%	2.3%	0.4%	2.5%	2.1%	4.0%	1.4%
SD/House	1.1%	1.5%	1.8%	2.3%	2.3%	4.7%	3.2%
Average Candidate Error	0.6%	1.3%	2.3%	1.2%	0.8%	2.8%	2.1%
Average Candidate Error for all 2002 Governor/Senator Polls (N = 159)[b]							2.4%

[a] Following NCPP convention, the forecast error reported here is based on major party candidates. Applying the same calculation to the margin between Pataki and Golisano, an independent candidate, generates a candidate error of 5.6 percent for the stratified RBS poll, 5.0 percent for the registered voters RDD poll, and 2.1 percent for the likely voters RDD poll.

[b] O'Neill, Mitofsky, and Taylor 2002. Polls in the NCPP sample were all conducted on or after October 20.

gains in accuracy associated with excluding nonvoters with heterogeneous preferences.

Across all six races, the average candidate error among the likely voter RDD forecasts was 2.2 percent, which is slightly better than the 2.4 percent average reported by the National Council on Public Polls for the 159 gubernatorial and senatorial polls conducted during the last two and half weeks of the campaign (O'Neill, Mitofsky, and Taylor 2002). There appears to be nothing unusual about the performance of this set of RDD polls.

Turning to the RBS polls, we find the stratified polls appear to provide more accurate forecasts than the RDD likely voter poll in four of the six contests, and in the two remaining cases the level of predictive accuracy was about the same. The forecasting edge of RBS may stem in part from the fact that each of the RDD samples of likely voters has approximately one hundred fewer respondents. The contrasting sample sizes reflect the greater proportion of likely voters in RBS samples. The RDD samples of registered voters are larger than the corresponding RBS samples, but the likely voter screen causes this number to drop sharply.

Interestingly, the forecasting accuracy of RBS does not improve noticeably when we correct the RBS strata for response bias (column 2) or focus attention exclusively on respondents who claim to be likely voters (column 4). As Mann (2005) points out, patterns of survey nonresponse are in some sense their own likely voter screen; those who respond to surveys are disproportionately drawn from the ranks of regular voters. The strong performance of the stratified RBS polls is fortunate for survey researchers because it means that they need not discard "unlikely" voters from RBS samples that are already preweighted in favor of likely voters.

Nor does forecasting accuracy improve when we limit the sample to those RBS respondents who actually voted (column 5). This finding implies that RBS nonvoters were sufficiently similar to RBS voters that the loss of those observations reduces bias only minimally, while increasing sampling error.

Finally, our results offer a word of caution regarding the use of random RBS as opposed to stratified RBS. When we weight our RBS sample so that it represents a simple random sample from the population of registered voters (column 3), we find that forecasting accuracy diminishes.[10] The average forecasting error of random RBS samples turns out to be no better than that of RDD samples of registered voters.

Table 3 offers another analysis of forecasting error, this time using the log-odds ratio method recommended by Martin, Traugott, and Kennedy (2005). This approach produces results that are quite similar to what we report in table 2. The average absolute log-odds ratio is .035 for stratified RBS and .097 for RDD with likely voters. This table also sheds light on whether the difference

10. As Morin (2003) points out, tinkering with the unweighted Maryland sample to adjust for its underrepresentation of African-American respondents does not improve the poll's predictive accuracy.

