Getting Pre-Election Surveys Right: The Effects of Advance Letters on Pre-election Forecasting

Christopher B. Mann
Department of Political Science
Yale University

July 11th, 2003

The author gratefully acknowledges the assistance of Voter Contact Services in providing voter registration data, and the collaboration with the Washington Post and Quinnipiac University Polling Institute in conducting the surveys. I would like to thank Brett Clifton, Donald Green, Rachel Milstein, and David Nickerson for their suggestions that have improved this article. Any remaining errors of fact or judgment are my responsibility. Previous versions of this article were presented at the American Association of Public Opinion Research Annual Conference, Nashville, TN, May 2003 and the New England Political Science Association Conference, Providence, RI, May 2003. This research was supported by a grant from the Smith-Richardson Foundation, which bears no responsibility for the conclusions.

Abstract:
The survey methodology literature has debated whether advance letters to potential survey respondents will reduce non-response bias – which should improve the accuracy of pre-election forecasts. This article analyzes the results of experiments conducted in Maryland, New York and Pennsylvania in which advance letters were sent to a random sample of potential survey respondents for 2002 pre-election surveys. The survey samples were drawn from voter registration records containing demographic information not usually available to researchers. I find a significant increase in the overall response rate, but important variation in the effect across the demographic sub-groups. The sub-sample that was sent an advance letter more accurately predicts Republican candidate vote share than Democratic candidate vote share, while the reverse is true in the sub-sample not sent a letter. The sub-sample that was sent a letter consistently over-predicts the winner's margin, while the sub-sample that was not sent a letter consistently under-predicts the winner's margin. The results suggest that splitting the sample between letter and no letter samples balances the opposing partisan biases in forecasting candidate vote share and will allow researchers to use the pattern of over- and under-prediction to bracket the most accurate forecast of election outcomes.

Public opinion surveys conducted by media organizations and academic research institutes prior to elections are among the most high profile surveys in the United States. Pre-election surveys are not merely a neutral snapshot of the world as it is, but interact with the on-going political process in important ways. Candidates with strong survey showings have an easier time raising money, gain bandwagon supporters, and receive many other advantages. Conversely, weak survey showings may damage a candidate. Therefore it is imperative that pre-election surveys be as accurate as possible so that survey biases do not unfairly affect election outcomes.

Pre-election surveys are subject to methodological problems that can unintentionally introduce bias into any public opinion survey. Given the sophistication and care of most media and public survey organizations in conducting telephone surveys, the primary methodological concern is non-response bias. This article tests the hypothesis that advance letters to potential survey respondents will reduce non-response and the problems that arise from it.

Past studies have found strong evidence that respondents and non-respondents may differ in their interest in politics and their political preferences (e.g. Brehm 1993; Voogt and Van Kempen 2002). The existence of such differences undermines the external validity of survey data for drawing inferences about the population of voters as a whole rather than just the portion of voters like the respondents.

A variety of techniques can be used to reduce non-response rates: increasing numbers of callbacks, interviewer quality, refusal conversions, tailoring, paid incentives, extended fielding periods, and advance letters (Church 1993; Dillman 2000; Groves and

Couper 1998; Keeter et al. 2000; Lynn 2001; Singer et al. 1999; Singer, Van Hoewyk, and Maher 2000).[1] These techniques have been found effective in increasing the response rates to mail surveys in series of studies dating back to the 1960s (for a meta-analysis of mail surveys, see Church 1993).

Past studies of the effects of an advance letter on non-response in telephone surveys are inconclusive.[2] Several studies have shown increases in the overall response rate of 5 to 13 percentage points (Dillman, Gallegos, and Frey 1976; Goldstein and Jennings 2002; Groves and Couper 1998; Traugott, Groves, and Lepkowski 1987). Other studies have shown no statistically significant effect (Singer, Van Hoewyk, and Maher 2000; Sykes and Hoinville 1985). An early study by Brunner and Carroll (1969) found a large positive effect for an advance letter on university letterhead, but a negative effect for a letter on market research firm letterhead. Each of the three experiments conducted in 2002 has a larger sample size than the largest past study. Therefore, the results of these experiments have more statistical power to determine the effect of advance letters.

Non-response rates are a function of both difficulty in contacting potential respondents and refusal to complete the survey. The early literature on non-response was primarily concerned with refusal rates (e.g. Dillman, Gallegos, and Frey 1976). However, times have changed so that difficulty in contacting potential respondents is

[1] Keeter et al's (2000) recent study found that cash incentives sent to potential respondents with listed numbers, refusal conversion attempts, increasing the number of call backs, and extending the time period for fielding a national survey significantly increased the overall response rate. However, there were only minor and statistically insignificant impacts on the distribution of political attitudes, behavior, knowledge and engagement, even though there were some significant differences in the demographic variables due to the treatments.

[2] There are several psychological reasons why we expect that an advance letter would increase participation in a survey: authority and legitimacy of the sponsor, consistency with respondents' commitments to

driving much of the continuing decline in response rates. Some of this decline may be attributable to busier contemporary lifestyles, but researchers have also found evidence that households with answering machines, caller-ID, and fax/modem lines are harder to contact (Link and Oldendick 1999; Tuckel and O'Neill 2002).

Techniques to increase the cooperation rate may be successful in reducing refusals but not in increasing the contact rate, or vice versa. Non-response reduction techniques may help improve the accuracy of surveys by cutting down bias due to non-response. Conversely, variation in the reduction of non-response may reinforce current biases or create new biases. Thus survey researchers must look beyond increasing the overall response rates to examine the magnitude and direction of effects on contact and refusal across demographic sub-groups (Groves, Singer, and Corning 2000; Groves and Couper 1998).

Using data from experiments conducted for 2002 pre-election surveys in Maryland, New York and Pennsylvania, I assess: 1) the effect of an advance letter on contact, cooperation and refusal rates; 2) the effect on the representativeness of survey participants; and 3) whether the advance letter improves the accuracy of the forecasts from pre-election surveys. I begin with a description of the methodology used in the experiments on the effect of an advance letter, including situating these experiments within a larger research program on Registration Based Sampling (RBS) survey methodology. I will then present the findings from a combined data set of the three states.[3] I will discuss several important concerns raised by the findings before concluding with what the findings suggest for future practices in pre-election telephone surveys.

---

participate in the community and/or politics, and reciprocity. See Groves, Cialdini and Couper (1992) for a detailed discussion.

Methodology

The experiments testing the effect of an advance letter are part of a larger experimental research program on survey methodology. The main focus of this broader research program is comparing Registration Based Sampling (RBS) with traditional Random Digit Dialing (RDD) samples while using parallel and identical survey instruments, calling protocols, phone houses, and other methodology. The experiments on the effect of an advance letter were conducted within the RBS samples in each state.

Since RBS samples are drawn from voter registration rolls, they contain more demographic information about potential respondents than RDD samples of computer generated phone numbers. Previously used methods of studying advance letters such as matching RDD generated phone numbers to phone company records (e.g. Singer, Van Hoewyk, and Maher 2000) or drawing a sample from among listed phone numbers (e.g. Dillman, Gallegos, and Frey 1976; Traugott, Groves, and Lepkowski 1987) provide phone numbers, names and addresses, but no demographic information about potential respondents. RBS captures name, address, date of birth, registration date, gender, and past voting history[4] from the voter registration rolls. Unlike past studies, the demographic information available from voter rolls allows an examination of differences in the effect of an advance letter across demographic sub-groups of respondents *and* non-respondents (see also Goldstein and Jennings 2002).

Sampling and Stratification Using Registration Rolls

Voter Contact Services (VCS), a commercial vendor of voter registration data, maintains a list of all registered voters and the vote history of individual registered voters

[3] The results for each state can be found in the appendices.

in Maryland, New York and Pennsylvania. VCS's data are gathered from the public agency responsible for maintaining the list of registered voters in every county in each state.

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 in New York 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, I 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, I looked at the 1998 electorate because it was the last similar (mid-term) election.[5] 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. I determined the proportions from each of the 5 strata in the 1998 electorate.[6] I then drew a random sample of approximately 40,000 registered voters for

[4] Voter history records indicate that an individual cast a ballot in a particular election, but reveals nothing about how the voter cast their ballot.

[5] 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.

[6] 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

the 2002 general election distributed among the strata according to the 1998 proportions.[7] 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 I 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 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.

Conducting the Letter Experiment

In order to select those voters in the survey sample, the data file was sorted according to those with phone numbers and those without phone numbers. VCS's voter registration lists include any phone number provided by voters as part of the registration process. VCS also matches the voter registration data to listed phone numbers to capture

---

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.

additional valid phone numbers quarterly. In each state the percentage of registered voters with valid phone numbers was: Maryland 64.7%; New York 68.9%; and Pennsylvania 66.3%. The survey sample was 10,000 records selected randomly from among the registered voters in the RBS sample with valid phone numbers. Selecting among only the records with phone numbers introduces a potential bias as past studies have found that people with listed and unlisted phone numbers respond differently to surveys (e.g. Camburn et al. 1995).

