

**Improving Pre-election Forecasts
From Registration Based Sampling:
Using Voter Registration Data to Predict Partisan Vote
Intention and to Allocate Undecided Voters**

Christopher B. Mann
Yale University

This paper was prepared for the 2003 Annual Conference of the American Political Science Association, Philadelphia, PA. I would like to thank Don Green for his inspiration, advice and generosity as a partner in our research into Registration Based Sampling. I would like to thank Kathy Frankovic and Anthony Salvanto of CBS News, Rich Morin and Claudia Deane of the Washington Post, and Doug Schwartz of Quinnipiac University Polling Institute for being enthusiastic partners in testing Registration Based Sampling. All errors of fact or judgment contained in this paper are solely my responsibility.

The inclusion and allocation of 'undecided' respondents has vexed pollsters for decades. The problem of undecided respondents is particularly acute for survey researchers conducting pre-election surveys, since the voting behavior of undecided voters can have a significant effect on the outcome of an election.

This paper will assess whether information drawn from voter registration rolls can be useful in allocating undecided voters. Voter registration rolls contain two types of information which could be valuable in allocating undecided voters. First, in many states voter registration information includes the declared partisan affiliation of each voter. Second, in most jurisdictions, the past voting history of registered voters is available. This history indicates which elections voters have participated in, but not how they voted. Since voting is habit forming, this past voting behavior is useful for identifying likely voters in the current election (Gerber, Green and Shachar 2003; Green & Gerber 2003). Furthermore, voting history from primary elections may be a valuable measure of active partisanship which contains more information than dichotomous party membership measures.

This paper will test the information from voter registration rolls using a series of pre-election surveys conducted prior to the 2002 General Election. These surveys were part of an effort to test the Registration Based Sampling (RBS) technique proposed by Green and Gerber as an alternative to traditional Random Digit Dialing (RDD) sampling (Green and Gerber 2003). Using data from three of RBS surveys, I will examine whether past general election voting history is a good predictor of turnout for the 2002 General Election, whether there are correlations between party registration and reported vote preference and between past primary voting history and reported vote preference, and finally whether using this information from the voter file to allocate undecided respondents provides more accurate estimates than current approaches to allocating undecided respondents.

Approaches to Identifying Likely Voters

The process of identifying likely voters in order to produce an accurate forecast of election results has been the topic of an extensive methodological literature (e.g. Anderson and Silver 1986; Bolstein 1991; Petrocik 1991; Monson 1998; Siegelman 1982; Siegelman 1985; Silver, Anderson and Abramson 1986; Traugott and Tucker 1984; Voss, Gelman and King 1995). One point of consensus in this literature is that being registered to vote is the most effective criteria for identifying likely voters. However, this consensus is not a solution since the mis-reporting of registration status (and other criteria for selecting likely voters) is quite common and introduces bias into RDD pre-election polls. If survey researchers could know with greater certainty whether voters were registered and then whether they were likely to turnout, pre-election surveys could be substantially less prone to error.

In the past decade, many jurisdictions in the US have begun to keep their voter registration records in electronic databases rather than the registration books of political lore. At a minimum, these databases contain the basic identifying information on voters: name, address, age, gender, and registration date. Often the voter registration data files included other information provided in the registration process such as party affiliation and phone numbers. Many of these databases also contain information about past voting history – that is in which past elections voters have cast a ballot. Some public agencies record all past elections since the creation of their computerized database, while others maintain only one or a few recent elections. When public agencies do not maintain a comprehensive voter history, commercial voter file vendors have often retained voter history for many past elections.

One of the difficulties in dealing with voter registration data is that it is often decentralized at the local level. Collecting data from dozens of local election clerks and compiling

it in a standard format is time consuming and expensive. Fortunately, due to candidates' demand for the ability to contact voters, commercial voter file vendors have sprung up to gather the voter registration data from local officials and make it available in a standardized statewide format. By spreading the costs of gathering the data among many candidates and other clients, voter file vendors can make voter registration data available at reasonable prices. As noted above, they have added value to their databases by maintaining voter history over time even when local public agencies only keep voting records for recent elections. Voter file vendors also add other valuable information to the voter file. The two most valuable pieces of information are phone numbers and updated addresses. Phone numbers are added by frequently matching voter registration records to phone company records to increase the number of phone numbers available. Current addresses are maintained by matching voter registration lists to the US Post Office's National Change of Address database to track voters who have moved.

At present, the coverage of the US with high quality electronic voter registration data is uneven. However, the recently passed federal election reform law requires that every state create a statewide electronic voter registration database. Thus, in the next few years, this resource should be available everywhere in the US – and probably at less cost than at present.

Private political polling firms have not been as oblivious to the technological development of electronic voter registration files as most public survey organizations appear to have been. For more than a decade, political pollsters have been drawing samples from voter files whenever they are available. Some political pollsters simply draw a random sample from among registered voters, thereby eliminating reliance on self-reported voter registration in identifying likely voters. Other pollsters utilize the past voting history as a way to enhance the identification of likely voters by selecting only among registered voters who have voted in one

or more past election plus new registrants. Since political pollsters keep their exact methods confidential to protect their advantages in the marketplace (Crespi 1988; Voss, Gelman and King 1995), it is impossible to test or even to know how well variations of this procedure work.

