

**Voter Lists as Sampling Frames for Telephone and Mail Samples:
Validating Two Studies by Focusing on Vote Choice and Turnout**

Nicoleta Corbu

College of Communication and Public Relations

National School of Political Studies and Public Administration

Bucharest, Romania

Qingmei Qing & Lee B. Becker

James M. Cox Jr. Center for International Mass Communication Training and Research

Grady College of Journalism and Mass Communication

University of Georgia

Abstract

Most concerns about sampling in survey-based research focus on the lack of accuracy and the high costs of the sampling methods generally used. The growth of cell-phone-only households and increasingly lower return rates in surveys in general have brought new challenges in recent years. This study examined some options for drawing inexpensive and valid samples that take advantage of new sources of information, such as telephone online directories. Address-based sampling using voter lists and mail contact has been shown in earlier research to be an effective alternative to random digit-dialing techniques. This research analyzes two Georgia surveys using address-based sampling techniques using voter registration lists. The two studies show that voter list-based sampling is a good, inexpensive method to obtain accurate samples at the county level. Both of the studies provided good estimates of elections outcome, not only in a telephone survey, but also in a mail survey. Additionally, voter lists offer a great advantage in providing voting history, as well as demographic characteristics for each voter, which can be a good base for testing the samples accuracy, and for refining the forecasting in order to obtain a more accurate outcome prediction.

Key words: sampling technique, voter registration list, turnout, election outcome

Introduction

Sampling accuracy has been a key issue in studies using public opinion survey. Most concerns were related to possible demographic or overreporting biases and the high costs of commonly used sampling methods. Cell-phones only households and low return rates brought new challenges, especially for large populations. Issues associated with small, community surveys have largely been ignored or considered less important than national ones. Local political office holders, small media organizations, and community organizations, however, rely on surveys based on what are often small community samples.

Random telephone number generation and random digit dialing are the most widely used method in telephone surveys. This paper focuses on sampling from voter lists, as an inexpensive alternative for both telephone and mail surveys, using the advantages of new resources, such as online telephone directories.

One of the main problems related to sampling in public opinion surveys is overreporting and how this can affect other variables, especially elections outcome. The findings in two local surveys show that sampling from voter lists provides valid, reliable samples and that overreporting vote does not affect accuracy of the predictions of elections outcome.

Registration-Based Sampling: An Alternative to Random Digit Dialing Technique

Accuracy has been one of the main concerns of researchers using survey methods. The Mitofsky-Waksberg method of random digit dialing (RDD) is the most widely used survey method (Brick et al, 1995). It ensures that all residential telephones have an equal chance of selection. On the other hand, people living in institutions may have no chance

to be selected (Mitofsky et al., 2005). The disadvantages of RDD sampling method are related to the fact that pollsters may discard or down-weight the interviews conducted with people whom they consider as unlikely voters (Green & Gerber, 2003).

In recent years, researchers who conducted pre-election surveys have experimented with samples drawn from voter registration lists. They are used because researchers can identify potential voters, learn past voting history, and acquire demographic information of voters (Mitofsky et al., 2005). These attributes can be used as a source of parameters for improving the estimates, and in the same time they can provide a good accuracy test for the samples drawn. In addition, they can be used as stratification variables, which can improve the design of the sample.

As recent studies mention (Green & Gerber, 2006; McDonald, 2007), relatively little has been written about the practical details of registration-based sampling (RBS). Although relatively rarely used in research and political surveys in general, voter registration list sampling is shown to have considerable advantages as compared to the random digit dialing procedure (Green & Gerber, 2006; McDonald, 2007): useful background information about the respondents, simpler interview protocols, higher response rates (particularly in low-salience elections), reduction of the costs associated with identifying likely voters, etc. There are also some possible drawbacks, such as incomplete coverage across and within states, lack of phone number information and unwillingness to disclose vote intentions when the survey is not anonymous.

Little is also known about the sample accuracy in different types of surveys using voter registration lists. If investigated at all, differences emphasize the advantages or disadvantages of different types of surveys and the possible biases related to different

aspects of the sampling procedures. Visser, Krosnick, Marquette and Curtin (1996) suggest that mail surveys not only may be viable alternatives to telephone surveys but may also be more effective under some conditions, based on their research findings, which shows a more accurate forecasting for mail surveys (using voter registration lists in Ohio) as compared to random digit dialing surveys. Green and Gerber (2003, 2006) found a more accurate prediction of midterm election outcomes for registration-based sampling (registration list for Maryland, Pennsylvania, New York and South Dakota) versus random digit dialing sampling. In addition, they found that RBS is more money-saving than RDD.

Public Opinion Surveys: Estimating Turnout and Predicting Election Outcome

The main focus of public opinion surveys during election times has always been forecasting an accurate election outcome. Additionally, turnout was important not only by itself, but also because of the challenges it raises. Research shows that, when discussing accuracy of public opinion surveys, overreporting vote is investigated in terms of possible influences on predicting outcome.

Validity of survey data has been a constant concern in public opinion research. The earliest large-scale data set that systematically investigated validity issues is the 1949 Denver Community Study. The data was consequently used to address specific validation issues for several decades (Presser, 1984). Overreporting thus started to be investigated in terms of causes and characteristic of people who declared they voted and did not.

The main causes of misreporting are memory and social desirability (Katosh & Traugott, 1981; Presser, & Traugott, 1992; Belli et al., 1999; Fullerton, Dixon & Borch, 2007), and, for local issues, exposure to local television news (Volgy & Schwarz, 1984).

Social desirability has an important impact on estimating turnout because it influences the way people report vote, especially for those who believe voting is a civic duty that should be filled in a healthy democracy. Therefore, one of the key questions related to overreporting vote is who misreports. In other words what are the characteristics of people who declared they voted but did not, as compared to true voters.

Research shows that the most salient characteristic of overreporters is high education (Silver, Anderson & Abramson, 1986; Granberg & Holmberg, 1991; Bernstein, Chadha & Montjoy, 2001; Cassel, 2003; Karp & Brockington, 2005; Fullerton, Dixon & Borch, 2007), which is explained by the belief that they feel a pressure to misreport because they usually think they are the kind of people who vote or should vote. Other characteristics are related to the degree of partisanship. The more partisan and the more religious people are more likely to overreport (Bernstein, Chadha & Montjoy, 2001), as well as people who accept the norm of voting and have an interest in the outcome of the election (Karp & Brockington, 2005). Middle-aged people are less likely to tell the truth about voting, while young people would rather acknowledge not have been voted (Granberg & Holmberg, 1991). Race has proven an important indicator rather for misreporting registration (Fullerton, Dixon & Borch, 2007).

The study of causes and characteristics of overreporters has led to a constant search for solutions and a further investigation of the impact of misreporting on other variables, with a focus on forecasting the outcome of elections. Different solutions have been proposed to decrease the overreporting bias: different wording, with options stating people have thought about voting but did not in the end (Belli et al., 1999), secret ballot

technique as a solution to avoid misreports in exit-polls (Bishop, Fisher, 1995), post-debates polls to estimate likely-voters (Traugott, 2005).