Within the survey sample, records were randomly selected to be sent an advance letter. The letter identified the sponsors of the survey (the Washington Post and Yale University in Maryland and the Quinnipiac University Polling Institute in New York and Pennsylvania), described the purpose of the survey as providing public information about the upcoming gubernatorial election, stated that completion of the survey would take less than 10 minutes, and advised the recipient to expect a call to complete a survey during the calling period of each survey.[8] In Maryland and Pennsylvania, 3,000 records were selected randomly from the survey sample and sent advance letters via first class mail a week before the survey began. In New York, 5,000 records were randomly selected from the survey sample and were sent advance letters via bulk rate mail ten days before the survey began.[9] All selected registered voters were sent a personalized letter in a white

[7] Due to rounding off our proportions in each stratum, we did not get exactly 40,000 records in each state.
[8] The Pennsylvania survey was in the field October 21-27, 2002. The Maryland survey was in the field October 21-25, 2002. The New York survey was in the field October 28-November 3, 2002.
[9] In Maryland, 3,000 records were randomly selected to be sent a letter and then attempts were made by randomly selecting from this sub-sample. Attempts were also made by drawing randomly from among the 7,000 records not sent a letter, such that equal numbers of surveys were completed in the letter and non-letter sub-samples. In Pennsylvania, the 3,000 records were randomly selected to be sent a letter. Another 3,000 records were randomly selected to create a sub-sample of 6,000 records evenly split between being sent a letter and not being sent a letter which was used for randomly selected attempted dialings. The remaining 4,000 records in the survey sample were a "reserve" sub-sample for attempts which was never released for dialing. In New York, 5,000 records were randomly selected to be sent a letter, and then

envelope with an address window.[10] The envelope provided only the name of the sponsor and return address.[11] (See Appendix A for copies of the letters)

Conducting the Survey

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 standard pre-election questionnaire on the gubernatorial race in each state designed by the Washington Post (Maryland) and the Quinnipiac University Polling Institute (New York and Pennsylvania).[12]

TNS Intersearch in Horsham, PA conducted the interviews for the Maryland survey. The Quinnipiac University Polling Institute in Mt. Carmel, CT conducted the interviews for the New York and Pennsylvania surveys. The interviewers were blind to whether or not potential respondents had received an advance letter.[13] Interviewers were

---

records for attempted dialings were selected randomly from among the entire survey sample of 10,000 voters. The Maryland results have not been adjusted to reflect the advantages of contacting those who were sent a letter that was seen in New York and Pennsylvania. In the future, the practice of creating an evenly divided pool of letter and non-letter sub-samples is preferred to the method of balancing completed surveys used in Maryland.

[10] The letters for New York and Pennsylvania were printed by Yale University's printing office and mailed using a Connecticut postal permit. The letters for Maryland were printed by a commercial vendor and mailed using a Florida postal permit.

[11] Approximately 5% of the first class letters were returned as un-deliverable, usually because a forwarding address had expired. In Maryland, 149 of 3,000 letters were returned. In Pennsylvania, 144 of 3,000 letters were returned. In New York, only 6 letters were returned but the post office does not normally provide return service for bulk mail, so this is not a comparable indicator.

[12] The New York survey instrument also included questions on the Attorney General and Comptroller races in the 2002 election. These down-ballot races will not be discussed in this article due to the difficulty of comparing high and low information races.

[13] Supervisors at the Quinnipiac calling facility reported that some respondents volunteered this information to the interviewers at the beginning of the interview. In addition, interviewers in Maryland became aware that a respondent had been sent a letter on the final question of the RBS survey when they were prompted ask if the respondent received a letter.

aware of the parallel RDD and RBS surveys being conducted and were rotated between the two surveys on a night-to-night basis. For the New York and Pennsylvania surveys, 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. In all three surveys, the names of the major party candidates were rotated to avoid ordering effects.[14]

Calculating Participation and Effects of the Letters

The tables in this article do not precisely conform to the American Association for Public Opinion Research [AAPOR] definitions for contact, cooperation, or refusal rates due to the inclusion of all attempted calls in the denominator. Many 'bad' numbers - wrong numbers, disconnected numbers, fax/modems lines, etc - are excluded from the denominator in the standard AAPOR formulas used for RDD surveys (American Association of Public Opinion Research 2000). Since the process for gathering data for registration based sampling assigns 'bad' numbers to individuals about whom significant amounts of information is known, it seems inappropriate to exclude them as 'unknown' in case the association of 'bad' numbers is in some way systematic.[15] In the analysis below, 'bad' numbers are added to the denominator of AAPOR's basic formulas.

[14] In New York, third party candidate Thomas Golisano was listed last for all respondents. The over-estimation of his vote share may be in part explained by response order effects (Chang and Krosnick 2001; Visser et al 2000)

[15] There is a similar possibility of introducing bias because 'bad' numbers are systematically related to some characteristic exists for RDD, despite the random generation of phone numbers for RDD (which seems to be the justification for excluding 'bad' numbers from RDD calculations).

Recalculating the percentages by the standard APPOR formulas does not change the substantive findings below. (See Appendix B for dispositions).

The analysis below consists of a series of bivariate logit analyses of the effect of the experimental treatment (being sent an advance letter) on a single demographic variable (e.g. gender, age, party registration). I have not provided a multivariate analysis because it would not provide additional analytic insights or statistical leverage. The series of bivariate analyses of the experiments provides unbiased and statistically efficient estimates of the effects of advance letters because of the random assignment of the experimental treatment (i.e. the random selection of who is sent a letter). If properly specified, a multivariate model would yield exactly the same estimates of the effects of the advance letters as the bivariate analyses.

Findings

Effects on Cooperation Rates

As expected from most previous studies, the advance letter had a positive effect on the cooperation rate. The first column of Table 1 presents the cooperation rate from the combined data set of all three states (expressed in percentage points).[16] These combined effects were calculated using fixed effects for the differing baseline cooperation rate in each state. The discussion below will focus on the combined data set because its larger size provides the most statistical efficiency (i.e. accuracy) in estimating the effect of the advance letter.

[*INSERT TABLE 1 ABOUT HERE*]

[16] The percentage point effects listed in Tables 1 and Appendix C are converted from bivariate logit regression analyses of the data. The logit coefficients are available from the author on request.

In the combined data set of the three states, the advance letter caused an overall increase of 3.5 percentage points in the cooperation rate.[17] Maryland and Pennsylvania each had a highly statistically significant increase of 4.8 percentage points in the cooperation rate among those who received an advance letter. In New York, the advance letters generated only a 1.3 percentage point effect that failed to reach statistical significance (see Appendix C Table 1).[18] Statistically, it is quite conceivable that the insignificant effect of the letter observed in New York in 2002 lies in the outer edges of the distribution of potential observations if the true effect of advance letters is close to the 3.5 percentage point increase found in the combined data set.

The combined data set reveals a few noteworthy items about cooperation rates in the demographic sub-groups. The effect of the advance letter on cooperation among men and women is approximately equal to the overall effect. However, the letter does appear to have a stronger effect on registered Republicans than registered Democrats or other voters (5.0 versus 3.0 percentage points). This difference in partisan registration suggests that the survey respondents who received the letter may express more support for Republican candidates than those survey respondents who did not receive a letter.

The advance letter has a greater effect on cooperation among older registered voters (55-64: 4.9 percentage points; 65+: 4.2 percentage points). The effect of the letter on the 45-54 age group drops off to less than half the effect among older registered

[17] Although most social science shies away from unequivocal discussion of causation, the random assignment of the letter treatment among the survey sample allows clear determination of causality in these experiments.

[18] One potential, although unlikely, explanation for the poor performance of the letters in increasing the cooperation rate in New York relative to the other two states is the use of bulk rate rather than first class mail. It is plausible that the bulk rate mail was either not delivered in a timely fashion or not delivered at all, but we have no evidence to prove or disprove this hypothesis (see Footnote 10). It seems unlikely that the bulk rate mail was read less often than first class mail because the only difference between the bulk rate

voters. The strong effect also shows up among 35-44 year old registered voters (3.6 percentage points) and 25-34 year old registered voters (4.7 percentage points). Advance letters appear to have no net effect (-0.7 percentage points, p-value 0.818) on cooperation among the youngest registered voters (18-25).

The effect of an advance letter appears to be roughly consistent across the voter registration categories.[19] The date of registration can be a good a proxy for length of residence in the community.

The effect across voting history categories presents an interesting pattern. The registered voters who have recently participated in the electoral process show the greatest increases in cooperation from the advance letter. Newly registered voters (6.1 percentage points), voters who cast a ballot in 2000 and 1998 (4.1 percentage points), and those who voted in 2000 but not 1998 (3.0 percentage points) demonstrated large positive effects from the letter. On the other hand, apathetic registered voters - those who were registered but did not vote in either of the previous general elections – and voters who had not participated in a general election in 4 years show no statistically significant effect on cooperation from the advance letter. Since recent participation in the electoral process is strongly indicative of the likelihood of future voting, the same dynamic appears to be at work promoting participation in surveys and voting.[20]

---

letters in New York and the first class letters in Pennsylvania was the amount listed in the imprint from the postage machine.

[19] Although there is some variation in the reported estimates, none of the estimates are significantly different from one another at the 95% level.

[20] In Maryland, there does not appear to be a metropolitan/rural pattern in the baseline cooperation rate (i.e. the no letter column). Furthermore, no metropolitan/rural pattern appears in the effect of the advance letter either. Higher cooperation rates do appear in less metropolitan areas does appear in New York and, to a lesser degree, in Pennsylvania. However, New York and Pennsylvania show no metropolitan/rural pattern in the effect of an advance letter across regions. (See Appendix C).

Breaking Down Cooperation: Contact and Refusal Rates

With the finding of a statistically and substantively significant 3.5 percentage point effect of an advance letter, I now turn to understanding the mechanisms underlying this finding. One important question is the degree to which the advance letter increases the contact rate, decreases the refusal rate, or some combination of the two. The second column of Table 1 shows the effect of the advance letter on contact rates and the third column of Table 1 shows the effect on refusal rates.