Green and Gerber have developed a Registration Based Sampling (RBS) methodology which utilizes past voting history from voter registration rolls to create a sample which reflects the likely electorate. This procedure is more sophisticated than the relatively crude threshold measures used by most political pollsters. The tests of the RBS method in the 2002 General Election show that it outperforms traditional RDD methods in the accuracy of election forecasts (Green and Gerber 2003). However, the basic premise underlying this model – the more frequently voters have voted in the past the more likely they are to vote in the current election – has not been analyzed yet from the 2002 election tests of their method.

Current Approaches to Allocating Undecided Respondents

Survey researchers have proposed a number of ways of reducing the problem of undecided respondents ranging from simply dropping them to random allocation to complex statistical allocation procedures. Each approach is grounded in theoretical arguments, and each has been supported with empirical evidence that it improves pre-election forecasts.

It has been common practice for several decades to encourage respondents who are reluctant to express their vote intention if they 'lean' towards a candidate. This reflects the belief that some people who have real preferences are reluctant to express them for some reason. For example voters may state they 'don't know' or are undecided because they do not want to do the cognitive work to think about it (Krosnick 1991) or because they believe that their candidate preference is not shared by others (Noelle-Neumann 1984). Visser et al (2000) report that allocating respondents who report leaning towards a candidate improve the

accuracy of pre-election forecasts. Thus for the analysis of the surveys here, I will categorize all respondents who stated that they lean towards one candidate as intending to vote for that candidate.

Survey researchers use a variety of approaches to handle those respondents who remain undecided even after they are encouraged to express a whether they are leaning towards a candidate. Crespi (1988, p109) found in his survey of 430 pre-election polls that polls which used some procedure to allocate undecided respondents were more likely to be accurate than polls which did not allocate undecided respondents. The most common approach to allocating undecided respondents was using reported party identification, with 19% of the pre-election polls in his sample using this technique (p110). Alternatively, allocation of undecideds may use candidate assessments or issue items from the same survey rather than the self-reported party identification. Crespi's interviews with survey researchers found that the primary reason not to allocate respondents was a concern with journalistic ethics - just reporting the facts – rather than concerns about survey methodology (p111).

Along these lines, Fenwick, Wiseman, Becker and Heiman (1982) proposed a multivariate modeling strategy in which the parameter estimates generated from a discriminant analysis model of the decided respondents is used to allocate the undecided respondents.¹ Despite Fenwick et al's claims that this process is objective, its weakness is the selection of the survey items to be used as independent variables in the model which involves a good deal of subjective judgment by the researcher. Furthermore, the survey items selected vary from survey to survey according to which will provide the most leverage, making the procedure unsystematic and difficult to know how to replicate.

¹ Vote intention is the dependent variable and a variety of survey items are the independent variables.

Another strategy is to allocate undecided respondents evenly between the candidates. This effectively assumes that respondents who are undecided reflect a portion of the electorate which will cast their ballots randomly (Visser et al 2000). This is the same as dropping them from both the numerator and denominator in calculating the percentages supporting each candidate.

Undecided respondents may be allocated as though they would split as the rest of respondents who expressed a vote intention are split. For example, if those who express a vote intention are split 60% to 40% the undecided respondents would be allocated 60-40. This approach will cause the allocated election forecast to show a wider winning margin than evenly splitting the undecideds.

Traugott and Tucker (1984) propose a more complex version of the last technique for allocating undecideds. They argue that partisan identification is very useful in predicting the vote, so undecideds should be divided by partisan affiliation then allocated according to the vote intentions of equivalent partisans who expressed a vote intention. For RDD, they find that this use of additional information about respondents to allocate undecideds provides more accurate pre-election forecasts.

Methodology and Data

The surveys analyzed in this paper were part of a series of pre-election surveys conducted by CBS News, Quinnipiac University Polling Institute, and the Washington Post² to test the Registration Based Sampling methodology (RBS) proposed by Green and against the traditional Random Digit Dialing (RDD) methodology Gerber (Green and Gerber 2003; Deane and Morin 2003; Mann 2003; Salvanto and Butterworth 2003; Schwartz and Richards 2003). In

² The now defunct Voter News Service also participated in these tests, but its dissolution immediately after the election has prevented anyone from accessing the data at this time.

each state, the above survey organization conducted a simultaneous surveys using procedures that were as identical as possible between RDD and RBS except for the sample called. The three surveys examined in this paper have slight differences in procedures since each survey organization used their standard procedures for the states in which they conducted the surveys. CBS News conducted the survey in South Dakota, Quinnipiac University Polling Institute conducted the Pennsylvania survey, and the Washington Post conducted the Maryland survey.

This paper analyzes the use of data available from voter registration records to improve survey forecasts, so only the RBS portions of the surveys are used. The RBS samples are drawn from voter registration lists maintained by two different sources. For Pennsylvania and Maryland, we used the voter registration data maintained by Voter Contact Services (VCS), a commercial vendor of voter registration data. VCS's data are gathered from the public agency responsible for maintaining the list of registered voters in every county in these two states. In South Dakota, the Secretary of State's office maintains an electronic database of voter registration information and past voting history that is regularly updated by each county clerk. I purchased a copy of this database in September 2002 from which to draw the RBS sample.

To improve efficiency in reaching likely voters, RBS samples can be pre-stratified by past voting behavior to reflect the likely composition of voter turnout for the upcoming election. For example, using the vote history and registration information, each registered voter South Dakota was placed into one of five strata for the 2002 general election: 1) voted in both of the 2000 and 1998 general elections; 2) voted in the 1998 general election, but not the 2000 general election; 3) voted in the 2000 general election, but not the 1998 general election; 4) registered to vote but had not voted in either the 2000 or 1998 general elections; 5) newly registered since the 2000 general election.