Table 3. Forecasting Error by Type of Sample, Using Log-Odds Ratio Formula for Assessing Candidate Error (standard errors in parentheses)

	Registration-Based Sampling					Random Digit Dialing	
	Stratified But Not Reweighted	Weighted to Original Strata Proportions	Weighted to All Registered Voters	Using "Definitely Vote" Screen	Among Actual 2002 Voters	Registered Voters	Using Likely Voter Screen
MD/Governor	0.003	–0.084	0.185	0.011	0.003	0.010	–0.062
	(0.077)	(0.077)	(0.085)	(0.081)	(0.080)	(0.066)	(0.075)
PA/Governor	–0.005	–0.050	–0.228	–0.032	–0.024	–0.264	–0.251
	(0.078)	(0.078)	(0.078)	(0.082)	(0.085)	(0.064)	(0.086)
NY/Governor[a]	0.034	0.034	0.050	–0.013	0.059	0.020	0.044
	(0.090)	(0.091)	(0.092)	(0.094)	(0.102)	(0.077)	(0.095)
SD/Senate	0.003	0.025	–0.153	–0.029	0.003	–0.123	–0.083
	(0.108)	(0.107)	(0.109)	(0.114)	(0.113)	(0.105)	(0.115)
SD/Governor	0.145	0.181	0.077	0.187	0.176	–0.134	–0.015
	(0.109)	(0.109)	(0.113)	(0.116)	(0.115)	(0.106)	(0.117)
SD/House	–0.021	–0.038	0.113	–0.082	–0.078	–0.196	–0.127
	(0.105)	(0.106)	(0.105)	(0.113)	(0.111)	(0.104)	(0.114)
Average Absolute Log-Odds Ratio	0.035	0.069	0.134	0.059	0.057	0.124	0.097

[a] Following Martin, Traugott, and Kennedy (2005), the forecast error reported here is based on major party candidates. The log-odds formula is given in Martin, Traugott, and Kennedy (2005) as equation 1: $\log[(r/d)/(R/D)]$, where r and d refer to the number of respondents supporting the Republican and Democratic candidates, respectively, and where R and D refer to the number of votes cast for each candidate.

in results across the two sampling frames is statistically meaningful. One question is whether the differential error rates could occur by chance. If we assume that the six results are independent—a questionable assumption given that we have four pairs of polls—we can conduct an F-test. The test makes use of the fact that the square of the log-odds ratio divided by the square of its standard error is distributed chi-square, and the ratio of chi-squares follows an F distribution. We find the $F(6,6) = 5.7$ to be significant at the .05 level. On the other hand, if we exclude the South Dakota races for offices other than governor on the grounds that these poll results are non-independent, the $F(4,4) = 4.9$ falls short of the .05 threshold. Thus, we regard the overall pattern of results to be of borderline significance.

Needless to say, one cannot form a definitive judgment about the superiority of RBS based on just four statewide polls. What is interesting about this exercise is not so much the superior performance of RBS, which may be a sampling fluke, but the fact that RBS performed adequately despite the potential for debilitating bias. We find no evidence that RBS generated estimates that were biased for or against Republicans, incumbents, or front-runners. In its first direct empirical test RBS acquitted itself reasonably well and clearly warrants further experimentation in the future.

Cost Comparison

When comparing the costs of RDD and RBS polls, we assume that the polls are conducted in the same manner (e.g., computer-assisted telephone interviewing), by the same staff, using a questionnaire of similar length. The two polls differ in cost as a function of the expense incurred while creating a sample, recruiting respondents, and interviewing. Currently, RDD enjoys a cost edge in terms of sampling. Obtaining random phone numbers is relatively inexpensive. For example, the Maryland sample of RDD numbers cost $614, and the New York and Pennsylvania samples cost $1,200 apiece. Commercial firms are just beginning to offer stratified samples from registration lists, and no such service was available at the time of this study. The average cost of the registration lists in each state was approximately $1,900 in each state, including the costs of the phone match.[11]

The costs of recruiting respondents depend on whether an RBS survey is preceded by a letter encouraging respondents to participate. This letter can be costly to print and mail. The cost of sending the letter to roughly half of the RBS target list in each state was $1,956 in Pennsylvania, $2,003 in New York, and $2,500 in Maryland.[12] As we discovered during our ill-starred attempt to

11. This figure does not count the extra names that we purchased but consigned to an uncontacted control group in order to facilitate a separate study examining the effects of polling on voter turnout.