The second column of Table 1 shows that the effect of the advance letter on the contact rate (1.3 percentage points, p-value 0.157) in the combined data set cannot be distinguished from zero at conventional significance levels. Thus, the change in contact rate due to the advance letter is unlikely to contribute much to the above finding of increased cooperation. It is worth noting that the advance letter causes a highly statistically significant 3.5 percentage point increase in the overall contact rate in Maryland, while the effects in New York and Pennsylvania are much smaller and fail to reach significance (see Appendix C Table 2).[21]

In the demographic sub-groups, the letter has a statistically significant effect on the contact rate among male registered voters (2.3 percentage points), but no significant effect among female registered voters. The letter also appears to have a sizable negative effect among voters 18-24 years old (-3.4 percentage points), which raises the possibility that younger voters utilize the letter as a warning to avoid contact rather than a positive inducement to participate.[22]

[21] Making 8 attempts to contact respondents in Maryland rather than only 5 attempts in New York and Pennsylvania is one possible explanation for this difference in contact rates.

[22] The regional contact rates are a mixed bag across the states (Appendix C). Pennsylvania demonstrates no discernable regional pattern in baseline contact rates or the effect of the letter on contact rates.

The third column of Table 1 shows a highly statistically significant reduction in the overall refusal rate of 2.2 percentage points. Among the states, Pennsylvania leads the way with a highly statistically significant 4.6 percentage point reduction in the refusal rate caused by the advance letters. Maryland and New York have smaller effects that fail to reach statistical significance (see Appendix C Table 3).

In the combined data set, the letter causes a larger drop in refusals among female registered voters than among male registered voters (2.8 versus 1.3 percentage points). The reduction of refusals is strongest on the youngest (3.5 percentage points) and the oldest (5.2 percentage points) registered voters with the intermediate age groups showing lesser effects. The reduction in refusals among younger voters indicates that although they appear to avoid calls when they get a letter (the negative effect on the contact rate noted above), those who are contacted will nonetheless be more cooperative if they are sent a letter in advance.[23]

Given the difference between the challenge of contacting a potential respondent and the challenge of securing cooperation once the potential respondent is on the phone, it seems likely that the advance letter informing the potential survey respondent will be more successful reducing refusals than increasing the contact rate. Nonetheless, despite supporting evidence from the combined data set, the contradictory evidence from Maryland and Pennsylvania (and the lack of statistically significant evidence from New

---

Maryland's baseline contact rates are the reverse of conventional expectations with higher contact rates in metropolitan areas than in more rural regions. However, the effect of the letter on contact rates in Maryland is higher the rural areas than in the metropolitan areas. New York's baseline contact rates have the expected pattern as one goes from NYC to Upstate, but the pattern does not extend to the effect of the letter.

[23] Again the regional patterns across the states are somewhat mixed (Appendix C). Maryland demonstrates no recognizable pattern in the baseline rate or the effect of the letter on refusals. New York and Pennsylvania show lower refusals rates in their metropolitan areas and larger refusal rates in other regions.

York) leaves the question of whether the letter is doing more work to increase contact or reduce refusals open for resolution by future replication of experiments on the effects of advance letters.

Representativeness

Since there is variation in the effects of an advance letter among sub-groups, I now turn to whether this variation makes the letter sub-sample of survey respondents more representative of the actual voting electorate than the no letter sub-sample. Table 2 presents the demographic profile of the sample at the major stages of the survey process for the combined data set.[24] The percentages in the table are weighted averages of the three states. The first column ('Original') lists the demographic profile of the combined stratified RBS samples of 40,000 registered voters from each state. The 'Survey' column is the profile of the 10,000 registered voters selected for the survey sample in each state. Since only about two-thirds of registered voters in all three states had valid phone numbers, some differences in the demographic profile might be expected due to the demographic characteristics of households with listed versus unlisted phone numbers. Within the demographic categories available from the voter registration information, the only change that seems to merit attention is the 5 percentage point increase in regular voters (voted in 2000 and 1998). The letter and no letter sub-samples and the sub-sample of attempts to contact are identical to the survey sample within the bounds of random chance - as expected from their random selection.

[*INSERT TABLE 2 ABOUT HERE*]

There does not appear to be a regional pattern to the effect of the advance letter on refusal rates in any of the states.

[24] Tables with the demographic breakdowns for each state are in Appendix D Tables 1-3.

Contacting potential respondents is the next stage where changes in the demographic profile might occur.[25] Gender and party registration have only very small shifts towards females and towards Republicans going from attempts to successful contacts. Voter registration history also shows only small shifts. There is a large shift in the age distribution so that the sample of registered voters that is successfully contacted for the survey is somewhat older than the sample of registered voters that we attempted to contact. There is also a large jump in the share of regular voters (68% versus 61%). This increase appears to come at the expense of apathetic voters (registered but not voting in 1998 or 2000) and newly registered voters.

The final stage is cooperation among those who were contacted. Those who cooperated are separated into two sub-samples based on whether or not they were sent a letter. Since it was established above that the letter had caused an increase in the cooperation rate and that there are the differences in that effect among demographic groups, there should be notable demographic differences between the letter and no letter sub-samples. Indeed, the sub-sample that was sent a letter is more evenly balanced between male and female, is slightly more Republican, and has some minor differences in the distribution across age groups and past voting history. Recalling the effects of an advance letter on cooperation rates across the sub-groups presented above, these differences between the letter and no letter sub-samples are as expected.

I now turn attention from the evolution of the original sample to the motivating question of whether these final sub-samples of actual respondents were representative of

[25] The contact rate is not broken down by letter and non-letter samples since we saw above that there is no statistically significant difference between these sub-samples for the contact rate.

the actual turnout (final column of Table 2).[26] The letter sub-sample of registered voters who cooperated in completing a survey is closer to the actual turnout in the distribution of male and female. On partisan registration, both sub-samples slightly under-represent registered Democrats and over-represent Republicans relative to the actual turnout.

The differences across age sub-groups, voter registration history, and voting history sub-groups are minimal. Neither the letter sample nor the no letter sample is clearly closer to the actual turnout. The distributions of respondents with regard to age, voter registration history, and vote history in both sub-samples are noteworthy for how much they have shifted away from the original sample and how much they miss the distributions of the actual turnout. Regular voters (voted 2000 & 1998) make up an overly large share of the people who cooperate in completing a survey whether they were sent a letter or not. The age distribution of those who completed a survey has shifted conspicuously towards older voters relative to the actual turnout. Similarly, voters who have been registered for more than 20 years (registered before 1983) make up a disproportionately large share of survey participants.

Forecasting Accuracy

While the issues of cooperation rates and representativeness of the sample are interesting, these experiments set out to see if an advance letter could improve the forecasting accuracy of pre-election surveys. Pre-election surveys provide a rare opportunity in survey research to test the accuracy of survey results since the survey forecast of the election can be compared to the actual election outcome.

[26] With the demise of the Voter News Service in the 2002 election, the demographics of the actual turnout are a bit of a challenge to obtain. For region and party, the states have released a breakdown of the actual electorate. The best available data on actual turnout by age, gender, and vote history is drawn from the

In Tables 3 to 5, I examine whether survey results based on respondents who were sent a letter produce a better forecast of election outcomes than survey results among those who were not sent a letter. With observations in only three states, there is far too little data to draw conclusions with any certainty. Nonetheless, the data from these experiments is strongly suggestive for future survey practices – at least until further replications confirm, modify or reject the results in these three experiments.

Respondents who were "undecided" or refused to answer the candidate preference question have not been allocated in Tables 3 to 5. Any procedure for allocating undecided voters has an effect on forecast accuracy (Chang and Krosnick 2001; Visser et al. 2000), so the data have been left unaltered so that the effect of the advance letters is clear.[27] One important effect of advance letters is to reduce the number of "undecided" responses in the first place.

Table 3 shows the percentage of respondents who expressed a candidate preference in the gubernatorial race in each state. New York had the largest increase in the percentage of respondents expressing a candidate preference due to being sent a letter, 5.5 percentage points from 85.7% to 91.2%. Pennsylvania saw a 3.1 percentage point increase from 87.1% to 90.2%. Maryland had only a 0.3 percentage point increase from 92.4% to 92.7%. It is not surprising that as the baseline rate of respondents expressing a candidate preference approaches 100%, the effect of advance letters decreases. While

---

post-election data gathered from county elections officials by the commercial voter file firms Voter Contact Services (Maryland) and Labels & Lists (Pennsylvania and New York).

[27] The candidate preferences reported here include respondents who initially said they were undecided but expressed a preference when asked the follow-up question about which candidate they were leaning towards. These survey results are unscreened and unweighted for vote likelihood survey responses or other factors. Weighting or screening the data would make assessments of the effect of an advance letter on the predictive accuracy hard to separate from the effects of weights or a screen. This presentation of the data is also supported by Visser et al (2000) and Chang and Krosnick (2001) who found that forecast accuracy was improved both by including leaners and by using unweighted results.

advance letters do not completely resolve the "undecideds" problem, the reduction of "undecided" responses should improve the accuracy of the forecasts.