In order to make the sample as representative of likely voter turnout as possible, we drew a sample with the proportion of voters in each stratum reflecting the proportion of actual 2002 voters likely to fall into that stratum. To estimate the likely proportions, we looked at the 1998 electorate because it was the last similar (mid-term) election.³ We divided the 1998 list of registered voters into the same five strata, using the 1996 general election in place of the 2000 general election and the 1994 general election in place of the 1998 general election. We determined the proportions from each of the 5 strata in the 1998 electorate.⁴ We then drew a random sample of approximately 40,000 registered voters for the 2002 general election distributed among the strata according to the 1998 proportions.⁵ The sample size was large to ensure adequate control groups for other aspects of the broader research program on RBS.

In Maryland, the 2000 voter history was not available in many counties in the fall of 2002, so we stratified using the voting history of 4 and 6 years prior – 1992 and 1994 for 1998, 1996 and 1998 for 2002. This was not optimal, but was the best practicable option. In order to keep exposition and presentation of tables consistent and clear, the strata were recoded as described above using the 2000 voter history that became available after the 2002 election.

In Pennsylvania, 4 counties (making up 4.2% of registered voters in the state) have not released their 2000 voter history. For these 4 counties, three special strata were made up: voted in 1998, registered but did not vote in 1998, and registered since 1998. For the sake of clear exposition and presentation of tables, these special categories have been omitted from the

³ Ideally, we would have liked to make the following calculations for several prior mid-terms to get a more accurate estimate, but voter history is currently available only back far enough to make the calculation possible for the 1998 general election. Surveys in future elections will be able to draw on a longer record of voter history since commercial vendors and public agencies began maintaining computer readable voter history records for most jurisdictions in the 1990s.

⁴ We sought a complete list of all voters registered in 1998 from public agencies and commercial vendors, but were not able to locate one. The closest available approximation, which we used, was a list of registered voters from shortly before the 2000 general election. Some voters registered in 1998 were removed from this list prior to our calculation, but we could find no way of re-capturing them. However, the methodology should not require absolute precision in the proportions of each stratum to make a vast improvement in the efficiency and accuracy of the sample.

tables and discussion (the Pennsylvania vote history accordingly sums to 95.8% in the Tables below). Examining these special categories does not change the substance of any of the findings below.

During the calling, interviewers asked for the individual from the RBS sample by name. This was necessary to ensure that the individual level data from the voter registration records could be matched with the dispositions and responses from the survey. Traugott, Groves, and Lepkowski (1987) have shown that asking for a potential respondent by name does not alter response rates. The survey instrument used for each survey was the pre-election questionnaire on the 2002 General Election in each state designed by CBS News (South Dakota), Quinnipiac University Polling Institute (Pennsylvania), and the Washington Post (Maryland).⁶

TNS Intersearch in Horsham, PA conducted the interviews for the Maryland survey between October 21st and 25th, 2002. The Quinnipiac University Polling Institute in Mt. Carmel, CT conducted the interviews for Pennsylvania survey between October 21st and 27th. CBS News conducted the interviews for South Dakota between October 9th and 11th at their facilities in New York, NY. Interviewers were aware of the parallel RDD and RBS surveys being conducted and were rotated between the two surveys. For the Pennsylvania survey, each record received a maximum of 5 attempts. Callbacks were scheduled within the calling period whenever possible, but no refusal conversion attempts were made. In Maryland, each record received a maximum of 8 attempts, callbacks were scheduled within the calling period whenever possible, and refusal conversion was attempted after 48 hours. CBS used their standard procedures for

⁵ Due to rounding off our proportions in each stratum, we did not get exactly 40,000 records in each state.

⁶ The questionnaires were identical except for the introductions, because the individual selected for the RBS sample was requested by name while the RDD samples were randomized at the household level according to each survey organization's standard practice. In Maryland, the RBS questionnaire omitted a number of questions on candidate favorability and issues in the campaign between the initial horserace questions and the demographics battery. Since the number of terminations midway through the survey were similar for the long RDD and short RBS format and the calling house reported that almost all of these occurred in the first few questions, it seems unlikely that the survey length was a meaningful factor.

repeated attempts and callbacks, and did not attempt refusal conversions. In all three surveys, the names of the major party candidates were rotated to avoid ordering effects.

CBS News asked survey respondents in South Dakota if they had already voted via absentee. Any voter who had already voted was asked a separate set of questions regarding for whom they had voted. For this paper, absentee voters were pooled with traditional voters by coding absentee reported vote as the same thing as traditional voters' vote intentions. Quinnipiac University Polling Institute and the Washington Post asked all voters their vote intention without trying to identify voters who had already cast a ballot by absentee.

Past Voting History and Voter Turnout

Voter registration records provide two types of information that is useful to survey researchers. First, voter registration records contain demographic information such as age, gender, and date of birth.⁷ Presumably since this demographic information is provided for a legal purpose, it is more reliable than self-reported information from the survey. It can be used instead of self-reported information or to verify the self-reported information.

The second type of information available from public voter registration records is voting history. In most US jurisdictions, the public record includes whether or not each registered voter has cast a ballot in past elections. This voting history can be used for a number of purposes in pre-election surveys. Many commercial political survey firms use this information to select a sample of likely voters. Green and Gerber (2003) have developed a Registration Based Sampling (RBS) methodology which makes more refined use of past voting history.