12. The Maryland letter was sent first-class because the survey followed soon after the mailing list became available.

poll the New York gubernatorial primary, the expense of the letter may go for naught if a leading candidate withdraws from the election.

In terms of labor costs the balance sheet favors RBS. TNS Intersearch, which conducted the *Washington Post* poll in Maryland, estimated that the cost per completed interview with a registered voter averaged $44.91 for RDD and $22.19 for RBS (Morin 2003), although, in fairness, the RBS instrument was much shorter than its RDD counterpart. The Quinnipiac poll, on the other hand, used questionnaires of comparable length and measured efficiency in terms of total interviewer hours associated with each poll. RDD (using CATI) consumed 633 hours in Pennsylvania and 510 in New York. RBS (using paper questionnaires) consumed 365 hours in Pennsylvania and 295 in New York. RBS and RDD generated comparable numbers of likely voters, but the labor costs required to do so were substantially lower for RBS.[13]

The potential cost advantages of RBS may be magnified when pollsters attempt to study specialized populations. In most states, where absentee voters constitute a small but important fraction of the electorate, it is prohibitively expensive to conduct an RDD survey of absentee voters. If, for example, 50 percent of the adults in a state vote in federal midterm elections, and only 20 percent of all voters cast absentee ballots, RDD must plow through approximately ten screening interviews in order to reach a single absentee voter. Similar arguments apply to surveys of young voters, minority voters, or voters who have recently switched parties. RBS seems particularly valuable for journalists or political marketers who are intent on studying small segments of the electorate—provided that phone numbers are available for these populations.

A similar point holds for populations defined by geographic boundaries that do not coincide with telephone exchanges. Consider, for example, the case of the 5th Congressional District in Connecticut. In the wake of redistricting, the borders of this district changed substantially in 2002, and one could not necessarily count on RDD respondents to report accurately whether they resided in the newly drawn district. Even if respondents could be trusted to reliably identify their congressional district, the RDD poll would have incurred additional costs when screening ineligible respondents. Address information from the registration rolls, on the other hand, could be fed into GIS software containing the boundary files of the new district. Thus, the Quinnipiac poll was able to use RBS to gauge opinion in the 5th District during the 2002 campaign.

Depending on the costs of the target lists and the number of completed interviews, RBS may offer an overall cost advantage. The cost savings from RBS should properly be seen as contributing to the accuracy of RBS polls because these savings free up resources that can be invested in additional interviews. The apparent forecasting advantages of RBS are in some sense understated because they do not take into account the ways in which reallocated resources might enhance RBS surveys.

13. Cost-efficiency estimates are not available for South Dakota.

Conclusion

Although we have used the term "registration" when referring to registration-based sampling, the idea behind RBS extends to any sort of list-based sampling frame (Green 2005). For example, in polities without voter registration, postal lists could be used as the basis for sampling. The statistical advantages of RBS, however, stem from the ancillary information about the prospective respondents that one gathers before interviews are conducted. Voter registration lists provide the voter's age, date of registration, and whether other voters are registered at the same address—all important predictors of voter turnout. Moreover, registration lists include addresses, which enable the pollster to find out a great deal about the political and economic climate in which a respondent lives and, by extension, how the potential respondent is likely to behave on Election Day.

This point has special resonance when pollsters attempt to measure the opinion of a select subgroup, a situation that arises in closed party primaries. The cost advantages of RBS polling in closed party primaries are especially promising. Costly as it is to locate voters in federal midterm elections using RDD, it is far more difficult to locate Democratic or Republican primary voters. Registration databases typically provide information about who has voted in past primaries—a strong predictor of who will turn out to vote in the upcoming primaries.