Overreporting has been an important issue not only by itself, but also in relation to other variables. A recent study (Cassel, 2003) provides some comfort to researchers by showing that the possible influence of overreporting on other variables that serve as predictors for the outcome is lower than previously estimated.

Many scholars have used the approach of identifying likely voters to make an accurate forecast of election results (Anderson & Silver, 1986; Bolstein, 1991; Petrocik, 1991; Monson, 1998; Siegelman, 1982; Siegelman et al., 1985; Silver, Anderson & Abramson, 1986; Traugott & Tucker, 1984; Voss, Gelman & King, 1995). Using voter registration lists seems to be the most effective technique to identify likely voters (Mann, 2003). The Columbus Dispatch poll in Ohio using mail surveys based on voter registration lists has proven to be more accurate in forecasting outcome as compared to RDD similar polls (Visser et al., 1996).

Two main possible bias sources in forecasting election outcome are nonresponse and undecided respondents before elections. Nonresponse may affect the representativeness of the sample (Mann, 2005), because research provides evidence that respondents and nonrespondents may have different interests and preferences in politics (Groves & Couper, 1998; Taylor, 1997; Voogt & Van Kempen, 2002). Allocating undecided respondents has proven a better method than just treating them as missing cases, and therefore researchers have developed methods to deal with this issue (Mann, 2003). Possible solutions are allocating respondents who report leaning towards a candidate to improve the accuracy of pre-election forecasts (Visser et al., 2000), using

party identification reported by respondents or candidate assessments or issue items from the same survey (Crespi, 1988), allocating evenly between the candidates (Mann, 2003), assuming that undecided respondents reflect a portion of the electorate that will cast their ballots randomly (Visser et al, 2000). Traugott and Tucker (1984) propose a relatively complicated technique to allocate undecided respondents. They claim that undecided respondents should be divided by partisan affiliation then allocated based on the vote intentions of equivalent partisans who expressed a vote intention.

Other methods to increase forecasting accuracy are based on research on the length of time between the poll and the election showing that the closer to the election, the more accurate the poll (Crespi, 1988), the number of days the poll is in the field (Lau, 1994), and screening likely voters (DeSart & Holbrook, 2003).

Research Questions and Hypotheses

This research is based on two public opinion surveys using sampling procedures based on voter registration lists. One is a telephone survey conducted before the presidential elections in November 2008, the other is a mail survey conducted immediately after a local tax initiative election in March 2009.

The focus of these research studies concerns two main problems. The first is related to voter registration lists as a reliable sampling technique in both telephone and mail surveys, and the second to the accuracy of election turnout and outcome.

Although using online telephone directories in addition to voter lists to create the sample for the telephone surveys eliminate some voters for whom telephone numbers cannot be found, we argue that address-based sampling from voter registration lists will provide accurate samples in both telephone and mail surveys, in terms of demographic

characteristics of the voters, and therefore the representativeness of the samples would not be affected. However we expect some bias in terms of an increased activism of people who respond in public opinion surveys, regardless of the type of survey.

The research questions and hypotheses associated with these issues are

RQ1: Does address-based sampling using voter registration lists associated with online telephone directories provide a sample accurately representing population parameters in a telephone survey?

RQ2: Does address-based sampling using voter registration lists provide a sample accurately representing population parameters in a mail survey?

RQ3: What kind of biases appear as a result of not being able to interview all the persons whose telephone numbers were found (nonparticipation biases)?

RQ4: What kind of biases result from nonparticipation in a mail survey?

H1: People who take a telephone public opinion survey are more active in terms of voting behavior than the general population.

H2: People who take a mail public opinion survey are more active in terms of voting behavior than the general population.

The issue of accuracy of turnout and election outcome is partially influenced by people's activism. In addition, literature shows that people tend to overreport voting in favor of a more socially desirable response. We predict similar behaviors in both telephone and mail surveys. Therefore we expect some bias in estimating election turnout, but we argue that election forecasting is not influenced by it.

H3: People overreport voting when taking a telephone public opinion surveys.

H4: People overreport voting when taking a mail public opinion surveys.

H5: Forecasting election outcome in a telephone public opinion survey is not influenced by self-reporting voting behavior.

H6: Forecasting election outcome in a mail public opinion survey is not influenced by self-reporting voting behavior.

Design Study 1

The first study focused on the presidential elections of 2008 in Oconee County, Georgia. The voter registration list was obtained on September 8, 2008, from the Secretary of State Office and contained records through that date. The list included 21,352 names. Registration closed on Oct. 6. On election day, November 4, 2008, the Board of Election in Oconee County reported 21,579 registered persons. The number of people who voted on November 4 but who were not registered on September 8 was 1,143.

A simple random sample of 840 people was drawn via SPSS from the list on October 1, 2008. The numbers were randomly ordered and divided into groups of 30 for assignment to 21 interviewers, who were graduate students in a research methods class in the Grady College of Journalism and Mass Communication at the University of Georgia. The first two authors were enrolled in the class. The students were instructed to work them from top to bottom until they finished 10 completed interviews each.

The interviewers first looked for telephone numbers available for the assigned names, using the names and addresses as search criteria, using www.whitepages.com, www.switchboard.com, www.yellowbook.com, www.dexknows.com, and simple www.google.com searches on the Internet.

The interviewers were given new blocks of numbers as needed as they worked through the interview process and attempted to complete interviews with the assigned voters. In the end, from the 800 names and addresses, the student interviewers were able to locate 553 telephone numbers in the online directories. For 192 names, it was impossible to find corresponding telephone numbers using the Internet resources. For 55 names, there was no attempt to find telephone numbers. In these cases, the numbers had been assigned but never used.

A total of 200 surveys were completed by telephone from October 15 to November 3. The interviewers were allowed to use alternative modes of gathering data, such as mailing questionnaires or in-person interviews, but few did so. In the end, only an additional 14 interviews were completed through non-telephone methods. For the purpose of this paper, only the 200 interviews completed by telephone will be considered. Out of the 553 sampled registered voters for whom telephones numbers were available, 200 were interviews completed by telephone, another 14 by other methods, 32 were continuously answered by answering machines. Four numbers were always busy, and in 43 cases the person to be interviewed was never at home. In 69 cases, nobody ever answered the telephone. An additional 50 numbers were out of service. One person did not speak English and could not be interviewed. In one case, the interviewer made an illegal substitution, interviewing the wife of the selected registered voter. In 32 cases, the person to be interviewed no longer lived in Oconee County and was not interviewed. A total of 107 persons refused to complete the survey and could not be converted by subsequent calls.

These data are summarized in Table 1, with the resultant AAPOR calculations of return rate shown. Return Rate 1 was 38.0%.

The voter registration lists obtained from the Secretary of State contain a number of characteristics of the voters that can be used to compare the samples drawn with the population. In addition, the Secretary of State produces a voter history file that contains a record of voter turnout for each voter for each election during a given year. This voter history file is available online for download and contains a voter identification number that also was part of the voter list purchased from the Secretary of State. This voter history file was downloaded and merged with the purchased list.