Both the cruder threshold selection practices used by many commercial pollsters and the more refined methodology of Green and Gerber depend on the assumption that past voting

behavior reliably predicts future voter turnout. Table 1 shows that, for the two states examined in this paper, the more frequently voters have turned out in past general elections the more likely they were to turnout for the November 2002 general election.⁸

We can see that for both Maryland and Pennsylvania the likelihood of voting in the 2002 general elections rose with the number of prior elections in which a voter had cast a ballot. There is a clear difference in the turnout rate between the two states, but both exhibit the same pattern of rising turnout as the number of past general elections voted in increases. Newly registered voters fall neatly into this pattern between those who have not voted in any of the last four general elections and those who have voted in only one of the past four general elections. This seems appropriate in that newly registered voters are a mix of those who have been motivated to register but may never participate and those who will be or already are regular participants in elections. Therefore, from the evidence in these two states, it seems that past voting behavior can be a reliable indicator of future voting behavior.

Partisan Registration and Reported Partisan Vote Intention

While voting history for general elections provides information about likelihood of turnout, party registration and primary voting history also provide information about the partisan identification of registered voters. The meaning of self-reported partisan identification has been subject to considerable debate (e.g. Erikson, MacKuen and Stimson 2002; Green, Palmquist and Schickler 2003) which suggests that, at a minimum, other measures of partisanship without the problems of short term fluctuation and measurement could be useful. Even if we take self-reported partisan identification at face value, party registration and primary

⁷ Other information such as registration date can be used as a proxies for length of residence. Further, some states which are subject to Section 5 of the federal Voting Rights Act keep information about race and/or ethnicity in the voting records.

voting history provide behavioral rather than attitudinal measures of whether voters are active partisans. Using the data from Maryland, Pennsylvania and South Dakota, I will attempt to establish whether these behavioral measures of partisan identification can be used to improve the accuracy of pre-election survey forecasts.

First, I will look at the party registration. Party registration is expected to be a weaker indicator for two reasons. First, selection of a partisan affiliation occurs only when voters register (or re-register to change address, partisan registration or some other piece of information). Registration may have occurred some time ago for a substantial portion of registered voters. In the time since these voters registered, their partisan identification may have altered yet not be reflected in the public record. Second, partisan registration is expected to be a weaker indicator of partisan identification than primary voting history because party registration provides no indication about the strength of the voters affinity for the party while the frequency of primary voting history can be used as a measure of the strength of partisan identification.

Table 2 shows the cross-tabulation of respondents to the RBS survey who were registered Democrats, Republicans or Other⁹ and the partisan affiliation of the candidate they reported intending to vote for in the upcoming election. Maryland and Pennsylvania report on the gubernatorial race, while South Dakota reports the results from the gubernatorial race plus the US Senate and US House at large races.

The reported partisan vote intention is the partisan affiliation of the candidate for whom the voter said they intended to vote. This measure of partisan vote intention includes some error since voters decisions include consideration of the individual candidate as well as the

⁸ The vote validation process in South Dakota is not available at this time due to delays in the release of the 2002 vote history from the South Dakota Secretary of State's office.

⁹ Other includes those registered with minor parties, declared independents and voters who did not declare a partisan affiliation.

partisan affiliation. However, my concern here is only whether past partisan voting activity is a good predictor of future voting activity so this error is part of what I seek to assess.

For the three gubernatorial races, there appears to be wide variation in the amount of loyalty between voters registered with a party and that party's candidate. The Democratic candidate in South Dakota was garnering only 54% of the vote among Democrats, while the Republican gubernatorial candidate was racking up the support of 76% of his co-partisans. In Maryland, the Democrat was getting 63% of her co-partisans but the Republican was getting 78% of those registered with his party. In Pennsylvania, the Republican was getting 63% while the Democrat had a stronger performance with 73%. The wide range of performance suggests that voting based on partisan registration is not stable cross-sectionally in this election. It is also noteworthy that even the strongest performers among their co-partisans get barely more than three-quarters of the voters from voters registered with their party.

Although I have data on federal offices only from South Dakota, it appears that the difference in performance among voters registered to their party is smaller for federal office than for state (gubernatorial) office. If so, it suggests that partisan identities have a more consistent meaning at the national than the state level. However in South Dakota in 2002, it is equally possible that the more even performance of the candidates for the two federal offices is a function of these both being highly visible, hotly contested races while the governors race was a comparatively low key affair.

It is interesting to note in retrospect that in each case reported here, the candidate with the stronger support among the voters registered to his or her party was the winner of that race. This suggests that the range in performance among voters registered with a party is a function of the short term dynamics of the campaign rather than a longitudinally stable pattern.

Past Primary Voting History and Reported Vote Intention

Table 3 shows the cross-tabulation of the number of past partisan primaries voted in and the reported partisan vote intention. The pattern here is quite different than we saw in Table 1. In Table 1, increases in the number of past general elections in which a voter had cast a ballot was correlated with a higher likelihood of voting. In Table 3, voting in more of the last 4 partisan primaries does not appear to lead to a greater likelihood of support that party's candidate in the 2002 general election. Given the small cell sizes available, it appears that voting more frequently in past primaries has no effect on the likelihood of partisan vote choice. Here, the fact that no candidate is able to garner support from much over three quarters of his party's past primary voters becomes more significant: not even regular (4 of 4) past primary voters stand unanimously behind their party's nominee.