Less clear are the advantages of RBS in the context of high-turnout elections, such as presidential contests. Although the recent study by Mitofsky et al. (2005) suggests that random RBS performs adequately under these conditions, the cost savings of their statewide poll were less dramatic than in the midterm election studies reported here. Even more challenging is the prospect of conducting a nationwide survey using RBS. One could imagine conducting RBS surveys in battleground states (except for those states with same-day voter registration systems) because list vendors compete to provide registration lists and phone numbers to political campaigns in these states. Outside battleground states, lists are often spotty, forcing a nationally representative RBS survey to fall back on a clustered sampling design, which undercuts efficiency. Presidential election polls may eventually make use of dual-frame RBS/RDD designs, using RDD to fill in regional or demographic gaps within the RBS frame. Further research is needed in order to assess the accuracy and cost-effectiveness of these emerging sampling strategies.

Appendix

ANALYTICS OF USING POLLS FOR FORECASTING

This section describes the estimation problem that confronts random digit dialing (RDD) and registration-based sampling (RBS) polls. Our aim is to offer an accessible formal characterization of how the two sampling strategies differ. From this formal depiction we derive a series of observations about the conditions under which RBS and RDD provide accurate forecasts of the election outcome.

Suppose that an electoral jurisdiction has a traditional voter registration system that requires voters to register in advance of an election.[14] At the close of the registration period, the final list of registered voters is obtained and matched to a telephone directory, forming the sampling frame for an RBS survey. Meanwhile, a list of random telephone digits is appended to a random sample of telephone exchanges, forming the basis of an RDD survey. Let us now compare the attributes of the survey respondents with the attributes of actual voters.

The starting point for our analysis is a classification of the adult population into an exhaustive list of mutually exclusive categories, as shown in figure A1. Adults may be subdivided into registered and nonregistered voters. Registered voters may either vote or abstain, whereas nonregistered voters must abstain. This classification results in three categories: registered voters, registered abstainers, and unregistered abstainers. Within each of these three categories, we can distinguish between those whose telephone numbers are known to the pollster and those whose numbers are unknown.[15] Within each of these six groups, we may further classify people according to whether they will agree to be interviewed. This branching classification produces a total of 12 groups.

Each of these groups has a probability of supporting one of the two candidates $\{D, R\}$ running for office in a hypothetical election. We denote the probability that the kth group supports candidate D as p_k.[16] Let us consider the best-case scenario from the standpoint of polling, namely, that the p_k remains stable over time.

The actual vote outcome can be expressed as a weighted average of p_1, p_2, p_3, and p_4. In other words, actual voters comprise four groups: (1) those with known phone numbers who will participate in an interview if sampled, (2) those with known phone numbers who will not participate in an interview if sampled, (3) those with unknown phone numbers who will participate in an interview if sampled, and (4) those with unknown phone numbers who will not participate in an interview if sampled. Let α_i represent the share of the electorate that each group comprises. It follows that the actual vote (V) may be expressed:

$$V = \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + (1 - \alpha_1 - \alpha_2 - \alpha_3) p_4 \qquad [1]$$

To what extent do RBS and RDD surveys furnish information about the preference probabilities $\{p_1, p_2, p_3, p_4\}$ and the group weights $\{\alpha_1, \alpha_2, \alpha_3\}$? As figure A1 indicates, RBS results reflect a weighted average of p_1, the preferences of voters with listed numbers who accede to interviews, and p_5, the preference of nonvoters with listed numbers who accede to interviews. The weights that RBS accords to p_1 and p_5 depend on how the listed sample was constructed and whether likely voters are screened based on their survey responses. Consider the optimistic scenario whereby all of the voters pass the

14. This restriction excludes same-day registration states, which present a potential problem for RBS surveys because the registration list from which calls are made is not the effective registration list on Election Day. States with a same-day registration system or no registration might be candidates for RBS/RDD dual-frame designs.
15. By "known," we mean known correctly. If a caller asks for an RBS respondent by name, and the phone number is incorrect, we assume that no interview takes place.
16. For simplicity, we ignore third parties and "don't know" responses.