Measures Study 1

Characteristics from the voter registration and voting history files were examined and selected for analysis. Some overlap in these characteristics existed. All unique factors were selected. The first characteristic used was voter status. If a voter has not voted in two years and has not responded to a mailed challenge, the voter is labeled as inactive. If the voter does not go to the polls in two even-year elections, the voter is purged. The lists contain a full voter address, including Zip Code. Zip Code was recoded into a binary variable, reflecting delivery by the post office in the county seat of Watkinsville or by some other post office. Oconee County divides voters into 13 precincts. Precinct was recoded into a binary variable dividing the county into the southern, agricultural part of the county and the northern, more developed part of the county. The county has four small, incorporated cities, including the county seat of Watkinsville. These were collapsed to create a variable for incorporated vs. unincorporated areas of the county.

The election file also includes race and gender. Since 92 percent of the Oconee County voters classify themselves as white, race was also reduced to a binary variable of white and non-white.

Another characteristic was registration year, reduced to a binary distinction between those who registered before and after 2000. The original voting record contained a recording of the year of last voting, prior to the November 2008 election. The year people last voted was reduced to a binary variable, distinguishing between first time voters (at the address) versus those who had voted at the address before.

The voting history file downloaded for all of 2008 contained a record of voting or not voting for five elections. The first was the February 5, 2009, presidential primary, in which eight Democrats, including frontrunners Hillary Clinton and Barack Obama, and nine Republicans, including frontrunners Mike Huckabee, John McCain and Mitt Romney, appeared on the ballot. (Obama and Huckabee won their respective races.) The second was the July 15 primary for local and statewide offices. The third was the August 5 runoff election for the local and statewide offices. The fourth was the November general election. (McCain carried Georgia.) The fifth was a runoff election for a U.S. Senate seat. In addition, the record indicated whether the voter used an absentee ballot or provisional ballot, which included ballots cast as part of the early voting procedures allowed in the state. Georgia does not register voters by party and has open primaries, meaning that a voter decides on election day which ballot she or he wishes to cast. The record indicated whether the voter had asked for a Democratic or a Republican ballot in the February, July and August primaries.

Finally, the official vote record showed John McCain received 12,113 votes (70.8%) in Oconee County in November of 2008, Barack Obama received 4,824 votes (28.2%), and Libertarian Bob Barr received 177 votes (1.0%). The telephone survey contained a measure of vote intent (or actual vote, if the voter has already cast a ballot). The question measuring the likelihood to vote in the questionnaire on a 10-item Likert scale was recoded into a binary variable. People who chose the “definitely will vote” (or 10 on the scale) option or had already voted at the time of the interview were considered as likely to vote, all the others as not likely to vote.

The sampling procedures described above resulted in the creation of five samples that can be compared with the population on these characteristics. These samples are summarized below.

Sample 1: Sample Drawn (N=800). This is the random sample of voters assigned to interviewers.

Sample 2: Sample of Eligible Respondents (N= 703). This is the sample of voters actually used by interviewers. Excluded were numbers assigned but never used because the interviewer did not need them as well as those who were not registered at the address. In the latter case, this exclusion was confirmed by the interviewer who either reached the voter at a new address or reached someone else at the address. If the number was inoperative, was always busy, or only was answered by an answering machine and the interviewer could not confirm that the voter had moved, the individual remained in this sample.

Sample 3: Telephone Sample (N= 532). This is the sample for which telephone numbers could be located.

Sample 4: Sample Interviewed (N= 200). This is the sample of registered voters actually interviewed by telephone.

Sample 5: Telephone Sample Plus Confirmed Improperly Registered (N= 242). This is the sample of voters interviewed plus voters found to have moved and therefore to be ineligible to vote.

Sample 1 shows if any biases appeared in the originally, randomly selected sample. These biases would have to result from the error of sampling alone. A comparison of Sample 1 with the population would indicate if there were any error resulting from random sampling. Sample 2 differs from Sample 1 in two ways. The first is by eliminating randomly subjects not assigned to the interviewers. The second group eliminated were those who were improperly registered and this was confirmed by the interviewers. A comparison of Sample 2 with the population would indicate if any bias resulted from elimination of these two groups. Sample 3 is the sample of voters for which telephone numbers could be found. A comparison of this sample with the population will indicate the biases that can appear because of the design of the study itself, in other words, from using telephone directories to identify landline phone numbers. Sample 4 is the sample of registered voters actually interviewed. A comparison of it with the population will show the biases of nonparticipation. Sample 5 includes those who were interviewed as well as those who could have been interviewed but were not because they were improperly registered. A comparison of this sample with the population provides a second test of the biases of using a telephone sample based on numbers found in online directories.

Design Study 2

The second study was completed during the period following a vote for what is in Georgia is called a Special Purpose Local Option Sales Tax. This is a one cent on the dollar tax for designated capital projects that requires periodic approval by the voters. Such an election was held on March 17, 2009, in Oconee County.

For this study, a simple random sample of 500 was drawn via SPSS from a voter registration list obtained on February 3, 2009. At that time, 22,090 people were registered to vote. The registration closed on February 17, when a total of 22,113 people were officially registered to vote for the March 17 election. There were 1,457 voters who actually voted on March 17, according to the official returns. The voter history contained records for 1,438 voters, meaning by the time it was uploaded 19 persons who voted had been purged from the list, presumably because they moved by the time those records were created at the Secretary of State Office in May of 2009. Of the 1,438 voters, 83 had not been registered when the voter list was purchased on February 3, 2009.

Five hundred questionnaires were mailed on March 16, and a second mailing followed on April 3. Eighty-one valid completed questionnaires were returned after the first mailing, and another 43 after the second mailing and before the closing date of April 24. Another valid completed questionnaire was returned after that date but was not included among the completed questionnaires. An addition four questionnaires came back after the first mailing, but the persons who completed them removed their names from the instrument, making it impossible to identify them. The total completed returns thus was 128. One survey was returned blank after the second mailing. Thirty-nine letters were “returned to sender” by the postal services because the address was no longer valid.

These data are summarized in Table 2, with the resultant AAPOR calculations of return rate shown. Return Rate 1 was 26.6%.

Although there was no intent to conduct a telephone survey, the first two authors of this report looked for telephone numbers, using the same strategies as in the first study. A total of 321 numbers were found out of the sample of 500. For 26 names there were no telephone numbers at all in the online databases used. Another 38 were identified as unlisted, and for 115 names, the address did not match the name in the online records.

One hundred and twenty-nine voters returned questionnaires. In one case, the form was not filled out. A telephone number had been found for this respondent. Four additional questionnaires were returned with the voter identification number removed. Three of these were partials; all four have been treated as unusable. Of the 124 returned questionnaires that could be identified, 92 were among the cases for which telephone numbers were found. Four of them were cases without any address or telephone number, eight were unlisted numbers, and 20 were identified as wrong addresses in the telephone numbers search.

Thirty-nine came back returned by the U.S. Postal Service. For eight of these, a telephone number had been found. In seven cases, no telephone number had been found. One of the 39 was unlisted, and in 23 cases the address did not match with the voter on the voter list.