[*INSERT TABLE 3 ABOUT HERE*]

Table 4 shows the forecast errors for candidate vote share in the gubernatorial race in each of the three states. The errors were calculated as the forecast vote share minus the actual vote share. The final three rows in this table summarize the results from the three states. Contrary to expectations, the sub-sample of respondents that was not sent a letter has an average absolute forecast error for candidate vote share (4.4 percentage points) that is somewhat smaller than the error for the sample which was sent a letter (5.0 percentage points). This suggests that survey researchers might be better off forecasting election results without advance letters.[28]

[*INSERT TABLE 4 ABOUT HERE*]

However, there is an important partisan difference in the forecast errors of candidate vote share. The letter sub-sample has a smaller average absolute forecast error for Republican candidates (4.1 versus 5.1 percentage points), while the no letter sub-sample has a smaller average absolute forecast error for Democratic candidates (4.3 versus 5.7 percentage points). When the two sub-samples are pooled the opposing partisan skews of the errors balance one another: 4.6 versus 4.9 percentage point errors for the Democratic and Republican candidates respectively.

Table 5 presents an alternative way of looking at the accuracy of pre-election surveys. Pre-election surveys have difficulty predicting actual vote share because survey respondents have an option to be 'undecided' that is not available on the ballot.

[28] These forecast errors in candidate vote share are comparable to previous studies covering a larger number of races (Crespi 1988; Gelman and King 1993).

'Undecided' responses are discouraged by survey questions but are nonetheless unavoidable. Therefore it may be fairer to compare the margin between candidates predicted by pre-election surveys to the actual election margin.

The margins in Table 5 are calculated as support for the Democratic candidate minus support for the Republican candidate. The forecast errors are calculated as the forecast margin minus the actual margin. The average absolute size of the forecast error for the winner's margin for the no letter sub-sample is almost twice the average absolute size of the error for the sub-sample that was sent a letter (4.4 versus 2.5 percentage points). Measured by the accuracy of the margin, the evidence appears strong that survey researchers should send advance letters.

[*INSERT TABLE 5 ABOUT HERE*]

A careful look at Table 5 reveals a pattern that makes this improvement in the magnitude of the forecast error of the margin less easy to embrace. The letter sub-sample has a smaller forecast error but it consistently over-predicts the margin of the eventual winner. Conversely, the no letter sub-sample consistently under-predicts the actual winner's margin.[29]

Under-prediction of the actual winner's margin would likely be beneficial to the trailing candidate. The frontrunner would benefit less from a bandwagon effect among late deciding voters, would lose fundraising advantages due to a lesser apparent likelihood of victory, and would lose other advantages that are associated with leading in the polls. The trailing candidate would not only gain relatively from the frontrunner's losses, but also could increase their own fundraising due to the increased appearance of a

chance of victory, reduce the possibility of demoralization/low turnout among supporters, and get a boost in other areas where apparent competitiveness helps a campaign. This artificially increased competitiveness from the systematic under-prediction of the frontrunners margin by surveys that do not send a letter (which is all of the current RDD surveys conducted by the media and other public survey organizations) is troubling methodologically, but at least seems likely to promote debate and discourse which are valuable to a healthy representative democracy.

The consistent over-prediction of the actual winner's margin when an advance letter is sent seems more troubling to a healthy representative democracy. Advantaging the frontrunner and disadvantaging the trailing candidate may suppress debate, discourse, and accountability without compensatory gains.

The pooling of the two sub-samples again demonstrates excellent properties, because the over- and under-prediction in the two sub-samples largely balance each other out when pooled together.[30] The average absolute forecast error (0.8 percentage points) of the pooled sample is less than one-fifth the no letter sub-sample error and less than one-third the letter sub-sample error. The problems of over-predicting the winner's share remain, but are now much smaller and therefore proportionally much less troubling.

---

[29] The under-prediction of the winner's margin of victory using stratified RBS without sending an advance letters is opposite of what has been found in studies of public polls using RDD methods (e.g. Visser et al 2000).

[30] The pooled sample in each state was simply adding the two sub-samples together. In Maryland, the sub-samples were the same size because more attempts were intentionally made among the no letter sample to create this balance. In New York and Pennsylvania, there were more completed surveys in the letter sub-sample because of the higher cooperation rate due to the letter. Weighting Maryland to approximate the distribution of completed surveys between the two sub-samples in Pennsylvania and New York reduces the over-prediction.

Discussion

The overall 3.5 percentage point effect from advance letters on cooperation found in Maryland, New York, and Pennsylvania are smaller than the 5 to 13.5 percentage point effects found in previous studies with a positive finding (Dillman, Gallegos, and Frey 1976; Goldstein and Jennings 2002; Traugott, Groves, and Lepkowski 1987). Our study differs from these past studies on a number of dimensions that are impossible to sort out with the available data: 1) geography – the surveys were in three neighboring Mid-Atlantic states; 2) general election versus primary election[31]; 3) election versus other purposes. Replication in these states and in other locations, in both primary and general elections, and for pre-election and other purposes is necessary to get a more thorough understanding of the effects of advance letters across a variety of settings.

The possibility of extending these results to purely commercial surveys remains an open question. The Washington Post is a widely recognized public information source and simply the use of its name indicates the purpose of the survey to most residents of Maryland. The partnership with Yale University on this particular survey added the imprimatur of academic purpose. In New York and Pennsylvania, the name of the Quinnipiac University Polling Institute indicates an academic purpose in its name and the letter stated the survey's purpose of providing information about the election to the public. The academic legitimacy and public information purpose conveyed by the

[31] In the 2002 New York Democratic primary, we had a registration based sampling survey in the field and letters had been sent to gather data on the effects of advance letters in primary elections when Andrew Cuomo dropped out of the race in final week. Further, Goldstein and Jennings (2002) also note that the presence of Arizona's US Senator John McCain in the 2000 primary may make the circumstances of their experiment exceptional. They find that the only question on which the letter appears to have affected the distribution of responses was the horse-race question in the GOP primary between McCain and Bush.

institutions involved in these surveys may make the effect of an advance letter impossible to replicate for a commercial survey without such characteristics.

The cost of reducing the magnitude of the error in the survey's pre-election forecast of the margin between the candidates by one percentage point was about $525 in New York and Pennsylvania (using Yale's in-house printing and mailing facility) and about $735 in Maryland (using a commercial vendor). The best-case scenario for the future use of letters would include using an in-house printing and mail facility with enough lead time to use bulk rate postage and sending 3,000 letters for surveys of the size of our experiments – making the estimated cost per percentage point improvement in the forecast accuracy of the margin of victory about $350. The worst-case scenario would include using first class mail because of short lead-time and using a commercial vendor - leading to an estimated cost per percentage point improvement of around $700.[32] Comparing these costs to the costs entailed in other ways of reducing non-response (e.g. paid incentives, intensive refusal conversion efforts, large increases in sample size) the cost of sending an advance letter seems to be a good value.

Conclusion

The use of large samples drawn from voter registration lists that provide demographic information about potential respondents has expanded our understanding of the effects of advance letters on pre-election surveys. As previous experiments have found, sending an advance letter increases the overall cooperation rate for the survey. This finding suggests that an advance letter could improve the representativeness of pre-election surveys. However, the demographic profiles of the letter and no letter samples

[32] It should be noted that most of the cost difference between the best and worst case scenarios is due to first class versus bulk rate postage, not the use of a commercial firm to process the mailing.

of survey respondents clearly support this idea only for gender. The distribution of partisan registration is also somewhat supportive of sending an advance letter. The voter history, voter registration, and age group distributions in the letter and no letter sub-samples of respondents are nearly identical and not well representative of the actual turnout. So while an advance letter has distinct effects on overall cooperation, they are not decisively in the direction of a more representative sample.[33]

Since these experiments used pre-election surveys, the analysis does not have to stop with this inconclusive picture of the effects of advance letters on representativeness. Comparing the survey forecasts to actual election outcomes, advance letters have clear effects on the accuracy of survey forecasts. Surveys conducted without sending an advance letter – as is the current practice for RDD surveys – appear to consistently under-predict the winner's margin. When potential respondents are sent an advance letter, surveys consistently over-predict the winner's margin, which is normatively problematic. Furthermore there is a partisan skew in accurately predicting candidate vote share: surveys which do not send an advance letter do a better job of forecasting Democratic candidates' vote share, while sending an advance letter reverses this pattern to more accurately forecast Republican candidates' vote share.

The increased cooperation rate does increase the overall response rate, but this does not serve the end for which it is intended: accurate forecasts from pre-election surveys. The best practice for future pre-election surveys appears to be found in the serendipitous construction of this experiment: send advance letters to a sub-sample of half the potential respondents and send no advance letters to the other half. Pooling these

[33] Ironically, using the data from these three experiments and others Green and Gerber (2003) find that a more representative sample reduces forecasting accuracy.

two sub-samples balances the opposing partisan biases in accurately forecasting candidate vote share. By splitting the sample between letter and no letter, researchers can use the offsetting biases of over and under prediction in the respective sub-samples to bracket the best forecast of the margin between the candidates.

BIBLIOGRAPHY

American Association of Public Opinion Research. 2000. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Lexana, Kansas: AAPOR.

Brehm, John. 1993. *The phantom respondents : opinion surveys and political representation*, *Michigan studies in political analysis*. Ann Arbor: University of Michigan Press.

Brunner, G. Allen, and Stephen J. Carroll. 1969. The Effect of Prior Notification on the Refusal Rate in Fixed Address Surveys. *Journal of Advertising Research* 9 (2):42-44.

Camburn, Donald, Paul Lavrakas, Michael Battaglia, James Massey, and Robert Wright. 1995. Using Advance Letters in Random-Digit Dialing and Telephone Surveys. Paper read at American Association of Public Opinion Research Annual Meeting; as cited in Goldstein and Jennings 2002.

