Using Voter Lists To Create Telephone and Mail Samples:

Two Validation Studies from a Single Community

Nicoleta Corbu

(nicoleta.corbu@comunicare.ro)

College of Communication and Public Relations

National School of Political Studies and Public Administration

Bucharest, Romania

Qingmei Qing & Lee B. Becker

(cecilia2001@gmail.com) & (lbbecker@uga.edu)

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

Grady College of Journalism and Mass Communication

University of Georgia

Introduction

One of the main problems of survey methodology is related to possible biases due to the sampling method, especially biases related to low return rates. Currently, random telephone number generation is the most widely used methods in telephone surveys, which are still the most widely used sampling method in election polls. In recent years, Registration-based sampling (RBS) method has gain a lot of attention in public opinion research field and have been approved by many scholars for its effectiveness of gaining a relatively representative sample. The present research focuses on two validation studies from the same community using RBS method.

The present research has two main purposes . First, the study tried to compare the survey response rates between telephone survey and mail survey under the condition of the same sampling method- RBS. Second, the current research also focused on the capacity of address-based sampling using voter registration lists to provide an accurate sample for election polls, both in telephone surveys and in mail surveys. This paper presents the results of two different studies completed in November 2008 and March 2009. The first used a telephone survey based on a sample drawn from a voter registration list. The second used a mail survey based on a voter registration list as well.

Literature Review and Research Questions

Survey Response Rate

Nonresponse is a critical problem for survey researchers. High level of nonresponse can produce systematic or non-random error, which will affect researchers' estimation of significant population characteristics (Goldstein & Jennings, 2002). As Brehm (1993, p.93) argues: "Every conceivable method of analyzing survey data is at risk because of nonresponse." Bias is not simply a function of the number of sampling units that refused to take the survey or failed to be reached (Groves & Couper, 1998). Instead, it is a combined function of both the percentage of sampling units and the degree to which respondents and nonrespondents are different on important measures. Recent studies show that substantial robustness can be achieved even with very low response rates (e.g., Keeter et al, 2000).

The response rates depend on a number of survey design characteristics: the topic of the survey, the survey population, the efforts at refusal conversion, the duration of the interview period, information about the sample persons, and so on (Traugott, Groves & Lepkowski, 1987). Before making decisions on a survey design, a balance needs to be made between the research costs of design attributes and the likely nonresponse error reduction (Traugott, Groves & Lepkowski).

There is an increasing interest in improving return rates in different kinds of surveys. Research concerning particularly mail surveys focused on several important issues related to the return rates' improvement: follow-up letters effects (Sudman, 1982; Tedin & Hofstetter, 1982; James & Bolstein, 1990), incentive effects (Nederhof, 1983; Mizes,

1984; James & Bolstein, 1990; Church, 1993;Trussel & Lavrakas, 2004), business reply postage vs. first class postage (Armstrong & Lusk, 1987), and different options of completing mail surveys (Dillman, West & Clark, 1994).

Fox, Crask, and Kim (1988) found that prenotification, follow-ups, and stamped return postage were the most important determinants of mail survey response rates. In addition, sponsorship and small monetary incentives improve response rates. Worthen and Valcarce (1985) found that the effect size for personalization from previous studies was small, and in a follow-up study found that cover letter personalization has no effect on enhancing response rates. Armstrong and Lusk (1987) conducted a meta-analysis and found that the only effect of consequence on response rate was first-class postage compared with business reply, and the effect was weak. James and Bolstein (1990) argue that the two most effective methods of increasing response rates are: sending follow-up letters with replacement questionnaires, and enclosing incentives with the initial mailing. They also found that a combination of there two kinds of techniques is more effective than either technique used alone.

Comparisons between mail surveys and other types of surveys showed rather similar return rates in face-to-face and mail questionnaires (Goyder, 1985, Krysan, Schuman, Scott & Beatty, 1994), similar return rates in web vs. mail surveys, with age differences in the samples (Kaplowitz, Hadlock & Levine, 2004), and better forecasting accuracy of mail surveys compared to random digit dialing surveys (Visser, Krosnick, Marquette & Curtin, 1996).

Random Digit Dialing (RDD) and Registration-based Sample (RBS) Designs

The Mitofsky-Waksberg method of random digit dialing (RDD) is the best know and probably the most widely used survey method (Brick et al., 1995). This method 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). According to Waksberg (1978), the major attribute of this method is that it is a self-weighting sample of all residential telephone numbers that significantly reduces the proportion of nonresidential telephone numbers, which must be dialed below the amounts needed with simple random sampling.

There are some disadvantages of the RDD sampling method. Pollsters may discard or down-weight those interviews they conduct with people whom they considered as unlikely voters (Green & Gebner, 2003). Moreover, interviewing nonvoters can be meaningless for the research; however, there is no way to identify which respondents will be possible voters before conducting the interviews.