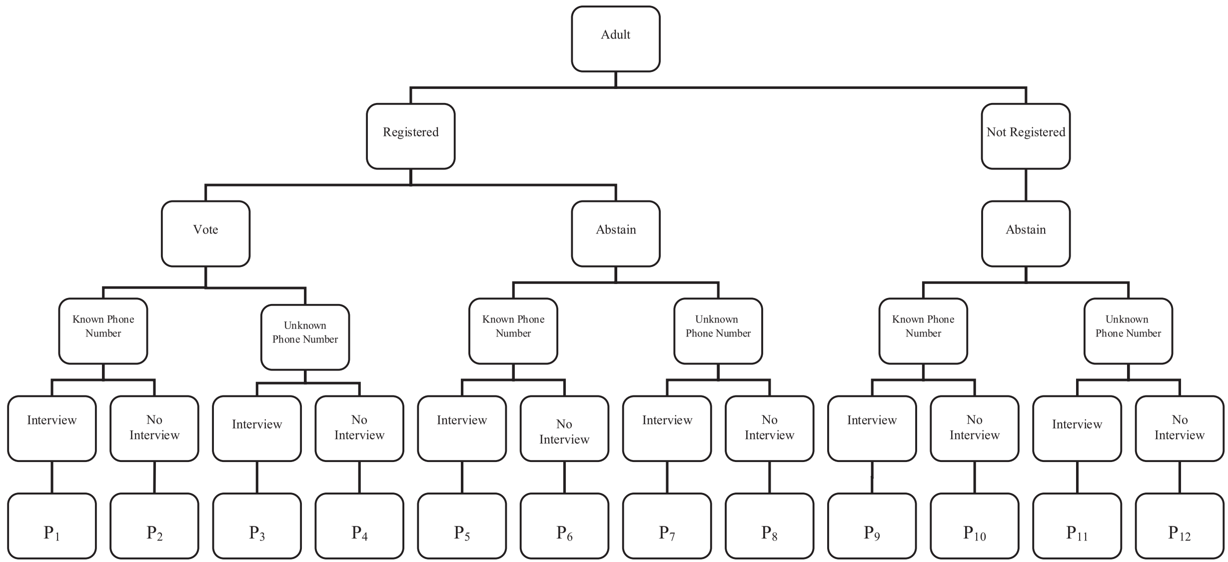

Figure A1. Schematic diagram of subgroups with RDD and RBS sampling frames.

screen and none of the nonvoters do. In that case, the RBS poll provides a consistent estimator for p_1. However, even in this special case, we see from equation 1 that p_1 equals V only when

$$p_1 = \frac{\alpha_2 p_2 + \alpha_3 p_3 + (1 - \alpha_1 - \alpha_2 - \alpha_3) p_4}{1 - \alpha_1} \qquad [2]$$

For example, an RBS poll would provide accurate results if $p_1 = p_2 = p_3 = p_4$. Thus, even under fairly optimistic assumptions, RBS produces unbiased results only under special conditions.

An analogous situation confronts RDD polls. RDD draws observations from a wider array of different population categories, all of which share a willingness to accede to interviews: (1) voters with known phone numbers, (2) voters with unknown numbers, (3) registered nonvoters with known numbers, (4) registered nonvoters with unknown numbers, (5) unregistered adults with known numbers, and (6) unregistered adults with unknown numbers. The main potential advantage of RDD is that it contains information about p_1 as well as p_3.[17] The main disadvantage is that its results include extraneous information about p_5, p_7, p_9, and p_{11}, which does not figure into the vote forecast except insofar as it is used to construct demographic sampling weights.