A total of 336 questionnaires were never returned. In 220 cases a telephone number had been found. In 15 cases no telephone number had been found. In 29 cases the number was unlisted, and in 72 cases the address found did not match with the voter list. For a summary of the cases, see Table 5.

The voter registration lists obtained from the Secretary of State contained the same information as was included in the voter registration list obtained for the November 2008 study. Those records were updated, of course, to reflect voting behavior and registration since that time. The Secretary of State voter history file for 2008 was downloaded and merged with the purchased list. In addition, the voter history file for the March 17, 2009, election was downloaded and merged with the voter registration file.

Measures Study 2

The same variables were taken from the voter registration lists as in the first study. A new variable for vote in the March 17, 2009, SPLOST election as well as use of absentee or a provisional ballot for that election also was added to the file.

The official vote record showed 1,037 persons voted in favor of the SPLOST and 420 voted against it. The mail survey contained a measure of vote intent (or actual vote, if the voter had already cast a ballot).

The sampling procedures described above resulted in the creation of five samples that can be compared with the population on these characteristics. These samples are summarized below.

Sample 1: Sample Drawn (N=500). This is the random sample of voters assigned to interviewers.

Sample 2: of Eligible Respondents (N= 461). These are registered voters minus those confirmed as ineligible when the questionnaire was returned by the U.S. Postal Service.

Sample 3: Telephone Sample (N=313). This is the sample for which telephone numbers could be located. The eight cases for which the mailing came back undelivered were eliminated.

Sample 4: Interviewed Sample (N= 124). This is the sample of registered voters who returned a completed, usable questionnaire and had not eliminated the voter registration number.

Sample 5: Interviewed Sample Plus Confirmed Improperly Registered (N= 163). This is the sample of voters who returned a questionnaire plus voters found to have moved and become ineligible to vote.

As in Study 1, Sample 1 was compared to the population to make sure the simple random sample fell into the 95% of confidence level. Sample 2 was created to see if any biases appeared if only eligible persons were considered from Sample 1. This sample actually eliminated improperly register people. Although they were never used, telephone numbers were looked for, in order to verify the biases that might appear in a telephone survey using online directories to identify land lines numbers, because of the address-based sampling design itself. This was Sample 3. The purpose of Sample 4 is see if biases resulted from participation. Just as in the first study, Sample 5 is used to show if improperly registered people in Sample 1 (and in the population) play a part in any biases of Sample 4.

Findings Study 1

In order to have a complete view of the possible biases, each variable in the comparison tables will be discussed. The voter status shows little variation across the samples, with a slightly increased value for the telephone sample, and a more significant

variation for the interviewed sample. Even if the percentage still falls into the confidence interval, the variation is explicable by the fact that people who agree to being interviewed are generally more interested in the elections and are more likely to vote, therefore they are more likely to be active voters.

There are insignificant variations for the zip code. The distributions for precinct and the municipal names vary starting with the sample drawn, but no variation is greater than 2.8%, and they all fall into the confidence interval.

The variation of the race is rather insignificant. However, it should be point out that the improperly registered people added to the sample of completed interviews seem to be the explanation for the 2% variation in the final interviewed sample, if one compares the sample of interviewed people and the sample obtained by adding the improperly registered people. The gender has a variation of 4.6% for the interviewed sample. A reasonable explanation is that women are generally more likely to agree to the telephone interviews. Still, the percentage falls into the confidence interval and all the other variations for the rest of the samples are insignificant.

The registration year is slightly leaning toward more people registered before 2000, for the telephone sample. This was to be expected as well, since people registered more than eight years ago are among the most stable in terms of changing addresses> Therefore, there are better chances that their landline telephone numbers are registered in the public data bases. The sample obtained by adding the improperly registered to the interviewed sample has a small variation leaning toward the recently registered voters, which indicates the fact that improper data in the registration list is to be found probably to a greater extent among recently registered people.

The variations for the variable measuring the distribution of new voters or people who never voted since their registration are to be found within the telephone sample, and the sample interviewed, although all the percentages fall within the confidence interval. This shows that people whose telephone numbers are listed in online directories and people who would agree to take surveys are generally slightly more active. Again, adding the improperly registered people to the interviewed sample reduces the variation.

The greatest variation among the samples, as compared to the population are related to the variables measuring whether or not people voted in the five 2008 elections. The greatest differences are in the interviewed sample. In four cases out of the five, the bias follows a trend toward an increased activism among people for whom telephone numbers were found, and was more prominent for the people who took the survey and were more likely to have voted during the year. Again, the sample containing the improperly registered people reduces somewhat the bias and it is a good indicator of one of the possible causes of the bias, the improper data in the whole population, which will never be found in the interviewed sample.

The three variables measuring the percentages of people who chose to vote in the Republican ballot on February, July, and August elections show an increasing trend toward a Republican choice in the samples, although the differences are smaller than the cases of the variables measuring the vote itself; most of them also fall into the confidence interval. The sample containing the improperly registered people slightly reduces the differences for this variable too.

The voter absentee measure has insignificant variations, except for the absentee in the November elections, for the interviewed sample and the interviewed and confirmed

improperly registered sample. The differences show that people who voted earlier are more likely to take a telephone survey, which is consistent with the slightly more active people in the interviewed sample overall.

The outcome of the elections, as a means to verify the accuracy of the final sample of interviewed people needs further explanations. Out of the 200 people, 178 actually expressed their vote preference for the elections, 2 declared they would not vote, 13 were undecided at the time of the survey, and 6 refused to reveal their options, and one just skipped this particular question when completing the survey. Out of the 178, 64.6% declared they would vote McCain, 33.7% declared for Obama, and 1.7% said other (Bob Barr). In fact, after the elections, the records show that 164 of them actually voted. Of these, 67.3% declared they would vote McCain, 31.5% Obama, and 1.2% Bob Barr.

The real outcome for Oconee County was 71.1% voted McCain, 27.7% Obama, and 1.2% Barr. The sample error calculated for the 178 persons sample is 7.4%, and the result falls into the confidence interval. However, as Murray, Riley and Scime (2009) suggested, in forecasting presidential elections outcome, the probability of voting increases with intent to vote and previous presidential vote. Therefore, before the elections and without knowing who will actually vote, a more reliable measurement of forecasting the vote outcome should have been obtained measuring the percentages within the group of people who voted in the previous presidential elections and declared they were likely to vote on November 4. There were 106 people who definitely intended to vote or had already voted in 2008 and also voted in 2004 presidential elections. Based

on this sample, the forecasting of the outcome is: 69.8% declared they would vote McCain, 28.3% Obama, and 1.9% Barr, a very accurate prediction for the sample size.

As far as the turnout of the presidential elections is concerned, 87.5% declared they would definitely vote or had already voted by absentee, even though not all of them declared how they would vote. Another 9.5% estimated more chances for voting than for not voting (chances of 6 to 9 on a 1 to 10 scale), although they were not sure about it. The real turnout for the interviewed sample is 91.5%. Most of the people (19 of 23) who were not sure about voting finally did vote, even people who estimated low chances (2 to 5 on a scale of 1 to 10). However, some of the people who declared they would definitely vote or had already voted did not actually vote. The real turnout for Oconee County was 80.2%. The comparisons indicate an increased activism among interviewed people, as compared with the population, and a slight tendency to overreport vote.