As an alternative, registration-based sample design (RBS) provides a way of differentiating possible voters and nonvoters before research money is spent for interviews (Green & Gebner). In recent years, researchers who conducted pre-election surveys have experimented with samples drawn from voter registration lists (RBS). RBS is used because researchers can indentify 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. In addition, they can be used as stratification variables, which can improve the design of the sample (Mitofsky, et

al., 2005). Whether the probabilities of selection for individuals are equal or unequal will depend on the rates of selection in various strata.

As recent studies mention (Green, Gerber, 2006; McDonald, 2007), relatively little has been written about the practical details of registration-based sampling. Although relatively rarely used in research and political surveys in general, the voter registration list sampling is shown to have considerable advantages as compared to the random digit dialing procedure (Green, Gerber, 2006; McDonald, 2007). The include the availability of useful background information about the respondents, simpler interview protocols, higher response rates (particularly in low-salience elections), and the reduction of the costs associated with identifying likely voters. 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.

Mitofsky, et al. (2005), developed a dual frame sample to achieve the advantages of voter registration lists (RBS) supplemented with a random digit dialing (RDD) sample to eliminate the undercoverage bias. Specifically, the frame was designed because the success rate for matching the registration list of a state to commercial phone lists typically ranges between 50 and 65 percent.

Adimora, et al. (2001), compare coverage of a state driver's license list and county voter registration lists as frames for sampling rural African Americans. It was found that the driver's list provided better coverage than did voter registration lists. Federal legislation that prohibits states from releasing driver's license lists for use in surveys, however, removes the driver's license as a valuable resource for population-based research.

Green and Gerber (2003) test the relative accuracy of RDD and RBS in predicting election outcomes. Their results suggest that RBS compares favorably to RDD in terms of forecasting accuracy. In addition, they also argue that voter registration lists often furnish useful background information that strongly predicts whether a person will vote. In addition, sampling from registration lists also reduces the costs associated with identifying likely voters.

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 then RDD.

Research questions of the present study were proposed as follows:

RQ1: Which survey method will provide higher return rates between mail-based and telephone-based survey?

RQ2: Will voter-registration-based sampling design provide a relatively accurate sample representing population parameters in election polls in mail survey?

RQ3: Will voter-registration-based sampling design provide a relatively accurate

sample representing population parameters in election polls in telephone

survey?

Research Design

Design of 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 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.

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.

Design of 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.

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 behaviour 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.

Measurement

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

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

Measures of 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 favour of the SPLOST and 420 voted against it. The mail survey contained a measure of vote intent (or actual vote, if the voter has already cast a ballot).

The sampling procedures described above resulted in the creation of four 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.

Findings

Findings of Study 1

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%.

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 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 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 preferences 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% Obama, and 1.7% other (Bob Barr). In fact, after the elections, the records show that 164 of them actually voted. Sixty-seven point three percent of them 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 of study 2

Eighty-one valid completed questionnaires were returned after the first mailing, and another 43 after the second mailing and before the closing data 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 numbers, and in 72 cases the address found did not match with the voter list. For a summary of the cases, see table 5.

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 SPOLSP election was 6.6%. Twenty-nine of them also responded how they voted, 69% for and 31% against the SPLOST tax. 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 pretended they voted actually did. 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 purposes of the two studies in the paper were to compare the response rates between telephone and mail survey methods, and to propose a reliable and not expensive 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.

The first research questions is about the comparison of response rates between telephone survey and mail survey. The response rates of the first study using telephone interview method was 38%; and the response rates of the second study using mail survey method was 26.6%. It is obvious that the response rate of telephone survey was much higher than that of the mail survey method, and the difference is 11.4%. One possible explanation is related to the topic. Traugott, Groves & Lepkowski (1987) argued that the survey topic is one factor that decides the response rates. In this present research, people might be more interested in talking about relatively larger social issues. The topic of the

first study was about voting decision of Presidential election, which was easier to draw more attention from the public. People might consider the survey as an additional channel for them to give their voice so that they had more motivation in this case. All in all, the present studies showed that telephone survey method had a higher response rate than mail survey since the matching rate of telephone numbers in the two studies were close (the first study was around 66%, and the second one was 64%). In addition, mail survey usually costs more than telephone survey, particularly when follow-up letter strategy is used. The current study findings indicated that mail survey can be an alternative to telephone surveys; however, future researchers should make a balance between research cost and possible increase of response rate.

The second and third research questions were about how accurate the samples in a voter-registration-based sampling design in terms of the representatives of population parameters using telephone and mail survey respectively. In both cases, the variables measuring demographics such as voter status, zip code, race, gender, registration year and so on 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 showed that telephone samples in both studies were accurate samples, with small variations, if compared to the population.