Suppose (optimistically) that it were possible to use survey responses to predict exactly who will vote. By excluding nonvoters, the RDD poll becomes a weighted average of p_1 and p_3. As sample size increases, the RDD likely voters poll result will converge to $\frac{\alpha_1 p_1 + \alpha_3 p_3}{\alpha_1 + \alpha_3}$. Although this RDD poll contains more information than the RBS poll, it nonetheless falls short of providing the information necessary to estimate V. The equality

$$V = \frac{\alpha_1 p_1 + \alpha_3 p_3}{\alpha_1 + \alpha_3} \qquad [3]$$

holds only under the special condition that

$$\frac{\alpha_1 p_1 + \alpha_3 p_3}{\alpha_1 + \alpha_3} = \frac{\alpha_2 p_2 + \alpha_2 p_4}{\alpha_2 + \alpha_4}. \qquad [4]$$

In other words, the accuracy of an RDD poll (with a perfect voter screen) depends on the relationship between the estimable quantities p_1 and p_3 and the unknown quantities p_2 and p_4. When the weighted average of p_1 and p_3 is the same as the weighted average of p_2 and p_4, an RDD poll that screens out nonvoters gives unbiased results. This result comports with the intuition that survey nonresponse leads to bias if the nonrespondents have different preferences than the respondents.

17. Another benefit of RDD is that it provides information about the proportion of voters whose numbers are known and who are willing to accede to an interview, which allows for consistent estimation of α_1, α_2, and α_3. However, these weights will estimated incorrectly if respondents do not accurately report whether they are registered.

Given the assumptions above, RDD seems to have an edge on RBS because the former provides unbiased results under somewhat more general conditions. This situation changes, however, when we relax the optimistic assumption that the likely voter screen functions perfectly.[18] A faulty voter screen allows registered nonvoters into the RBS poll and allows both registered and unregistered nonvoters into the RDD poll. A faulty voter screen may also mean that the category weights will be misestimated, insofar as respondents err on the side of exaggerating their probability of voting. Formally, the polls with imperfect screens may be represented as follows. The RBS poll converges to $\gamma_1 p_1 + (1-\gamma_1)p_5$, where γ_1 represents the expected weight of the registered voter category and $(1 - \gamma_1)$ represents the weight of registered nonvoter category. The bottom line is that RBS will provide accurate results if registered nonvoters with known phone numbers have preferences that are similar to voters with unknown numbers and voters who cannot be interviewed. The RDD poll is more complicated, because it comprises more groups:

$\gamma_1^* p_1 + \gamma_3^* p_3 + \gamma_5^* p_5 + \gamma_7^* p_7 + \gamma_9^* p_9 + (1-\gamma_1^* - \gamma_3^* - \gamma_5^* - \gamma_7^* - \gamma_9^*)p_{11}$. This time γ_k^* represents the expected weight associated with each sample category. Not surprisingly, given the number of parameters that generate the RDD poll result, this formula coincides with equation 1 only under special conditions. These conditions boil down to whether the weighted average of the extraneous groups (i.e., the nonvoters) pulls the weighted average of the voting groups away from the true outcome expressed in equation 1. If voters and nonvoters have similar preferences, the poll will be unbiased. If nonvoters have distinctive preferences and slip through the likely voter screen in large numbers, the bias can be quite serious.

In order to illustrate the potential biases of each sampling strategy, table A1 presents a series of scenarios in which different values of p_k and γ_k are assumed.

These examples affirm our general point that the bias associated with RDD and RBS polls can be large or small. In the first scenario, the population groups hold similar preferences, and both sampling strategies generate unbiased results. In the second scenario, respondents and nonrespondents have different preferences, and nonresponse bias throws off both RBS and RDD. In scenario 3, voters and nonvoters have different preferences, and bias is introduced because the likely voter screen is imperfect. In this case, RBS outperforms RDD because the RBS sample contains proportionally fewer nonvoters. In scenario 4, bias is introduced because people with known phone numbers have different preferences from those with unknown numbers. This situation leads to bias in the RBS poll but leaves the RDD poll unaffected.