Chang, LinChiat, and Jon A. Krosnick. 2001. Improving Election Forecasting. Center for Survey Research, The Ohio State University. Unpublished Manuscript.

Church, Allan H. 1993. Incentives in Mail Surveys. *Public Opinion Quarterly* 57 (1):62-79.

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

Dillman, D. A., J. G. Gallegos, and J. H. Frey. 1976. Reducing Refusal Rates for Telephone Interviews. *Public Opinion Quarterly* 40 (1):66-78.

Dillman, Don A. 2000. *Mail and internet surveys : the tailored design method*. 2nd ed. New York: Wiley.

Gelman, Andrew, and Gary King. 1993. Why are American Presidential Campaigns so Variable when Votes are so Predictable? *British Journal of Political Science* 23:409-430.

Goldstein, K. M., and M. K. Jennings. 2002. The effect of advance letters on cooperation in a list sample telephone survey. *Public Opinion Quarterly* 66 (4):608-617.

Groves, R. M., R. B. Cialdini, and M. P. Couper. 1992. Understanding The Decision to Participate in A Survey. *Public Opinion Quarterly* 56 (4):475-495.

Groves, R. M., E. Singer, and A. Corning. 2000. Leverage-saliency theory of survey participation - Description and an illustration. *Public Opinion Quarterly* 64 (3):299-308.

Groves, Robert M., and Mick Couper. 1998. *Nonresponse in household interview surveys*, *Wiley series in probability and statistics. Survey methodology section*. New York: Wiley.

Keeter, S., C. Miller, A. Kohut, R. M. Groves, and S. Presser. 2000. Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly* 64 (2):125-148.

Kish, Leslie. 1995. *Survey Sampling*. New York: John Wiley & Sons, Inc.

Link, M. W., and R. W. Oldendick. 1999. Call screening - Is it really a problem for survey research? *Public Opinion Quarterly* 63 (4):577-589.

Lynn, P. 2001. The impact of incentives on response rates to personal interview surveys: Role and perceptions of interviewers. *International Journal of Public Opinion Research* 13 (3):326-336.

Singer, E., J. Vam Hoewyk, N. Gebler, T. Raghunathan, and K. McGonagle. 1999. The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys. *Journal of Offical Statistics* 15 (2):217-230.

Singer, E., J. Van Hoewyk, and M. P. Maher. 2000. Experiments with incentives in telephone surveys. *Public Opinion Quarterly* 64 (2):171-188.

Sykes, Wendy, and Gerald Hoinville. 1985. *Telephone Interviewing on a Survey of Social Attitudes*. London: Social and Community Planning Research; as cited in Goldstein and Jennings (2002) and Traugott, Groves and Lepkowski (1987).

Thompson, Steven K. 1992. *Sampling*. 2nd ed. New York: John Wiley & Sons, Inc.

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.

Tuckel, P., and H. O'Neill. 2002. The vanishing respondent in telephone surveys. *Journal of Advertising Research* 42 (5):26-48.

Visser, Penny S., Jon A. Krosnick, Jesse Marquette, and Michael Curtin. 2000. Improving Election Forecasting: Allocation of Undecided Respondents, Indentification 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.

Voogt, R. J. J., and H. Van Kempen. 2002. Nonresponse bias and stimulus effects in the Dutch National Election Study. *Quality & Quantity* 36 (4):325-345.

Table 1
Effect of an Introductory Letter on Cooperation, Contact, & Refusal
Combined Dataset (Maryland, New York & Pennsylvania)

	Cooperation		Contact		Refusal	
	Effect (% pts)	p-value	Effect (% pts)	p-value	Effect (% pts)	p-value
Overall:						
Overall	3.5	0.000	1.3	0.157	-2.2	0.004
Gender:						
Male	3.5	0.000	2.3	0.040	-1.3	0.183
Female	3.4	0.000	0.5	0.624	-2.8	0.003
Party Registration:						
Democrat	3.0	0.002	1.2	0.345	-2.1	0.063
Republican	5.0	0.000	2.2	0.142	-2.2	0.082
Other Parties	1.3	0.487	-0.3	0.904	-2.7	0.173
Age:						
18-24	-0.7	0.818	-3.4	0.373	-3.5	0.292
25-34	4.7	0.031	3.4	0.214	-0.9	0.701
35-44	3.6	0.028	2.8	0.172	-0.2	0.889
45-54	1.5	0.314	1.7	0.385	-1.0	0.537
55-64	4.9	0.004	2.6	0.220	-1.6	0.379
65+	4.2	0.002	-1.0	0.545	-5.2	0.000
Voter Registration History:						
Registered 2001-2002	2.3	0.483	4.5	0.286	4.0	0.279
Registered 1999-2000	3.8	0.055	1.1	0.670	-4.0	0.066
Registered 1997-1998	3.3	0.251	-1.8	0.607	-4.5	0.152
Registered 1995-1996	1.0	0.619	-1.0	0.708	-2.9	0.198
Registered 1993-1994	3.9	0.207	7.4	0.062	3.1	0.372
Registered 1983-1992	4.1	0.002	4.0	0.016	-0.1	0.931
Registered Before 1983	3.4	0.006	-1.9	0.232	-5.0	0.000
General Election Voting History:						
2000 & 1998	4.1	0.000	0.5	0.665	-3.6	0.000
1998 (But Not 2000)	-0.8	0.815	-4.8	0.246	-4.6	0.205
2000 (But Not 1998)	3.0	0.049	3.5	0.074	-0.1	0.946
Registered but did not vote in 2000 or 1998	-0.4	0.870	-1.3	0.709	-0.5	0.871
Registered since 2000	6.1	0.051	6.2	0.121	1.9	0.592
Sample Size	5788		12381		5788	

Calculation of effect used fixed effects to control for differences in baseline cooperation, contact, and refusal rates in the three states.

Table 2
Demographic Profile of Sample at Various Stages - All States Combined

	Original	Survey	Letter	Attempt	Contact	Cooperate - No Letter	Cooperate - Letter	Actual Turnout[1]
Gender:								
Unknown	1%	1%	1%	1%	1%	0%	0%	0%
Female	54%	54%	53%	54%	55%	58%	52%	54%
Male	45%	46%	47%	45%	44%	42%	47%	46%
Party Registration:								
Democrat	51%	49%	49%	48%	47%	48%	49%	50%
Republican	35%	37%	36%	37%	40%	39%	40%	37%
Other Parties	15%	14%	16%	15%	13%	12%	11%	13%
Age:								
18-24	6%	5%	5%	6%	3%	3%	2%	5%
25-34	13%	10%	11%	10%	7%	6%	7%	11%
35-44	20%	18%	19%	19%	17%	17%	18%	19%
45-54	22%	22%	21%	21%	21%	26%	22%	22%
55-64	16%	18%	18%	17%	18%	19%	21%	18%
65+	24%	27%	27%	27%	34%	30%	30%	25%
Voter Registration History:								
Registered 2001-02	7%	5%	4%	5%	3%	3%	3%	7%
Registered 1999-2000	15%	13%	12%	13%	11%	11%	11%	9%
Registered 1997-1998	7%	6%	6%	6%	6%	5%	5%	7%
Registered 1995-1996	12%	12%	12%	12%	11%	10%	9%	11%
Registered 1993-1994	5%	5%	5%	5%	5%	4%	5%	5%
Registered 1983-1992	27%	29%	30%	29%	29%	30%	31%	28%
Registered Before 1983	27%	32%	30%	31%	36%	37%	37%	33%
General Election Voting History[2]:								
2000 & 1998	56%	61%	60%	61%	68%	74%	74%	55%
1998 (But Not 2000)	5%	5%	4%	4%	3%	2%	2%	2%
2000 (But Not 1998)	22%	21%	22%	21%	19%	16%	17%	26%
Registered but did not vote in 2000 or 1998	8%	7%	7%	7%	4%	3%	2%	8%
Registered since 2000	8%	5%	5%	5%	3%	3%	3%	9%
Sample Size	119,755	29,996	10,999	12,381	5,788	1,099	1,215	

1) Actual turnout is estimated from records available from voter file vendors and public records of each state's elections office.
2) Vote History does not total to 100% in Pennsylvania due to the exclusion of a few counties with inconsistent voter history, see footnote in Methodology section for further explanation.

Table 3
Expressed Candidate Preference

State-Office	Candidate	Party	Incumbent	No Letter	Letter
MD-Governor	Townsend	D		46.2%	42.9%
	Ehrlich	R		46.2%	49.7%
Total Expressed Preference				**92.4%**	**92.7%**
PA-Governor	Rendell	D		45.8%	50.7%
	Fisher	R		41.2%	39.5%
Total Expressed Preference				**87.1%**	**90.2%**
NY-Governor[2]	McCall	D		29.7%	25.7%
	Pataki	R	x	40.9%	43.9%
	Golisano	I		15.1%	21.6%
Total Expressed Preference				**85.7%**	**91.2%**

Sample Sizes

MD-Governor	424	414
PA-Governor	322	418
NY-Governor	353	383

Table 4
Forecasts of Candidate Vote Share

State-Office	Candidate	Party	Incumbent	Actual Vote[1]	No Letter	Letter	Pooled Results	*Forecast Error: No Letter*	*Forecast Error: Letter*	*Forecast Error: Pooled*
								(percentage points)		
MD-Governor	Townsend	D		47.7%	46.2%	42.9%	44.6%	*-1.5*	*-4.7*	*-3.1*
	Ehrlich	R		51.6%	46.2%	49.7%	48.0%	*-5.3*	*-1.8*	*-3.6*
PA-Governor	Rendell	D		53.4%	45.8%	50.7%	48.6%	*-7.6*	*-2.7*	*-4.8*
	Fisher	R		44.4%	41.2%	39.5%	40.3%	*-3.1*	*-4.8*	*-4.1*
NY-Governor[2]	McCall	D		33.5%	29.7%	25.7%	27.6%	*-3.8*	*-7.8*	*-5.9*
	Pataki	R	x	49.4%	40.9%	43.9%	42.4%	*-8.5*	*-5.5*	*-7.0*
	Golisano	I		14.3%	15.1%	21.6%	18.5%	*0.9*	*7.3*	*4.2*
Average Absolute Forecast Error for Candidate Vote Share								*4.4*	*5.0*	*4.7*
Average Absolute Forecast Error for Democratic Candidate Vote Share								*4.3*	*5.1*	*4.6*
Average Absolute Forecast Error for Republican Candidate Vote Share								*5.7*	*4.1*	*4.9*

Sample Sizes	Actual Vote[1]	No Letter	Letter	Pooled Results
MD-Governor	1,704,560	424	414	838
PA-Governor	3,581,989	322	418	740
NY-Governor	4,579,078	353	383	736

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 - 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.