However, there is a clear difference in the voting behavior of those who have voted in at least one of the past primaries of the major parties and those voters who have not voted in any of the past four primaries. Table 4 distills Table 3 into those who voted in at least one Democratic primary, no past primaries, or at least one Republican primary. A comparison of Tables 2 and 4 reveals that past primary voting history has about the same predictive power as partisan registration for partisan candidate choice. The loyalty of past primary voters from each party to their party's nominee in the 2002 general election was almost identical to looking at the loyalty of those who were registered with each party. This suggests that party registration – as it appears on the voter rolls – and past primary voting history are roughly equivalent in predicting future partisan vote choice.

The finding that frequency of past primary voting history is not as strong a predictor of future partisan vote choice as past vote history is of future voter turnout is something of a

disappointment to advocates of RBS. However, the silver lining is that reliable party registration data is far more widely available than past primary voting history.

Accuracy of Reported Vote Intention

Green and Gerber (2003) have demonstrated that the election outcome predictions from RBS surveys are generally more accurate than predictions from traditional RDD methodologies. Table 5 is drawn from Green and Gerber (2003, Table 3) and reports the comparison of the RBS and RDD pre-election forecasts in each of these states.¹⁰ It displays the RBS and RDD results from simultaneous identical surveys which have been weighted or screened in various ways. Below the survey results for each race are two measures of pre-election forecast accuracy: 1) forecast error; and 2) margin error. The definition of forecast error used here follows Visser et al (2000) in calculating the average difference between the predicted and actual vote share of the major candidates in each race. Thus, forecast error is a measure of how exactly the survey forecasts the actual election results. Margin error is the difference between the predicted margin between the candidates and the actual margin between the candidates (Democrat minus Republican). This error indicates whether the survey captures who is leading and by how much – a fundamental concern of pre-election surveys.

Green and Gerber (2003) were primarily concerned with the margin error because the presence of undecided respondents meant that the forecast errors would be high in all of the races. Using this measure of pre-election forecasts, the unweighted RBS results are more accurate than any weighting or screening of these results in 4 out of 5 of the races analyzed here. More importantly, the unweighted RBS results are better predictions of the margin than

¹⁰ Green and Gerber (2003) included survey results from New York on the gubernatorial, attorney general and comptroller races. The New York data has not been included in this paper because the past voting history data currently available for New York is complete for general elections in the 1990s but not for primary elections in the same period.

the RDD results released by the survey organization in 4 of the 5 elections. Only in the South Dakota gubernatorial race does RDD prove to be more accurate – but only by 0.3 of a percentage point, while RBS was more accurate by 3 percentage points or more in the other 4 races.¹¹

For all of the results in these two sections, presence of undecided voters make the forecast errors are quite high. Since the ballot presents no option to be undecided and all but a few voters cast votes in all of the races as high on the ballot as those considered here, researchers in quest of a pre-election forecast that is accurate both for candidate vote share and for the margin must somehow allocate the undecided respondents. Table 6 displays the results when various allocation procedures are used.

Splitting the undecideds evenly between the candidates of the two parties has the same effect on the predicted margin as dropping the undecideds – that is the margin and, hence, the margin error remain unchanged. However, an even division significantly improves the forecast error of candidate vote share. Thus, even allocation creates a more accurate forecast of the actual election results than failing to allocate.

Relative to evenly splitting undecided respondents, allocating the undecided respondents according to the split in the preferences of the respondents who expressed a vote intention provides a better forecast error and margin error in Maryland and Pennsylvania but a worse forecast error in all of the South Dakota races and one tie, one increase and one decrease in the margin error. The difference between Maryland and Pennsylvania versus South Dakota could be a result of the fact that the Maryland and Pennsylvania polls were conducted in the last week and two weeks of the campaign respectively, while the South Dakota poll was conducted a month out from election day. In South Dakota, undecideds may have had more

¹¹ Unweighted RBS is also more accurate than unweighted RDD in 4 of the 5 races. Only in the Maryland gubernatorial race does unweighted RDD make a more accurate prediction of the margin – by a mere 0.4 percentage

time to make up their mind on other grounds than those used by the voters who expressed a vote intention in early October.

Traugott and Tucker (1984) proposed that rather than using the overall results to allocate undecideds, the electorate should be broken up by partisanship and undecideds within each partisan group should be allocated according to the vote intentions of equivalent partisans. Within the RBS data, we have found that there are two roughly equal options for parsing partisanship: party registration and past primary voting history.

As expected from the similarity seen in Tables 2 and 4, the results when undecideds are divided based on the vote intentions among party registrants or past primary voters are quite similar. The results when undecideds are allocated according to vote intentions of equivalent party registrants or equivalent past primary voters are nearly identical in Maryland and Pennsylvania, but have some variation in the South Dakota races that result in varying margin errors but are small enough to have little affect on the forecast errors. This difference again raises the possibility of survey timing playing a role in the results reported here.

Comparing allocation of undecideds by party registration or past primary voting history to allocating evenly or by the overall results reveals a mixture of results. Looking at the bottom two rows of Table 5 shows that the average absolute value of both the forecast and margin errors are little difference among these four approaches to allocation when aggregated among the five races analyzed here. If anything, the neutral approach of simply splitting the undecided voters evenly seems to have an edge over the other approaches. Overall, these results confirm that allocation of undecided enhances the accuracy of candidate share forecasts, but occasionally at a small cost of reducing the accuracy of predicting the margin between the two candidates.

points.