Overall, the small biases due to the mode of interview seem to follow the rule of an increased activism among people who usually complete opinion polls, and, to a significantly smaller extent, among people listed with workable telephone numbers in the online directories. All the variables for which the percentages do not fall into the confidence interval are related to previous elections. For all these cases the improperly registered people seem to play a part in the biases observed. The comparison between the population, the sample drawn, the sample of eligible respondents and the telephone sample indicates accurate samples in general, which shows no biases due to the mode of interview itself. The people left out of the sample, for lack of telephone numbers, are evenly distributed and do not cause significant biases in the final sample.

Vote behavior also indicates a good sample. Self-reported behavior shows that overreporting vote influences turnout estimations, which are already influenced by the increased activism of the interviewed sample. However, outcome seems to be very accurately predicted, especially following the rule of refining predictions based on vote behavior in previous presidential elections and vote intention.

Findings for study 2

The samples for the second study were built as to be replicas of the samples in the first study; the only difference is that interviewed sample is no longer a sub-sample of the telephone sample. The differences between the samples will be discussed for each variable.

The voter's status shows small differences for the telephone sample and the interviewed sample, within the confidence interval. The value of the variable in the sample of interviewed and improperly registered people shows that it is probable that a good share of the inactive people is to be found among the improperly registered people in the population, and therefore it would have been impossible to interview most of them.

There is a small difference, within the confidence interval, for the zip code and the precinct distribution in the interviewed sample. Percentages in all the other samples are very close to the distribution in the population. No significant differences were found for the municipal name.

People interviewed, as well as people improperly registered are white in a greater proportion than people in the population. Both differences still fall into the confidence interval. Gender is evenly distributed in all the samples, with a slightly greater proportion of males among the people for whom telephone numbers were found. It is possible that

males in the household would rather list their telephone numbers in the online directories, although in most cases both husband and wife were found in the directories used.

The registration year shows expected small biases for the telephone sample, the interviewed sample, and the interviewed and improperly registered sample. For the first two samples, it was most likely for people who have been living in Oconee County for a longer period of time to have their names listed in the online telephone directories, as well as to be more interested in the SPLOST local issue than for people who had recently moved in the county. Therefore, it was expected that both samples would have a variation for this particular variable. The variation for the interviewed and improperly registered sample suggests that the difference between the interviewed sample and the population is even smaller than what the actual numbers show.

The same increased activism can be observed in the variable measuring the distribution of people newly registered to vote or who had never voted since registration, in both telephone sample and interviewed sample. However, the percentages fall into the confidence interval, and the sample of interviewed and improperly registered people reduces the variation, thus offering a new explanation for the differences observed.

All the variables measuring whether or not people voted in the 2008 (February, July, August, November, December) and 2009 (March) elections show more active people in the telephone sample and the interviewed sample. An interesting finding (also observed, to a smaller extent, in the first study) is that people listed in the telephone directories are more likely to be active voters than the population in general. Interviewed people are even more interested in voting issues in general and therefore more likely to vote in both local and national elections. For all the six elections, the sample of

interviewed and improperly registered voters considerably reduces the biases, showing that another explanation for the differences between the interviewed sample and the population is the fact that voters' list contains a good share of improperly registered persons in general.

The percent of people using a Republican ballot in the three primary elections (February, July, and August) suggest a slightly more Republican preference for all the samples, with the exception of the Presidential Primary Election for the interviewed sample. However, all the percentages fall into the confidence interval.

The comparison for the "Absentee" variables shows significant and consistent variation only for the interviewed sample, which suggests again that people who are more likely to respond to questionnaires are more active in general, particularly more likely to vote earlier. The interviewed and confirmed improperly registered sample offers again a secondary explanation. In all six cases the percentages in this sample significantly reduce the difference in the interviewed sample, if compared with the population.

As far as the vote itself in the SPLOST election is concerned, the results raise interesting issues. If one tries a forecast of the outcome based on what people declared in the questionnaires, there are 32 people who declared they voted (26.2% of the 122 interviewed people how answered that particular question). The real outcome of the SPLOST election was 6.6%. Twenty-nine of them also indicated how they voted, 69.0% for and 31.0% against. The real outcome of the election was 71.2% for and 28.8% against the tax, which would indicate a very accurate prediction for such a small number of respondents. However, voters' registration list shows that only half of the respondents who reported they voted actually did so. Fourteen of the 29 respondents voted, 9 for

(64.3%) and 5 against (35.7%) the SPLOST tax. Even though only half of the people who declared they voted actually did, results show that had the other half actually voted, they would have voted in a very similar way with those who actually participated in the election.

Overall, the comparisons for the second study show a replication of what was observed in the first one. All significant but small biases are related to the degree of activism of people who are listed in the online telephone directories on the one hand, and of people who are more likely to take a public opinion survey on the other hand. All the demographics show very accurate samples, for all five samples built for this study. Just as in the first study, the interviewed and improperly registered sample suggests that the biases are even smaller than the numbers show, because the percentages for this particular sample always seem to reduce the observed differences in the interviewed sample.

The findings also show that even though predictions for turnout are influenced by a slightly increased activism of people responding to the survey and even more by the overreported vote behavior, outcome can be accurately forecast, regardless of the real vote behavior of interviewed people.

Discussion

The purpose of the two studies presented in the paper was to propose a very reliable and inexpensive sampling procedure for local public opinion surveys. Both modes of interview, telephone and mail surveys based on voter registration lists, provided accurate interviewed samples in general and needed very limited resources. RQ1 and RQ2 were thus answered, the results showing that address-based sampling using both

telephone numbers found in online directories and mailings provide accurate samples and represent a very good alternative to RDD techniques.

In both cases, the variables measuring demographics showed little variation across the samples analyzed. Although mail survey should have the advantage of covering people without landline telephones, and thus providing a more accurate sample, results show that telephone samples in both studies are accurate samples, with small variations, if compared to the population. This provides a new, free and accurate sampling methods using online resources for telephone surveys.

The small biases related to nonparticipation observed in both studies are all related to previous voting behavior. Telephone samples indicate that slightly more active people in general are to be found in the online telephone directories. Nevertheless, people without landlines telephone numbers publicly listed have roughly the same demographic characteristics with people for whom telephone numbers could be located in the online data bases.

Interviewed samples are generally accurate samples (if compared with the random sample drawn initially), with rather insignificant variations for the demographics, and small biases observed in the variables related to previous voting behavior, all showing an increased activism in general for people who agree to take the surveys, regardless of the mode of the interview. Nonparticipation biases are related to activism as well in both modes of interview, and with gender only for the telephone surveys, women being more likely to take telephone surveys than men. At the same time, hypotheses H1 and H2 were confirmed. Nevertheless it should be stated that the biases observed are rather small.