Of course, more complex scenarios can be created with other combinations of parameters. Because the number and plausible range of unknown theoretical quantities is so large, it is difficult to say *ex ante* which sampling approach is superior. Typically—but not always—heterogeneity in preferences between voters and nonvoters will tend to favor the RBS approach, while heterogeneity in preferences among those with known and unknown phone numbers will favor RDD.[19] How much heterogeneity to expect from various types of election contests remains an open empirical question.

18. A parallel argument would be to relax the assumption that who is registered is known with certainty, whether from registration lists or questions put to the respondent.

19. Another consideration is preference change. Some segments of the electorate, perhaps those with little information, are more prone to change than others. If RDD samples contain a larger proportion of such respondents than RBS samples, the former will tend to produce more volatile estimates.

Table A1. Simulation of Survey Bias under Different Assumptions about Heterogeneity across Subgroups

			Scenario 1		Scenario 2		Scenario 3		Scenario 4	
Parameters and Outcomes	Surveyed by RBS?	Surveyed by RDD?	Homogeneity	Survey Results	Nonresponse Bias	Survey Results	Flawed Likely Voter Screen[a]	Survey Results	Bias Due to Unknown Phone Numbers	Survey Results
p_1	Yes	Yes	60%		60%		60%		60%	
p_2			60%		50%		60%		60%	
p_3		Yes	60%		60%		60%		50%	
p_4			60%		50%		60%		50%	
p_5	Yes	Yes	60%		60%		40%		60%	
p_6			60%		50%		40%		60%	
p_7		Yes	60%		60%		40%		50%	
p_8			60%		50%		40%		50%	
p_9		Yes	60%		60%		40%		60%	
p_{10}			60%		50%		40%		60%	
p_{11}		Yes	60%		60%		40%		50%	
p_{12}			60%		50%		40%		50%	
α_1			42%		42%		42%		42%	
α_2			18%		18%		18%		18%	
α_3			28%		28%		28%		28%	
Election Result				60%		57%		60%		56%
RBS				60%		60%		56%		60%
RBS: Likely Voters				60%		60%		59%		60%
RDD				60%		60%		52%		56%
RDD: Likely Voters				60%		60%		58%		56%

NOTE.—All of the scenarios assume the same basic population structure and response rates: 75 percent of the population is registered to vote, 60 percent of the population votes, 60 percent of all population groups have known phone numbers, and 70 percent consent to an interview.

[a] The likely voter screen admits 95 percent of voters, 30 percent of abstainers, and 10 percent of the unregistered.

With regard to the empirical comparison of the two methods, we would add one further consideration to the analytic discussion. To this point, we have focused exclusively on bias and have ignored the issue of sampling variance. There is potentially a trade-off between the two. Consider, for example, an RDD poll of 1,000 adults. In this sample, 750 people are registered, and 450 of them will vote. The sample of registered voters may produce biased results, but its larger size means that its margin of error is 23 percent smaller. Thus, by the criterion of mean-squared error (sampling variance plus bias), the sample of registered voters may provide more accurate results than the sample of actual voters.

Sampling variance is an important consideration to bear in mind when comparing RBS and RDD. Because the unit cost of an RBS interview tends to be lower, RBS tends to generate larger sample sizes under a given budget constraint. Furthermore, as we note below, RBS surveys need not rely exclusively on a likely voter screen administered during the course of an interview; instead, those conducting RBS polls can use information from the voter file to predict whether respondents are likely to vote. Because respondents are screened before they are interviewed, a very high proportion of RBS respondents are deemed likely voters when asked about their vote intentions. RDD, on the other hand, tends to drop a sizable fraction of the respondents who fail to meet the profile of a likely voter. Thus, consistent with our empirical results, the final RBS sample will tend to be substantially larger than the corresponding RDD sample of likely voters. This gives RBS a potentially important advantage in terms of forecasting accuracy.

References

American Association for Public Opinion Research (AAPOR). 2000. *Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Surveys*. Lenexa, KS: AAPOR.