Table 5
Forecasts of Margin of Victory

State-Office	Actual Vote[1]	No Letter	Letter	Pooled Results	*Forecast Error: No Letter*	*Forecast Error: Letter*	*Forecast Error: Pooled*
					(percentage points)		
MD-Governor	-3.9%	0.0%	-6.8%	-3.4%	*3.9*	*-2.9*	*0.5*
PA-Governor	9.0%	4.6%	11.2%	8.3%	*-4.4*	*2.2*	*-0.7*
NY-Governor	-15.9%	-11.1%	-18.2%	-14.8%	*4.8*	*-2.3*	*1.1*
Average Absolute Forecast Error in Margin of Victory					*4.4*	*2.5*	*0.8*

mple Sizes	Actual Vote[1]	No Letter	Letter	Pooled Results
MD-Governor	1,704,560	424	414	838
PA-Governor	3,581,989	322	418	740
NY-Governor	4,579,078	353	383	736

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.

Appendix A

(*Note: The letters for New York and Pennsylvania were identical except for the state name in the text.)*

Quinnipiac University
Polling Institute

NAME
ADDRESS
CITY/STATE/ZIP

You have been selected at random from among registered voters in New York as a potential participant in a telephone survey regarding the upcoming election. You may be contacted by phone during the last two weeks in October to complete a brief survey. The survey will take less than 10 minutes. We will not be selling anything nor asking for anything other than your answers to a few questions.

All information you provide will be kept confidential. Your responses to the survey will be used only to tabulate the results of the survey. All personal information will be kept confidential and will not be released for any purpose. We maintain the highest professional standards of confidentiality.

Quinnipiac University conducts independent public opinion surveys in New York, Connecticut, New Jersey, Pennsylvania and nationwide. For more than 12 years, our polling has been widely known as accurate and thorough in providing an understanding of elections and public policy issues. News media organizations and academic institutions turn to us for the most accurate information about public opinion. We are conducting this survey of New York voters in conjunction with Yale University. Your cooperation in completing the survey will assist us in continuing to provide accurate and thorough information for the public.

If you would like to learn more about the Quinnipiac University Polling Institute, please visit our website at http://www.quinnipiac.edu/.

Thank you in advance for participating in our poll!

Sincerely,

Doug Schwartz
Director of the Quinnipiac University Poll

Yale University's
Institution for Social and Policy Studies
in partnership with
The Washington Post

NAME
ADDRESS
CITY/STATE/ZIP

You have been selected at random from among registered voters in Maryland as a potential participant in a telephone survey regarding the upcoming election for Governor. You may be contacted by phone during the next week to complete a brief survey. The survey will take less than 10 minutes. We will not be selling anything nor asking for anything other than your answers to a few questions.

All information you provide will be kept confidential. Your responses to the survey will be used only to tabulate the results of the survey. All personal information will be kept confidential and will not be released for any purpose. We maintain the highest professional standards of confidentiality.

The Washington Post conducts public opinion polls in the Washington, DC area and nationally to provide current information to readers and civic leaders about the opinions of the public. For more than 20 years, our polling has been widely known as accurate and thorough in providing an understanding of elections and public policy issues. Yale University's Institution for Social and Policy Studies is partnering with the Washington Post as part of its mission to conduct research into important public policy arenas. Your cooperation in completing the survey will assist us in continuing to provide accurate and thorough information for the public.

Thank you in advance for participating in our survey!

Sincerely,

Donald Green
Director
Institution for Social and Policy Studies
Yale University

Richard Morin
Polling Director
The Washington Post

Appendix B
Survey Dispositions

	Maryland			New York			Pennsylvania			Combined		
	No Letter	Letter	Total	No Letter	Letter	Total	No Letter	Letter	Total	No Letter	Letter	Total
Complete Interviews	424	414	838	353	383	736	322	418	740	1099	1215	2314
	20.2%	**25.0%**	**22.3%**	**15.2%**	**16.5%**	**15.9%**	**16.2%**	**20.9%**	**18.5%**	**17.2%**	**20.3%**	**18.7%**
Partial Interviews	25	21	46							25	21	46
	1.2%	**1.3%**	**1.2%**							**0.4%**	**0.4%**	**0.4%**
Refusal or Break off	281	214	495	697	659	1356	706	616	1322	1684	1489	3173
	13.4%	**12.9%**	**13.2%**	**30.1%**	**28.4%**	**29.2%**	**35.4%**	**30.8%**	**33.1%**	**26.3%**	**24.9%**	**25.6%**
Ans. Machine/ Respondent Not Available	365	300	665	692	673	1365	546	541	1087	1603	1514	3117
	17.4%	**18.1%**	**17.7%**	**29.8%**	**29.0%**	**29.4%**	**27.4%**	**27.1%**	**27.2%**	**25.0%**	**25.3%**	**25.2%**
No Answer/ Busy	583	407	990	203	217	420	176	180	356	962	804	1766
	27.8%	**24.6%**	**26.4%**	**8.8%**	**9.4%**	**9.1%**	**8.8%**	**9.0%**	**8.9%**	**15.0%**	**13.5%**	**14.3%**
Other	29	30	59	37	30	67	17	16	33	83	76	159
	1.4%	**1.8%**	**1.6%**	**1.6%**	**1.3%**	**1.4%**	**0.9%**	**0.8%**	**0.8%**	**1.3%**	**1.3%**	**1.3%**
Bad Number	290	206	496	324	333	657	202	196	398	816	735	1551
	13.8%	**12.4%**	**13.2%**	**14.0%**	**14.4%**	**14.2%**	**10.1%**	**9.8%**	**10.0%**	**12.7%**	**12.3%**	**12.5%**
Unresolved Callbacks	99	64	163	13	23	36	24	32	56	136	119	255
	4.7%	**3.9%**	**4.3%**	**0.6%**	**1.0%**	**0.8%**	**1.2%**	**1.6%**	**1.4%**	**2.1%**	**2.0%**	**2.1%**
Sample Size	2096	1656	3752	2319	2318	4637	1993	1999	3992	6408	5973	12381

Appendix C Table 1

Effect of an Introductory Letter on Cooperation Rates (by state)

	Maryland			New York			Pennsylvania		
	No Letter	Letter	Effect (% pts)	No Letter	Letter	Effect (% pts)	No Letter	Letter	Effect (% pts)
Overall:	20.2%	25.0%	4.8***	15.2%	16.5%	1.3	16.2%	20.9%	4.8***
Gender:									
Male	17.9%	26.2%	8.3***	14.2%	18.5%	4.2***	15.7%	19.3%	3.5**
Female	22.1%	24.0%	1.9	16.0%	15.0%	-1.0	16.8%	22.6%	5.9***
Party Registration:									
Democrat	21.5%	25.8%	4.3**	15.7%	14.3%	-1.4	14.6%	21.6%	7.0***
Republican	21.3%	26.2%	5.0**	15.3%	20.8%	5.6***	16.9%	21.4%	4.5**
Other Parties	13.2%	18.6%	5.4	14.2%	15.0%	0.8	21.4%	12.9%	-8.4*
Age:									
18-24	10.9%	11.2%	0.3	8.0%	6.6%	-1.472	5.1%	3.8%	-1.3
25-34	9.5%	16.6%	7.1	9.0%	10.2%	1.2034	10.5%	17.8%	7.3*
35-44	21.2%	21.2%	0.0	12.6%	17.6%	4.9**	15.6%	20.8%	5.2*
45-54	21.6%	24.8%	3.2	19.4%	18.4%	-0.954	19.7%	21.9%	2.2
55-64	19.4%	31.5%	12.1***	20.4%	16.8%	-3.609	17.2%	24.5%	7.3**
65+	26.1%	30.1%	4.0	15.1%	19.5%	4.4**	17.0%	21.5%	4.5**
Voter Registration History:									
Registered 2001-02	13.3%	13.3%	0.0	10.8%	8.8%	-1.9	7.4%	16.7%	9.3
Registered 1999-2000	16.2%	17.7%	1.5	13.2%	16.4%	3.2	14.8%	20.1%	5.3*
Registered 1997-1998	15.1%	20.7%	5.6	15.4%	12.7%	-2.6	10.8%	19.4%	8.6*
Registered 1995-1996	15.1%	20.4%	5.3	11.4%	8.5%	-2.9	16.0%	18.4%	2.4
Registered 1993-1994	14.6%	15.8%	1.2	12.7%	16.6%	3.8	18.9%	27.1%	8.2
Registered 1983-1992	20.8%	29.2%	8.4***	16.6%	17.1%	0.5	15.6%	21.1%	5.5**
Registered Before 1983	25.5%	28.8%	3.2	17.3%	21.7%	4.4**	19.5%	22.4%	3.0
General Election Voting History:									
2000 & 1998	24.0%	29.8%	5.8*	20.0%	21.2%	1.2	19.0%	24.3%	5.3***
1998 (But Not 2000)	8.1%	11.3%	3.2	9.9%	4.8%	-5.0*	9.8%	8.8%	-0.9
2000 (But Not 1998)	18.4%	21.0%	2.6	11.5%	14.2%	2.7	12.6%	17.0%	4.4
Registered but did not vote in 2000 or 1998	9.8%	11.3%	1.5	4.8%	4.0%	-0.9	9.3%	7.6%	-1.7
Registered since 2000	9.5%	15.6%	6.1	10.5%	11.9%	1.4	8.0%	18.0%	10.0*
Region:									
Metro DC	21.5%	25.4%	3.8*						
Baltimore Metro	19.5%	25.4%	5.9*						
Eastern Shore	19.3%	25.6%	6.3						
Southern MD	16.4%	19.4%	3						
Western MD	23.0%	35.2%	12.3*						
New York City				9.9%	11.4%	1.5			
Suburbs				14.6%	15.5%	0.9			
Upstate				18.8%	20.4%	1.6			
Allegheny (Pittsburgh)							14.5%	17.2%	2.7
Philadelphia							13.9%	11.4%	-2.4
Northeast							12.4%	21.7%	9.2**
Southeast							15.0%	21.2%	6.2**
Northwest							19.4%	22.2%	2.8
Northeast							15.9%	21.6%	5.7
Central							19.4%	25.6%	6.1**
Sample Size	829	713		1063	1065		1052	1066	