The samples of eligible respondents are generally very accurate samples, just slightly leaning toward more active people. In the same time, the samples of interviewed and confirmed improperly registered people considerably reduce activism biases observed in the interviewed samples. Both observations argue for another possible explanation of the differences observed in the variables measuring voting behavior: improperly registered people in the voter registration lists.

Interviewed samples in both studies tend to overreport vote. Correlated with the slightly increased activism of people in these samples, estimations for turnout lean toward overreporting vote. Hypotheses H3 and H4 were supported. We emphasize that survey mode does not affect in any way overreporting, and thus socially desirable behavior influences self-reported vote behavior both in a direct (telephone) and indirect (mail) communication. However, outcome forecast had proven a very accurate prediction in both cases, regardless of self-reported vote behavior. Hypotheses H5 and H6 were supported. Address-based sampling is a reliable technique for public opinion survey, offering good forecasting of elections outcome. Voter registration lists offer information that provides tools for a more refined and accurate prediction.

Conclusions

Probably the most important result of this research is related to the new sampling technique using address in voters' list and online telephone directories to obtain good samples for both telephone and mail surveys. This technique links voter lists with telephone numbers, thus providing a viable alternative to RDD in public opinion surveys.

Telephone numbers were found for about two-thirds of the people in the lists. However, the biases related to this mode of interview are only related to the activism of

people for whom telephone numbers could be found. Of them, even more active people actually take surveys (for both telephone and mail surveys), and women are more likely to respond in a telephone survey than men.

Another important finding is related to turnout. People tend to overreport vote in public opinion surveys. Biases in estimating turnout are due not only to the increased activism of people taking surveys (that was observed in both studies), but also to a general tendency to declare voting even when this is not true, probably because of the socially desirable dimension of voting itself. However, both studies showed that overreporting vote does not influence predictions of election outcome.

Overall, sampling within the population provided by voter registration lists has proven a reliable and inexpensive alternative to random digit dialling for local public opinion surveys. It also offers important information regarding people interviewed in the surveys, thus being a rich resource for alternative investigations related to the profile of people who vote in general, or for people who respond to public opinion surveys.

References

- Anderson, B. A., & Silver, B.D. (1986). Measurement and mismeasurement of the validity of self-reported vote. *American Journal of Political Science*, 30 (4), 771-785.
- Belli, R. F., Traugott, M. W., Young, M. & McGonagle, K. (1999). Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source. *Public Opinion Quarterly*, 63(1), 90-108.

- Bernstein, R., Chadha, A. & Montjoy, R. (2001). Overreporting Voting. Why It Happens and Why It Matters. *Public Opinion Quarterly*, 65(1), 22-44.
- Bishop, G. F., Fisher, B. S. (1995). "Secret Ballots" and Self-Reports in an Exit-Poll Experiment. *Public Opinion Quarterly*, 59(4), 568-588.
- Bolstein, R. (1991). Comparison of the likelihood to vote among preelection poll respondents and nonrespondents. *Public Opinion Quarterly*, 55(4), 648-650.
- Brick, M. J., Waksberg, J., Kulp, D. & Starer, A. (1995). Bias in list-assisted telephone samples. *The Public Opinion Quarterly*, 59(2), 218-235.
- Cassel, C. A. (2003). Overreporting and Electoral Participation Research. *American Politics Research*, 31(1), 81-92.
- Crespi, I. (1988). *Pre-election polling: Sources of accuracy and error*. New York Russell Sage Foundation.
- DeSart, J., & Holbrook, T. (2003). Campaigns, polls, and the states: Assessing the accuracy of statewide presidential trial-heat polls. *Political Science Quarterly*, 56(4), 431-39.
- Fullerton, A. S., Dixon, J. C. & Borch, C. (2007). Bringing Registration Into Models Of Vote Overreporting. *Public Opinion Quarterly*, 71(4), 649-660.
- Granberg, D. & Holmberg, S. (1991). Self-Reported Turnout and Voter Validation. *American Journal of Political Science*, 35(2), 448-459.
- Green, D. P. & Gerber, A. S. (2003). Enough already with random digit dialing: Can registration-based sampling improve the accuracy of election forecasts? Paper presented at the *Annual Meeting of the American Association for Public Opinion Research*, May 16-19, 2003, Nashville, TN.

- Green, D. P. & Gerber, A. S. (2006). Can registration-based sampling Improve the accuracy of midterm election forecasts?. *The Public Opinion Quarterly*, 70(2), 197-223.
- Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.
- Karp, J. A. & Brockington, D. (2005). Social Desirability And Response Validity: A Comparative Analysis Of Overreporting Voter Turnout In Five Countries. *The Journal Of Politics*, 67(3), 825-840.
- Katosh, J. P., Traugott, M. W. (1981). The Consequences of Validated and Self-Reported Voting Measures. *Public Opinion Quarterly*, 45(4), 519-535.
- Lau, R. R. (1994). An analysis of 'trial heat' polls during the 1992 presidential election. *Public Opinion Quarterly*, 58(1), 2-20.
- Mann, C. B. (2003). Improving pre-election forecasts from registration based sampling: Using voter registration data to predict partisan vote intention and to allocate undecided voters. Paper prepared for the 2003 Annual Conference of the American Political Science Association, Philadelphia, PA.
- Mann, C. B. (2005). Do advance letters improve preelection forecast accuracy? *Public Opinion Quarterly*, 69(4), 561-571.
- McDonald, M. P. (2007). The true electorate: A cross-validation of voter registration files and election survey demographics. *The Public Opinion Quarterly*, 71(4), 88-602.
- Mitofsky, W., Bloom J., Lenski, J., Dingman, S., & Agiesta, J. (2005). A dual frame RDD/registration-based sample design: Lessons from Oregon's 2004 National

- Election Pool Survey. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, Alexandria, VA, 3929–36.
- Monson, Q. (1998). Predicting voter turnout in local elections: A vote validation study. Paper presented at the *Annual Meeting of the Midwest Association for Public Opinion Research*, Chicago, IL.
- Murray, G. R., Riley, C. & Scime, A. (2009). Pre-Election Polling: Identifying Likely Voters Using Iterative Expert Data Mining. *Public Opinion Quarterly*, 73(1), 159-171.
- Petrocik, J. R. (1991). An algorithm for estimating turnout as a guide to predicting turnout. *Public Opinion Quarterly*, 55(4), 643-647.
- Presser, S. (1984). Is Inaccuracy on Factual Survey Items Item-Specific or Respondent-Specific?. *Public Opinion Quarterly*, 48(1), 344-355.
- Presser, S. & Traugott, M. (1992). Little White Lies and Social Science Models. Correlated response errors in a panel study of voting. *Public Opinion Quarterly*, 56(1), 77-86.
- Siegelman, L. (1982). The nonvoting voter in voting research. *American Journal of Political Science*, 26(1), 47-56.
- Siegelman, L., Roeder, P. W., Jewell, M. E. & Baer, M. A. (1985). Voting and nonvoting: A multi-election perspective. *American Journal of Political Science*, 29(4), 749-765.
- Silver, B. D., Anderson, B. A., Abramson, P. R. (1986). Who Overreports Voting?. *The American Political Science Review*, 80(2), 613-624.