Conclusion

The apparent failure of voter registration data on past primary voting to show significant advantages in predicting partisan vote choice or in allocating undecided voters is a disappointment to proponents of the RBS method. However, the failure of registration data to resolve these problems only means that it shares the problems which have long plagued RDD. This paper shows that RBS is not a panacea to all of the difficulties in RDD – some of these are difficulties of pre-election surveys, not the sampling methodology. However, the inability to resolve some common problems of pre-election polling should not detract from the fact that RBS has many advantages in identifying likely voters, making accurate election forecasts, and reducing survey cost. RBS moves forward, but leaves plenty of room for other innovative methodologists to resolve our remaining problems such as optimal allocation of undecided respondents.

BIBLIOGRAPHY

- Anderson, Barbara A. and Brian D. Silver. 1986. "Measurement and Mismeasurement of the Validity of Self-Reported Vote". *American Journal of Political Science* 30 (4):771-785.
- Bolstein, Richard. 1991. "Comparison of the Likelihood to Vote Among Preelection Poll Respondents and Nonrespondents". *Public Opinion Quarterly* 55:648-650.
- Chang, LinChiat, and Jon A. Krosnick. 2001. "Improving Election Forecasting". Center for Survey Research, The Ohio State University. Unpublished Manuscript.
- Crespi, Irving. 1988. *Pre-Election Polling: Sources of Accuracy and Error*. New York: Russell Sage Foundation.
- Deane, Claudia and Richard Morin. 2003. "Polls in Black and White: Examining the Differences in the Demographics from RDD and Registration Based Sampling". Presented at the Annual Meeting of the American Association of Public Opinion Research. Nashville, TN.
- Erikson, Robert S., Michael B. MacKuen, and James A. Stimson. 2002. *The macro polity*. New York: Cambridge University Press.
- Fenwick, Ian, Frederick Wiseman, John F. Becker, and James R. Heiman. 1982. "Classifying Undecided Voters in Pre-election Polls". *Public Opinion Quarterly* 46:383-391.
- Green, Donald and Alan Gerber. 2003. "Enough Already with Random Digit Dialing: Using Registration-Based Sampling to Improve Pre-Election Polling". Presented at the Annual Meeting of the American Association of Public Opinion Research. Nashville, TN.
- Green, Donald, Bradley Palmquist, and Eric Schickler. 2002. *Partisan hearts and minds: political parties and the social identities of voters*. New Haven, CT: Yale University Press.
- Gerber, Alan S., Donald P. Green, and Ron Shachar. 2003. Voting May be Habit Forming: Evidence from a Randomized Field Experiment. *American Journal of Political Science* 47(3): 540-550.
- Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213-236.
- Mann, Christopher B. 2003. "Getting Pre-Election Surveys Right: The Effects of Advance Letters on Pre-election Forecasting". Presented at the Annual Meeting of the American Association of Public Opinion Research. Nashville, TN.
- Monson, Quin. 1998. "Predicting Voter Turnout in Local Elections: A Vote Validation Study". Presented at the Annual Meeting of the Midwest Association for Public Opinion Research, Chicago, IL.
- Noelle-Neuman, Elisabeth. 1984. *The Spiral of Silence: Public Opinion, Our Social Skin*. Chicago, IL: University of Chicago Press.
- Petrocik, John R. 1991. "An Algorithm for Estimating Turnout As A Guide to Predicting Turnout". *Public Opinion Quarterly* 55:643-637.
- Salvanto, Anthony and Michael Butterworth. 2003. " ". Presented at the Annual Meeting of the American Association of Public Opinion Research. Nashville, TN.
- Schwartz, Douglas, and Clay Richards. 2003. "How Likely is a Likely Voter to Vote: A New Model of Likely Voter Screens". Presented at the Annual Meeting of the American Association of Public Opinion Research. Nashville, TN.
- Siegelman, Lee. 1982. "The Nonvoting Voter in Voting Research". *American Journal of Political Science* 26 (1):47-56.
- Siegelman, Lee, Phillip W. Roeder, Malcolm E. Jewell and Michael A. Baer. 1985. "Voting and Nonvoting: A Multi-Election Perspective". *American Journal of Political Science* 29 (4):749-765.

- Silver, Brian D., Barbara A. Anderson, and Paul R. Abramson. 1986. "Who Overreports Voting?" *American Political Science Review* 80 (2):613-624
- Traugott, M. W., R. M. Groves, and J. M. Lepkowski. 1987. Using Dual Frame Designs to Reduce Nonresponse in Telephone Surveys. *Public Opinion Quarterly* 51 (4):522-539.
- Traugott Michael W. and Clyde Tucker. 1984. "Strategies for Predicting Whether a Citizen Will Vote and Estimation of Electoral Outcomes". *Public Opinion Quarterly* 48:330-343.
- Visser, Penny S., Jon A. Krosnick, Jesse Marquette, and Michael 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*, edited by M. W. Traugott. New York: Chatham House Publishers.
- Voss, D. Stephen, Andrew Gelman, and Gary King. 1995. "Preelection Survey Methodology: Details from Eight Polling Organizations, 1999 and 1992". *Public Opinion Quarterly* 59 (1):98-132.