* Significant at $p < .10$ ** Significant at $p < .05$ *** Significant at $p < .01$

Appendix C Table 2
Effect of an Introductory Letter on Contact Rates (by state)

	Maryland			New York			Pennsylvania		
	No Letter	Letter	Effect (% pts)	No Letter	Letter	Effect (% pts)	No Letter	Letter	Effect (% pts)
Overall:	39.6%	43.1%	3.5***	45.8%	45.9%	0.1	52.8%	53.3%	0.5
Gender:									
Male	37.2%	42.9%	5.7**	42.6%	0.4684	4.2	50.5%	53.3%	2.8
Female	41.4%	43.2%	1.7	48.3%	45.3%	-3.0	54.7%	53.7%	-0.9
Party Registration:									
Democrat	39.6%	44.4%	4.8**	44.4%	41.9%	-2.4	52.7%	54.4%	1.7
Republican	44.2%	44.8%	0.7	48.2%	53.4%	5.2**	52.7%	53.5%	0.8
Other Parties	28.9%	33.5%	4.5	45.4%	43.8%	-1.6	82.5%	92.2%	9.7
Age:									
18-24	28.5%	20.4%	-8.1	29.9%	32.0%	2.0	28.6%	23.1%	-5.5
25-34	22.9%	27.8%	4.9	28.0%	28.5%	0.6	33.1%	40.2%	7.1
35-44	37.6%	37.5%	-0.1	40.6%	46.4%	5.8*	46.8%	48.6%	1.8
45-54	39.8%	43.3%	3.5	46.8%	47.0%	0.2	52.4%	53.6%	1.1
55-64	40.4%	49.2%	8.8**	48.8%	46.9%	-1.9	53.5%	55.2%	1.6
65+	50.7%	54.1%	3.4	58.9%	56.7%	-2.2	65.1%	62.1%	-3.0
Voter Registration History:									
Registered 2001-02	32.2%	27.8%	-4.4	28.0%	41.2%	13.2*	30.9%	41.0%	10.2
Registered 1999-2000	31.3%	33.0%	1.7	40.3%	43.8%	3.5	49.7%	47.6%	-2.1
Registered 1997-1998	31.4%	31.0%	-0.4	45.0%	40.1%	-4.8	47.1%	47.3%	0.2
Registered 1995-1996	34.7%	40.1%	5.5	41.2%	37.1%	-4.1	51.4%	50.0%	-1.4
Registered 1993-1994	30.5%	34.2%	3.7	40.6%	46.7%	6.1	47.3%	62.7%	15.4*
Registered 1983-1992	37.5%	48.2%	10.7***	44.1%	45.6%	1.4	50.7%	52.3%	1.5
Registered Before 1983	49.6%	47.9%	-1.7	56.1%	53.6%	-2.6	61.4%	60.0%	-1.4
General Election Voting History:									
2000 & 1998	44.8%	48.2%	3.4*	53.5%	51.4%	-2.1	59.3%	59.4%	0.1
1998 (But Not 2000)	23.0%	30.7%	7.7	46.5%	27.4%	-19.1**	37.6%	34.0%	-3.6
2000 (But Not 1998)	36.1%	42.0%	5.9	40.1%	44.2%	4.1	46.2%	46.4%	0.2
Registered but did not vote in 2000 or 1998	27.0%	22.6%	-4.4	30.3%	29.5%	-0.7	32.6%	34.2%	1.6
Registered since 2000	28.6%	31.3%	2.7	29.5%	41.7%	12.1*	31.8%	38.2%	6.4
Region:									
Metro DC	40.6%	41.9%	1.3						
Baltimore Metro	41.1%	44.1%	3.1						
Eastern Shore	40.9%	46.3%	5.4						
Southern MD	32.7%	39.2%	6.5						
Western MD	36.1%	53.5%	17.5**						
New York City				37.5%	34.8%	-2.7			
Suburbs				44.4%	47.9%	3.5			
Upstate				51.6%	52.0%	0.5			
Allegheny (Pittsburgh)							50.0%	53.3%	3.3
Philadelphia							46.2%	35.7%	-10.5**
Northeast							48.8%	53.9%	5.1
Southeast							51.0%	51.3%	0.3
Northwest							53.8%	54.1%	0.3
Northeast							52.3%	61.0%	8.7
Central							58.1%	56.7%	-1.4
Sample Size	2096	1656		2319	2318		1993	1999	

* Significant at $p < .10$ ** Significant at $p < .05$ *** Significant at $p < .01$

Appendix C Table 3
Effect of an Introductory Letter on Refusal Rates (by state)

	Maryland			New York			Pennsylvania		
	No Letter	Letter	Effect (% pts)	No Letter	Letter	Effect (% pts)	No Letter	Letter	Effect (% pts)
Overall:	13.4%	12.9%	-0.5	30.1%	28.5%	-1.6	35.4%	30.8%	-4.6***
Gender:									
Male	13.2%	11.7%	-1.5	27.7%	27.4%	-0.3	33.9%	32.3%	-1.6
Female	13.6%	14.0%	0.4	31.3%	28.9%	-2.4	36.3%	29.6%	-6.7***
Party Registration:									
Democrat	12.4%	12.9%	0.5	28.1%	26.7%	-1.4	36.9%	31.2%	-5.7***
Republican	15.2%	14.0%	-1.2	32.1%	31.6%	-0.5	34.6%	30.5%	-4.1*
Other Parties	13.2%	10.4%	-2.8	31.0%	27.7%	-3.3	31.6%	30.9%	-0.7
Age:									
18-24	11.7%	4.1%	-7.6**	21.9%	23.0%	1.1	22.4%	17.9%	-4.5
25-34	7.5%	6.7%	-0.8	17.9%	17.0%	-0.9	20.3%	20.7%	0.4
35-44	10.1%	11.6%	1.5	27.3%	28.2%	0.8	28.7%	25.7%	-3.1
45-54	13.3%	12.7%	-0.6	26.8%	26.5%	-0.3	31.5%	29.3%	-2.1
55-64	14.6%	13.5%	-1.1	28.1%	29.8%	1.7	35.4%	29.4%	-6.0
65+	18.2%	18.3%	0.1	43.3%	36.9%	-6.4**	47.6%	39.5%	-8.1***
Voter Registration History:									
Registered 2001-02	8.4%	11.1%	2.7	17.2%	29.4%	12.2*	22.2%	21.8%	-0.4
Registered 1999-2000	11.3%	9.1%	-2.2	26.7%	25.2%	-1.5	33.4%	26.2%	-7.3*
Registered 1997-1998	8.1%	5.2%	-3.0	28.4%	26.1%	-2.3	33.8%	25.6%	-8.2
Registered 1995-1996	12.1%	11.5%	-0.6	29.0%	26.9%	-2.2	34.4%	29.4%	-5.0
Registered 1993-1994	15.9%	15.8%	-0.1	27.3%	30.2%	2.9	27.0%	33.9%	6.9
Registered 1983-1992	12.1%	15.1%	3.0	27.1%	27.8%	0.6	34.3%	30.1%	-4.2
Registered Before 1983	17.2%	13.6%	-3.5*	38.2%	31.4%	-6.9***	40.8%	35.9%	-4.9*
General Election Voting History:									
2000 & 1998	15.0%	13.7%	-0.013	32.9%	29.2%	-3.7**	39.1%	33.3%	-5.8***
1998 (But Not 2000)	9.5%	14.6%	0.051	35.2%	22.6%	-12.6	27.8%	23.1%	-4.7
2000 (But Not 1998)	11.7%	13.3%	0.016	28.2%	28.8%	0.6	32.2%	28.8%	-3.4
Registered but did not vote in 2000 or 1998	12.3%	5.2%	-0.071	24.1%	25.1%	1.0	20.9%	25.3%	4.4
Registered since 2000	8.8%	11.4%	0.026	19.0%	27.3%	8.3	22.7%	19.1%	-3.6
Region:									
Metro DC	12.4%	12.3%	-0.1						
Baltimore Metro	15.2%	13.1%	-2.1						
Eastern Shore	14.8%	11.6%	-3.2						
Southern MD	13.9%	15.4%	1.5						
Western MD	9.8%	12.6%	2.8						
New York City				29.4%	32.2%	2.8			
Suburbs				32.4%	30.6%	-1.8			
Upstate									
Allegheny (Pittsburgh)							31.7%	34.5%	2.8
Philadelphia							31.8%	22.4%	-9.4*
Northeast							35.9%	30.4%	-5.5
Southeast							36.1%	30.2%	-5.9*
Northwest							33.9%	30.8%	-3.1
Northeast							36.4%	37.7%	1.3
Central							38.1%	30.2%	-7.9***
Sample Size	829	713		1063	1065		1052	1066	