Anderson, Andy, and Lydia K. Saad. 1996. “Checking Up on Respondents: A Voter Validation Study of the 1992 General Election.” Paper presented at the annual meeting of the American Association for Public Opinion Research, Salt Lake City, UT.

Bloom, Joel David. 2003. “Reliable Compared to What? Empirical Tests of the Accuracy of Election Polls, 2002.” Revised paper presented at the annual meeting of the American Association for Public Opinion Research, Nashville, TN.

Chang, LinChiat, and Jon A. Krosnick. 2001. “Improving Election Forecasting.” Unpublished manuscript, Ohio State University.

Crespi, Irving. 1988. *Pre-Election Polling: Sources of Accuracy and Error*. New York: Russell Sage.

Deane, Claudia, and Richard Morin. 2003. “Polls in Black and White: Examining the Differences in the Demographics of RDD and RBS Surveys in Maryland.” Paper presented at the annual meeting of the American Political Science Association, Philadelphia, PA.

Dimock, Michael, Scott Keeter, Mark Schulman, and Carolyn Miller. 2001. “Screening for Likely Voters in Pre-Election Surveys.” Paper presented at the annual meeting of the American Association for Public Opinion Research, Montreal, Canada.

Green, Donald P. 2005. “Registration-Based Sampling (RBS).” In *Polling America: An Encyclopedia of Public Opinion*, ed. Samuel J. Best and Benjamin Radcliff , pp. Westport, CT: Greenwood Press.

Groves, Robert M. 1990. “Theories and Methods of Telephone Surveys.” *Annual Review of Sociology* 16:221–40.

Lau, Richard R. 1994. “An Analysis of ‘Trial Heat’ Polls during the 1992 Presidential Election.” *Public Opinion Quarterly* 62:2–20.

Mann, Christopher. 2003. “Getting Pre-Election Surveys Right: The Effects of Advance Letters on Pre-Election Forecasting.” Paper presented at the annual meeting of the American Association for Public Opinion Research, Nashville, TN.

———. 2005. "Do Advance Letters Improve Preelection Forecast Accuracy?" *Public Opinion Quarterly* 69:561–71.

Martin, Elizabeth A., Michael W. Traugott, and Courtney Kennedy. 2005. "A Review and Proposal for a New Measure of Poll Accuracy." *Public Opinion Quarterly* 69:342–69.

Mitofsky, Warren. 1998. "The Polls—Review: Was 1996 a Worse Year for Polls than 1948?" *Public Opinion Quarterly* 62:230–49.

Mitofsky, Warren, Joel Bloom, Joseph Lenski, Scott Dingman, and Jennifer Agiesta. 2005. "A Test of a Combined RDD/Registration-Based Sampling Model in Oregon's 2004 National Election Pool Survey: Lessons from a Dual Frame RBS/RDD Sample." Paper presented at the annual meeting of the American Association for Public Opinion Research, Philadelphia, PA Miami Beach, FL.

Morin, Richard. 2003. "Smackdown in Maryland: RBS versus RDD." *Public Perspective* 6 (January/February): 7–9, 41.

O'Neill, Harry, Warren Mitofsky, and Humphrey Taylor. 2002. "National Council on Public Polls Polling Review Board Analysis of the 2002 Election Polls." December 19. Available online at http://www.ncpp.org/2002SenGovPoll/2002ElectionPolls.html (accessed April 11, 2006).

Schwartz, Doug, and Clay F. Richards. 2003. "The Big East Two: A Comparison of RBS and RDD Polls in the 2002 Elections in New York and Pennsylvania." Paper presented at the annual meeting of the American Political Science Association, Philadelphia, PA.

Thompson, Steven K. 1992. *Sampling*. New York: John Wiley and Sons.

Traugott, Michael W., Robert M. Groves, and James M. Lepkowski. 1987. "Using Dual Frame Designs to Reduce Nonresponse in Telephone Surveys." *Public Opinion Quarterly* 51:522–39.

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. New York: Chatham House.