Table 1
Using Past General Election Voting to Predict 2002 General Election Turnout

2002 General Election Turnout (%)

		Maryland		Pennsylvania		South Dakota	
		% Did Not Vote	% Voted	% Did Not Vote	% Voted	% Did Not Vote	% Voted
Number of Past General Elections Voted	0 <i>n</i>	66	34	94	6		
			3754		1974		
	Newly Registered <i>n</i>	56	44	76	24		
			4518		2552		
	1 of 4 <i>n</i>	53	47	66	34		
			4932		4838		
	2 of 4 <i>n</i>	38	62	51	49		
		5398		5666			
3 of 4 <i>n</i>	21	79	38	62			
		11748		9593			
4 of 4 <i>n</i>	12	88	19	81			
		9362		15358			
Total <i>n</i>	33	67	41	59			
		39712		39981			

Note: This data is from Registration Based Samples which are pre-stratified to reflect the likely composition of the 2002 general election turnout.

Table 2
Using Party Registration to Predict 2002 Partisan
Choice

State/ Office	Partisan Vote Intention (%)	Partisan Registration		
		Democrat	Republican	Other
Maryland Governor	Democrat	63	14	43
	Republican	30	78	51
	Undecided/ Other	7	8	6
	<i>n</i>	418	250	70
	<hr/>			
Pennsylvania Governor	Democrat	73	28	36
	Republican	16	63	36
	Undecided/ Other	11	9	29
	<i>n</i>	334	369	42
	<hr/>			
South Dakota Governor	Democrat	54	12	40
	Republican	19	76	35
	Undecided/ Other	26	12	26
	<hr/>			
	South Dakota US Senate	Democrat	69	17
Republican		9	65	30
Undecided/ Other		22	17	35
<hr/>				
South Dakota US House		Democrat	61	21
	Republican	18	65	47
	Undecided/ Other	21	15	16
	<i>n</i>	175	220	43
	<hr/>			

Table 3
**Using Frequency of Past Primary Participation to Predict 2002 Partisan
 Vote Choice**

State/ Office	Partisan Vote Intention (%)	Democrat					Republican				Total
		4 of 4	3 of 4	2 of 4	1 of 4	0	1 of 4	2 of 4	3 of 4	4 of 4	
Maryland Governor	Democrat	64	64	65	56	49	15	15	13	9	45
	Republican	31	30	23	38	44	77	74	76	88	48
	Undecided/ Other	4	6	13	6	7	8	11	11	3	7
	<i>n</i>	70	90	79	81	214	48	46	46	64	738
Pennsylvania Governor	Democrat	72	79	63	71	47	32	29	12	36	49
	Republican	17	16	22	15	40	65	60	76	51	40
	Undecided/ Other	12	5	15	13	13	3	10	12	14	11
	<i>n</i>	60	63	59	52	270	65	68	49	59	745
South Dakota Governor	Democrat	63	50	52	64	31	15	12	16	9	32
	Republican	13	19	13	18	49	68	76	71	83	49
	Undecided/ Other	25	31	35	18	20	17	12	13	9	19
	<i>n</i>	63	63	83	67	40	17	9	26	22	40
South Dakota US Senate	Democrat	63	63	83	67	40	17	9	26	22	40
	Republican	25	13	0	5	38	55	74	61	65	40
	Undecided/ Other	13	25	17	28	22	28	18	13	13	21
	<i>n</i>	56	56	70	62	38	23	21	23	26	39
South Dakota US House	Democrat	56	56	70	62	38	23	21	23	26	39
	Republican	31	19	9	18	45	62	65	61	57	44
	Undecided/ Other	13	25	22	21	17	15	15	16	17	17
	<i>n</i>	16	16	23	39	209	47	34	31	23	438

Table 4**Using Participation In At Least One Past Primary to Predict
2002 Partisan Vote Choice**

State/ Office	Partisan Vote Intention (%)	Democrat		Republican	
		At Least 1 of 4	0	At Least 1 of 4	1
Maryland Governor	Democrat	62	49	13	45
	Republican	31	44	79	48
	Undecided/ Other	7	7	8	7
	<i>n</i>	320	214	204	738
Pennsylvania Governor	Democrat	71	47	28	49
	Republican	18	40	62	40
	Undecided/ Other	11	13	10	11
	<i>n</i>	234	270	241	745
South Dakota Governor	Democrat	59	31	13	32
	Republican	16	49	73	49
	Undecided/ Other	26	20	13	19
	<i>n</i>				
South Dakota US Senate	Democrat	69	40	18	40
	Republican	9	38	63	40
	Undecided/ Other	22	22	19	21
	<i>n</i>				
South Dakota US House	Democrat	62	38	23	39
	Republican	18	45	61	44
	Undecided/ Other	20	17	16	17
	<i>n</i>	94	209	135	438

TABLE 5
 Comparison of Actual and Projected Results (From Green and Gerber 2003, Table 3)