* Significant at $p < .10$ ** Significant at $p < .05$ *** Significant at $p < .01$

Appendix D Table 1
Demographic Profile of Sample at Various Stages - Maryland

	Original	Survey	Letter	Attempt	Contact	Cooperate - No Letter	Cooperate - Letter	Actual Turnout
Gender:								
Unknown	0%	0%	0%	0%	0%	0%	0%	0%
Female	55%	55%	52%	55%	56%	61%	51%	54%
Male	45%	46%	48%	45%	44%	39%	49%	45%
Party Registration:								
Democrat	57%	55%	55%	54%	55%	56%	57%	57%
Republican	29%	32%	32%	32%	35%	34%	33%	30%
Other Parties	15%	13%	13%	14%	11%	9%	10%	13%
Age:								
18-24	7%	6%	5%	6%	4%	4%	3%	4%
25-34	13%	9%	9%	9%	6%	4%	6%	11%
35-44	21%	19%	19%	18%	17%	19%	16%	22%
45-54	23%	23%	23%	24%	24%	26%	24%	24%
55-64	17%	19%	19%	19%	20%	18%	24%	18%
65+	20%	24%	24%	23%	29%	29%	28%	20%
Voter Registration History:								
Registered 2001-02	11%	6%	6%	6%	5%	4%	3%	10%
Registered 1999-2000	16%	13%	13%	13%	10%	11%	9%	12%
Registered 1997-1998	4%	4%	4%	4%	3%	3%	3%	7%
Registered 1995-1996	10%	9%	10%	10%	9%	7%	8%	9%
Registered 1993-1994	4%	5%	5%	4%	3%	3%	3%	5%
Registered 1983-1992	26%	28%	28%	29%	29%	30%	33%	28%
Registered Before 1983	28%	35%	35%	35%	41%	42%	42%	29%
General Election Voting History:								
2000 & 1998	60%	67%	68%	68%	75%	78%	79%	63%
1998 (But Not 2000)	4%	4%	4%	4%	2%	1%	2%	2%
2000 (But Not 1998)	17%	17%	16%	16%	14%	14%	13%	21%
Registered but did not vote in 2000 or 1998	7%	6%	6%	6%	4%	3%	3%	3%
Registered since 2000	12%	7%	6%	7%	5%	3%	4%	11%
Maryland Region:								
Metro DC	47%	45%	43%	48%	48%	54%	43%	39%
Baltimore Metro	30%	31%	32%	30%	31%	27%	32%	34%
Eastern Shore	6%	6%	7%	6%	6%	4%	7%	8%
Southern MD	13%	13%	14%	14%	12%	11%	11%	15%
Western MD	4%	4%	4%	4%	4%	3%	6%	5%
Sample Size	39713	9996	2999	3752	1542	424	414	

Appendix D Table 2
Demographic Profile of Sample at Various Stages - New York

	Original	Survey	Letter	Attempt	Contact	Cooperate - No Letter	Cooperate - Letter	Actual Turnout
Gender:								
Unknown	3.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	
Female	53.5%	55.0%	54.3%	55.9%	57.0%	59.2%	50.4%	54.3%
Male	43.4%	44.9%	45.6%	44.0%	42.9%	40.8%	49.6%	45.7%
Party Registration:								
Democrat	45.6%	45.5%	46.0%	45.6%	42.9%	46.5%	39.9%	46.9%
Republican	29.4%	32.2%	31.8%	32.2%	35.5%	33.1%	39.4%	34.2%
Other Parties	25.0%	22.2%	22.2%	22.3%	21.6%	20.4%	20.6%	18.9%
Age:								
18-24	6%	6%	5%	6%	4%	3%	2%	3%
25-34	14%	12%	12%	13%	8%	7%	8%	8%
35-44	21%	19%	19%	19%	18%	16%	20%	17%
45-54	21%	21%	21%	20%	21%	27%	22%	23%
55-64	16%	17%	18%	17%	18%	21%	18%	20%
65+	23%	26%	25%	25%	32%	25%	29%	29%
Voter Registration History:								
Registered 2001-02	5%	4%	3%	4%	3%	3%	2%	5%
Registered 1999-2000	12%	10%	10%	10%	9%	9%	10%	8%
Registered 1997-1998	8%	8%	8%	7%	7%	7%	5%	7%
Registered 1995-1996	12%	12%	12%	12%	10%	9%	7%	10%
Registered 1993-1994	7%	7%	7%	7%	7%	6%	7%	6%
Registered 1983-1992	32%	33%	35%	34%	33%	35%	36%	33%
Registered Before 1983	23%	27%	26%	26%	32%	31%	33%	31%
General Election Voting History:								
2000 & 1998	48%	52%	53%	52%	60%	69%	67%	62%
1998 (But Not 2000)	3%	3%	3%	3%	2%	2%	1%	2%
2000 (But Not 1998)	32%	31%	30%	31%	29%	23%	27%	27%
Registered but did not vote in 2000 or 1998	11%	10%	10%	10%	6%	3%	2%	3%
Registered since 2000	6%	4%	4%	4%	3%	3%	3%	6%
New York Region:								
New York City	35%	30%	30%	29%	23%	19%	20%	30%
Suburbs	24%	23%	23%	24%	24%	22%	22%	23%
Upstate	42%	47%	47%	47%	53%	59%	57%	47%
Sample Size	40055	9998	5000	4637	2128	353	383	

Appendix D Table 3
Demographic Profile of Sample at Various Stages - Pennsylvania

	Original	Survey	Letter	Attempt	Contact	Cooperate - No Letter	Cooperate - Letter	Actual Turnout
Gender:								
Unknown	2.0%	1.6%	1.8%	1.8%	1.7%	0.9%	1.4%	0.1%
Female	51.5%	51.7%	50.1%	51.3%	52.4%	53.7%	55.0%	52.7%
Male	47%	47%	48%	47%	46%	45%	44%	48%
Party Registration:								
Democrat	48%	47%	47%	46%	46%	40%	48%	48%
Republican	45%	46%	46%	48%	48%	52%	47%	46%
Other Parties	7%	7%	7%	6%	6%	8%	4%	6%
Age:								
18-24	4%	4%	4%	5%	2%	2%	1%	3%
25-34	11%	9%	9%	9%	6%	6%	7%	8%
35-44	18%	17%	17%	18%	16%	16%	18%	17%
45-54	21%	21%	21%	20%	20%	24%	21%	23%
55-64	16%	17%	17%	16%	17%	17%	20%	19%
65+	29%	33%	33%	33%	39%	35%	33%	29%
Voter Registration History:								
Registered 2001-02	6%	4%	4%	4%	3%	2%	3%	6%
Registered 1999-2000	16%	15%	15%	15%	14%	13%	15%	8%
Registered 1997-1998	7%	7%	7%	7%	6%	5%	6%	7%
Registered 1995-1996	14%	14%	14%	14%	13%	14%	12%	13%
Registered 1993-1994	4%	3%	3%	3%	3%	4%	4%	3%
Registered 1983-1992	24%	24%	25%	24%	24%	23%	25%	23%
Registered Before 1983	29%	33%	33%	32%	37%	38%	35%	39%
General Election Voting History[1]:								
2000 & 1998	60%	64%	64%	64%	72%	75%	74%	61%
1998 (But Not 2000)	7%	7%	7%	7%	5%	4%	3%	1%
2000 (But Not 1998)	16%	15%	15%	14%	13%	11%	12%	28%
Registered but did not vote in 2000 or 1998	6%	4%	4%	4%	3%	2%	1%	3%
Registered since 2000	6%	4%	5%	4%	3%	2%	4%	7%
Pennsylvania Region:								
Allegheny (Pittsburgh)	13%	13%	13%	13%	13%	12%	11%	11%
Philadelphia	12%	9%	10%	10%	7%	7%	6%	11%
Northeast	12%	11%	11%	11%	11%	8%	11%	11%
Southeast	21%	22%	21%	21%	21%	20%	21%	22%
Northwest	9%	9%	9%	9%	9%	11%	10%	10%
Northeast	11%	11%	12%	11%	12%	11%	11%	10%
Central	23%	24%	24%	25%	27%	30%	30%	25%
Sample Size	39981	10000	3000	3992	2118	322	418	

1) Vote History does not total to 100% in Pennsylvania due to the exclusion of a few counties with inconsistent voter history, see Methodology section for further explanation.