- Taylor, H. (1997). Why most polls overestimated Clinton's margin. *Public Perspective*, 8(2), 45-48.
- Traugott, M. W. (2005). The Accuracy Of The National Preelection Polls In The 2004 Presidential Election. *Public Opinion Quarterly*, 69(5), 642-654.
- Traugott, M. W., & Tucker, C. (1984). Strategies for predicting whether a citizen will vote and estimation of electoral outcomes. *Public Opinion Quarterly*, 48(1), 330-343.
- Visser, P. S., Krosnick, J. A., Marquette, J. & Curtin, M. (1996). Mail surveys for election forecasting? An evaluation of the Columbus dispatch poll. *Public Opinion Quarterly*, 60(2), 181-227.
- Volgy, T. J., Schwarz, J. E. (1984). Misreporting and Vicarious Political Participation at the Local Level. *Public Opinion Quarterly*, 48(4), 757-765.
- Voogt, R. J., & Vam Kempoen, H. (2002). Nonresponse bias and stimulus effects in the Dutch national election study. *Quality and Quantity*, 36(4), 325-345.
- Voss, D. S., Gelman, A., & King, G. (1995). Preelection survey methodology: Details from eight polling organizations, 1999 and 1992. *Public Opinion Quarterly*, 59(1), 98-132.

Table 1. AAPOR return rates for study 1	
Final Disposition Codes for RDD Telephone Surveys (Adapted for ABS)	
1. Interview (1.0)	200
Complete (1.1)	199
Partial (1.2)	1
2. Eligible, Non-Interview (2.0)	198
Refusal and break-off (2.10)	107
Refusal (2.11)	
Household-level refusal (2.111)	
Known respondent refusal (2.112)	
Break-off (2.12)	
Non-contact (2.20)	75
Respondent never available (2.21)	43
Telephone answering device	32
(message confirms residential household) (2.22)	
Message left (2.221)	
No message left (2.222)	
Other (2.30)	16
Dead (2.31)	
Physically or mentally unable/incompetent (2.32)	
Language (2.33)	1
Household-level language problem (2.331)	
Respondent language problem (2.332)	
No interviewer available for needed language (2.333)	
Inadequate audio quality (2.34)	
Location/Activity not allowing interview (2.35)	
Miscellaneous (2.36) / Completed by other method + Illegal substitution	15
3. Unknown Eligibility, Non-Interview (3.0)	128
Unknown if housing unit (3.10)	128
Not attempted or worked (3.11)	55
Always busy (3.12)	4
No answer (3.13)	69
Telephone answering device (don't know if housing unit) (3.14)	
Telecommunication technological barriers, e.g., call-blocking (3.15)	
Technical phone problems (3.16)	
Ambiguous operator's message (3.161)	
Housing unit, Unknown if eligible respondent (3.20)	
No screener completed (3.21)	
Unknown if person is household resident (3.30)	
Other (3.90)	
4. Not Eligible (4.0)	274
Out of sample (4.10) / Not at the address anymore	32
Fax/data line (4.20)	
Non-working/disconnected number (4.30)	232
Non-working number (4.31) Number not found	182
Disconnected number (4.32)	
Temporarily out of service (4.33)	50
Special technological circumstances (4.40)	10
Number changed (4.41) / Wrong address	10
Call forwarding (4.43)	
Residence to residence (4.431)	

Nonresidence to residence (4.432)	
Pagers (4.44)	
Cell phone (4.45)	
Landline phone (4.46)	
Nonresidence (4.50)	
Business, government office, other organization (4.51)	
Institution (4.52)	
Group quarters (4.53)	
Person not household resident (4.54)	
No eligible respondent (4.70)	
Quota filled (4.80)	
e=never tried/total	32,0%
RR1	38,0%
RR2	38,0%
RR3	45,6%
RR4	45,6%
RR5	50,0%
RR6	50,3%
COOP1	61,6%
COOP2	61,9%
COOP3	64,8%
COOP4	65,1%
REF1	20,3%
REF2	24,4%
REF3	26,9%
CON1	61,4%
CON2	73,6%
CON3	81,2%

Table 2: Comparison between the samples for study 1

	Popu- lation	Sample Drawn	Sample of Eligible Respondents	Tele- phone Sample	Sample Interviewed	Tel. Sample + Confirmed Improperly Registered
Voter status - Active	92,4%	91,9%	92,6%	95,3%	98,5%	95,0%
Zip Code - Watkinville	52,3%	53,5%	53,1%	53,4%	52,0%	53,7%
Percinct - Rural Southern	19,1%	16,3%	16,1%	16,7%	17,0%	19,0%
Municipal name - Incorporated Area	13,5%	14,5%	14,1%	13,2%	11,0%	11,1%
Race - White	92,1%	91,0%	91,5%	93,1%	94,0%	92,6%
Gender - Male	47,6%	46,4%	46,2%	48,0%	43,0%	46,7%
Registration Year 2001 through 2008	54,4%	54,0%	53,2%	50,9%	52,0%	56,2%
Last Vote Year - new voters/never voted since registration	16,7%	18,3%	17,5%	13,6%	11,5%	14,5%
Voted Feb 5 2008 Presidential Primary Election- Yes	45,5%	45,0%	46,9%	52,2%	60,0%	54,5%
Voted July 15 2008 Local/State Primary Election - Yes	32,7%	32,3%	33,6%	38,0%	45,0%	39,7%
Voted Aug 5 2008 Local/State Primary Runoff - Yes	11,4%	11,1%	11,5%	12,5%	15,0%	13,2%
Voted Nov 4 2008 General Election - Yes	77,0%	75,8%	76,7%	82,4%	91,5%	86,0%
Voted Dec 2 2008 Senatorial Runoff - Yes	49,0%	47,0%	48,1%	53,0%	58,5%	55,0%
Pct. Voters using Republican Ballot - Feb 5 2008	66,4%	69,8%	70,9%	72,9%	69,2%	68,1%
Pct. Voters using Republican Ballot - July 15 2008	90,8%	91,9%	93,2%	94,8%	96,7%	95,7%
Pct. Voters using Republican Ballot - Aug 5 2008	86,7%	87,5%	88,7%	91,2%	93,3%	90,9%
Voted Absentee Feb 5 2008 - Yes	5,9%	4,1%	4,1%	4,3%	5,0%	5,4%
Voted Absentee July 15 2008 - Yes	8,4%	7,4%	7,4%	7,4%	7,5%	6,6%
Voted Absentee Aug 5 2008 - Yes	2,5%	2,5%	2,6%	2,4%	1,5%	1,2%
Voted Absentee Nov 4 2008 - Yes	52,6%	49,1%	49,6%	54,2%	61,5%	58,3%
Voted Absentee Dec 2 2008 - Yes	18,3%	16,1%	16,5%	18,5%	18,0%	17,8%
N	21352	800	703	552	200	242
Sample error		3,5%	3,7%	4,2%	6,9%	6,3%