<u>State-Office</u> (total ballots)	<u>Candidate</u>	<u>Party</u>	<u>Incumbent</u>	Actual Vote¹	Registration Based Sampling (Undecideds Not Allocated)					Random Digit Dialing (Undecideds Not Allocated)	
					<u>Unweighted</u>	<u>Weighted to Original Strata Proportions</u>	<u>Weighted to All Registered Voters</u>	<u>Using "definitely vote" screen</u>	<u>Among Actual 2002 Voters</u>	<u>Unweighted</u>	<u>Using likely voter screen²</u>
MD-Governor	Townsend	D		47.7%	44.6%	45.9%	33.0%	44.6%	45.1%	45.6%	49.0%
	Ehrlich	R		51.6%	48.0%	45.3%	42.6%	48.4%	47.6%	49.4%	49.4%
	Forecast Error				3.3	4.0	11.8	3.1	3.3	2.1	0.4
Margin Error					-0.5	-4.5	5.7	-0.1	-1.4	-0.1	-3.5
PA-Governor	Rendell	D		53.4%	48.6%	50.1%	55.3%	49.8%	49.8%	51.4%	54.0%
	Fisher	R		44.4%	40.3%	39.7%	36.7%	39.2%	40.5%	32.9%	35.0%
	Forecast Error				4.5	4.0	2.9	4.4	3.7	6.7	4.4
Margin Error					0.7	-1.3	-9.5	-1.6	-0.3	-9.4	-10.0
SD - Governor	Abbott	D		41.9%	31.5%	31.1%	30.2%	31.3%	Data Not Available	37.4%	36.3%
	Rounds	R		56.8%	49.3%	50.4%	44.2%	51.1%	Data Not Available	44.3%	48.4%
	Forecast Error				8.9	8.6	12.2	8.1	Data Not Available	8.5	7.0
Margin Error					3.0	4.5	-0.9	5.0	Data Not Available	-7.9	-2.7
SD-Senate	Johnson	D	x	49.6%	39.5%	39.3%	42.0%	41.3%	Data Not Available	43.8%	44.8%
	Thune	R		49.5%	39.5%	40.1%	35.9%	40.0%	Data Not Available	38.6%	41.1%
	Forecast Error				10.0	9.8	10.6	8.9	Data Not Available	8.3	6.6
Margin Error					0.2	1.0	-5.9	-1.1	Data Not Available	-5.0	-3.5
SD - House	Herseith	D		45.6%	38.6%	38.5%	36.7%	40.0%	Data Not Available	43.4%	43.1%
	Janklow	R		53.5%	44.3%	43.4%	48.2%	43.2%	Data Not Available	41.8%	44.5%
	Forecast Error				8.1	8.6	7.1	7.9	Data Not Available	6.9	5.7
Margin Error					-2.1	-2.9	3.6	-4.6	Data Not Available	-9.4	-6.4
Average Absolute Forecast Error (All Races)					7.0	7.0	8.9	6.5	Data Not Available	6.5	4.8
Average Absolute Margin Error (All Races)					1.3	2.9	5.1	2.5	Data Not Available	6.4	5.2

Sample Sizes

	Registered Only									
MD	738	738	738	738	657	590 (preliminary)	960	725		
PA	745	745	745	745	687	620	1214	636		
SD	438	438	438	438	380	-	438	353		

Notes:

- 1 - The Actual Vote reflects the percentage of the vote received by the major candidates. It may not sum to 100% because minor party candidates were excluded. The results were downloaded from the state elections office website in each state.
- 2 - This column reflects post-survey weighting of the data for vote likelihood and other factors as done by the Washington Post (MD), and the Quinnipiac Poll (PA) for public release. An index of several questions about likelihood to vote was used in Pennsylvania, while a single question about likelihood to vote ("How likely are you to vote...") was used in Maryland, and South Dakota.

TABLE 6

Comparison of Actual and Projected Results with Undecideds Allocated

<u>State-Office</u> (total ballots)	<u>Candidate</u>	<u>Party</u>	<u>Incumbent</u>	Actual Vote¹	Registration Based Sampling (Undecideds Allocated)			
					<u>Undecideds Allocated by Past Primary Groups</u>	<u>Undecideds Allocated by Party Registration</u>	<u>Undecideds Allocated by Overall Results</u>	<u>Undecideds Split Evenly</u>
MD-Governor	Townsend	D		47.7%	47.8%	47.8%	47.7%	48.1%
	Ehrlich	R		51.6%	51.6%	51.6%	51.3%	51.5%
Forecast Error					-0.1	-0.1	0.1	-0.2
Margin Error					-0.1	-0.1	-0.3	-0.5
PA-Governor	Rendell	D		53.4%	54.1%	54.1%	53.9%	54.1%
	Fisher	R		44.4%	44.6%	44.5%	44.7%	45.8%
Forecast Error					-0.5	-0.4	-0.4	-1.1
Margin Error					-0.5	-0.6	-0.2	0.7
SD - Governor	Abbott	D		41.9%	38.2%	38.9%	37.6%	41.1%
	Rounds	R		56.8%	57.9%	56.9%	58.8%	58.9%
Forecast Error					1.3	1.4	1.1	-0.7
Margin Error					4.9	3.2	6.4	3.0
SD-Senate	Johnson	D	x	49.6%	48.0%	48.3%	47.8%	50.0%
	Thune	R		49.5%	47.6%	47.0%	47.8%	50.0%
Forecast Error					1.7	1.9	1.7	-0.5
Margin Error					-0.2	-1.1	0.2	0.2
SD - House	Herseth	D		45.6%	45.4%	45.7%	45.2%	47.1%
	Janklow	R		53.5%	51.6%	51.3%	51.9%	52.9%
Forecast Error					1.0	1.0	1.0	-0.5
Margin Error					-1.6	-2.2	-1.1	-2.0
Average Absolute Forecast Error (All Races)					0.9	1.0	0.9	0.6
Average Absolute Margin Error (All Races)					1.5	1.4	1.6	1.3

Sample Sizes

MD	738	738	738	738	738
PA	745	745	745	745	745
SD	438	438	438	438	438

Notes:

1 - The Actual Vote reflects the percentage of the vote received by the major candidates. It may not sum to 100% because minor party candidates were excluded. The results were downloaded from the state elections office website in each state.