Table 3: AAPOR return rates for study 2	
Final Disposition Codes for Mail Surveys of Specifically Named Persons	
1. Returned questionnaire (1.0)	124
Complete (1.1)	124
Partial (1.2)	
2. Eligible, "Non-Interview" (2.0)	2
Refusal & Break-off (2.10)	1
Refusal (2.11)	
Other person refusal (2.111)	
Known respondent-level refusal (2.112)	
Blank questionnaire mailed back, "implicit refusal" (2.113)	
Break-off questionnaire too incomplete to process (2.12)	1
Non-Contact (2.20)	1
Other notification that respondent was unavailable during field period (2.26)	
Completed questionnaire, but not returned during field period (2.27)	1
Other (2.30)	0
Death (including USPS category: deceased) (2.31)	
Physically or mentally unable/incompetent (2.32)	
Language (2.33)	
Respondent language problem (2.332)	
Wrong language questionnaire sent for needed language (2.333)	
Literacy problems (2.34)	
Miscellaneous (2.36)	
3. Unknown eligibility, "non-interview" (3.0)	370
Nothing known about respondent or address (3.10)	331
Not mailed (3.11)	
Nothing ever returned (3.19)	331
Unknown if eligible respondent in unit (3.20)	9
No screener completed (3.21)	
USPS category: refused by addressee (3.23)	
Refused to accept (3.231)	
Refused to pay postage (3.232)	
USPS category: returned to sender due to various USPS violations by addressee (3.24)	
USPS category: cannot be delivered (3.25)	7
USPS Category: Illegible Address (3.251)	
USPS Category: Insufficient Address on Mail from One Post Office to Another Post Office (3.252)	
USPS Category: No Mail Receptacle (3.253)	2
USPS Category: Delivery Suspended to Commercial Mailing Agency (3.254)	
Unknown Whereabouts, Mailing Returned Undelivered (3.30)	30
Cannot Be Delivered as Addressed (3.31)	23
USPS Category: Attempted — Addressee Not Known (3.311)	
USPS Category: Postal Box Closed (3.312)	
No Such Address (3.313)	
USPS Category: No Such Number (3.3131)	1
USPS Category: No Such Office in State (3.3132)	
USPS Category: No Such Street (3.3133)	1
USPS Category: Vacant (3.3134)	
Not Delivered as Addressed (3.314)	
USPS Category: Unable to Forward (3.3141)	3
USPS Category: Outside Delivery Limits (3.3142)	

USPS Category: Returned for Better Address (3.3143)	
USPS Category: Moved, Left No Address (3.32)	2
USPS Category: Returned for Postage (3.33)	
USPS Category: Temporarily Away, Holding Period Expired, Unclaimed (3.34)	
USPS Category: Unclaimed -- Failure to Call for Held Mail (3.35)	
USPS Category: No One Signed (3.36)	
Returned with Forwarding Information (3.40)	0
Returned Unopened — address correction provided (3.41)	
Returned Opened — address correction provided (3.42)	
Other (3.9)	
4. Not Eligible, Returned (4.0)	4
Selected Respondent Screened Out of Sample (4.10)	4
No eligible respondent (4.70)	
Quota Filled (4.80)	
Duplicate Listing (4.90)	
e=	89,5%
RR1	26,6%
RR2	26,6%
RR3	28,8%
RR4	28,8%
RR5	98,4%
RR6	98,4%
COOP1	99,2%
COOP2	99,2%
COOP3	99,2%
COOP4	99,2%
REF1	0,2%
REF2	0,2%
REF3	0,8%
CON1	26,8%
CON2	29,1%
CON3	99,2%

Table 4. Comparison between the samples for study 2

	Population	Sample drawn	Eligible Respondents	Telephone sample	Mail Sample	Mail Sample + Confirmed Improperly Registered
Voter status - Active	94,2%	93,6%	96,5%	99,4%	100,0%	90,2%
Zip Code - Watkinsville	52,0%	51,8%	50,3%	53,7%	47,6%	52,8%
Precinct - Rural Southern	19,1%	20,2%	19,7%	20,8%	16,1%	18,4%
Municipal name - Incorporated Area	13,4%	13,4%	13,2%	12,8%	15,3%	15,3%
Race - White	91,5%	93,6%	93,1%	94,6%	97,6%	98,2%
Gender - Male	47,5%	48,2%	48,2%	50,2%	48,4%	48,5%
Registration Year 2001 through 2008	56,5%	57,2%	56,8%	48,9%	44,4%	48,5%
Last Vote Year - new voters/never voted since registration	10,1%	11,4%	10,4%	6,4%	2,4%	7,4%
Voted Feb 5 2008 Presidential Primary Election - Yes	43,5%	42,4%	44,9%	53,7%	54,8%	44,8%
Voted July 15 2008 Local/State Primary Election - Yes	31,5%	30,2%	32,1%	39,0%	41,9%	33,7%
Voted Aug 5 2008 Local/State Primary Runoff - Yes	11,0%	12,6%	13,4%	17,6%	21,8%	17,2%
Voted Nov 4 2008 General Election - Yes	77,6%	76,4%	80,7%	87,9%	94,4%	77,9%
Voted Dec 2 2008 Senatorial Runoff - Yes	48,8%	48,2%	51,6%	61,3%	69,4%	54,6%
Voted Mar 17 2008 SPLOST	6,1%	5,4%	5,9%	7,3%	12,9%	9,8%
Pct. Voters using Republican Ballot - Feb 5 2008	66,7%	70,8%	70,0%	70,2%	66,2%	68,5%
Pct. Voters using Republican Ballot - July 15 2008	90,8%	91,4%	91,2%	90,2%	94,2%	94,5%
Pct. Voters using Republican Ballot - Aug 5 2008	86,5%	87,3%	87,1%	87,3%	92,6%	92,9%
Voted Absentee Feb 5 2008 - Yes	5,6%	7,2%	7,8%	9,9%	14,5%	11,0%
Voted Absentee July 15 2008 - Yes	8,0%	7,4%	7,8%	9,6%	13,7%	11,0%
Voted Absentee Aug 5 2008 - Yes	2,4%	2,4%	2,6%	3,5%	4,8%	3,7%
Voted Absentee Nov 4 2008 - Yes	53,0%	48,8%	52,3%	58,8%	66,1%	52,1%
Voted Absentee Dec 2 2008 - Yes	18,1%	17,8%	19,1%	22,7%	30,6%	23,9%
Voted Absentee Mar 17 2008 - Yes	1,9%	1,8%	2,0%	2,9%	4,8%	3,7%
N	22094	500	461	313	124	163
Sample error		4,4%	4,6%	5,5%	8,8%	7,7%

Table 5. Telephone numbers found distributed by mailing outcome

Telephone numbers	Return Received	Returned Refused	Returned to Sender	No Response	Total
Found	92	1	8	220	321
Not found	4	0	7	15	26
Address matched voter but "unlisted" in white pages	8	0	1	29	38
Address did not match voter	20	0	23	72	115
Total	124	1	39	336	500