The small biases 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 characteristics with people for whom telephone numbers could be located in the online data bases.

Interviewed samples are generally accurate samples, 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.

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 over report vote. Correlated with the slightly increased activism of people in these samples, estimations for turnout lean toward overreporting vote. However, outcome forecast had proven a very accurate prediction in both cases, regardless of self-reported vote behavior.

Overall, sampling within the population provided by voter registration lists has proven a reliable and not expensive alternative to random digit dialing 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

Adimora A. A. , Schoenbach V. J., Martinson F. E., Stancil, T. R., & Donaldson, K. H.

    (2001). Driver's license and voter registration lists as population-based sampling

frames

    for rural African Americans. *Ann Epidemiol 2001, 11,* 385–388.

Brick, M. J., Waksberg, J., Kulp, D., Starer, A. (1995). Bias in list-assisted telephone

samples.

    The Public Opinion Quarterly, *59*, 218-235.

Armstrong, S. J., & Lusk, E. J. (1987). Return postage in mail surveys a meta-analysis.

    The Public Opinion Quarterly, *51 (2)*, 233-248.

Brehm, J. (1993). *The phantom respondents.* Ann Arbor: University of Michigan Press.

Church, A. H. (1993). Estimating the effect of incentives on mail survey response rates:

    A meta-analysis. *The Public Opinion Quarterly*, 57 (1), 62-79.

Dillman, D. A., West, K. K., & Clark, J. R. (1994). Influence of an invitation to answer

    by telephone on response to census questionnaires. *The Public Opinion Quarterly*, *58

    (4),* 557-568.

Fox, R. J., Crask, M. R., & Kim, J.(1988). Mail survey response rate: A meta-analysis of

    selected techniques for inducing response. *Public Opinion Quarterly, 52,* 47-91.

Goldstein, K. M., & Jennings, M. K. (2002). The effect of advance letters on cooperation

    in a list sample telephone survey. *Public Opinion Quarterly, 66,* 608-617.

Goyder, J. (1985). Face-to-face interviews and mailed questionnaires: The net difference

    in response rate. *The Public Opinion Quarterly*, *49(2),* 234-252.

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*

studies. New York: Wiley.

James, J. M., & Bolstein, R. (1990). The effect of monetary incentives and follow-up

mailings on the response rate and response quality in mail surveys. *The Public*

Opinion Quarterly, 54 (3), 346-361.

Kaplowitz, M. D., Hadlock, T. D., Levine, R. (2004). A comparison of web and mail

survey response rates. *The Public Opinion Quarterly*, *68(1),* 94-101.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of

reducing nonresponse in a national telephone survey. *Public Opinion Quarterly,*

64(2), 125-48.

Krysan, M., Schuman, H., Scott, L. J., Beatty, P. (1994). Response rates and response

content in mail versus face-to-face surveys. *The Public Opinion Quarterly*, *58(3),*

381-399.

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.

Mizes, S. J., Fleece, E. L., & Root, C. (1984). Incentives for increasing return rates: Magnitude levels, response bias, and formats. *Public Opinion Quarterly*, *48,* 784-800.

McDonald, M. P. (2007). The true electorate: A cross-validation of voter registration files and election survey demographics. *The Public Opinion Quarterly*, *71(4),* 588-602.

Murray, Gregg R., Riley, Chris, Scime, Anthony (2009). Pre-Election Polling: Identifying Likely Voters Using Iterative Expert Data Mining. *Public Opinion Quarterly*, 73(1), 159–171.

Nederhof, A. J. (1983). The effects of material incentives in mail surveys: Two studies. *The Public Opinion Quarterly*, *47 (1),* 103-111.

Sudman, S. (1982). Estimating response to follow-ups in mail surveys. *The Public Opinion Quarterly*, *46(4),* 582-584.

Tedin, K. L., Hofstetter, C. R. (1982). The effect of cost and importance factors on the return rate for single and multiple mailings. *The Public Opinion Quarterly*, *46 (1),* 122-128.

Traugott, M. W., Groves, R. M., & Lepkowski, J. M. (1987). Using dual frame designs to reduce nonresponse in telephone surveys. *Public Opinion Quarterly, 51,* 522-539.

Trussell, N., & Lavrakas, P. J. (2004). The influence of incremental increases in token cash incentives on mail survey response: Is there an optimal amount? *Public Opinion Quarterly, 68(3)*, 349-367.

Visser, P. S., Krosnick, J. A., Marquette, J., Curtin, M.(1996). Mail surveys for election

   forecasting? An evaluation of the Columbus dispatch poll. *The Public Opinion*

   Quarterly, *60(2),* 181-227.

Waksberg, J. (1978). Sampling methods for random digit dialling. *Journal of the*
American

   Statistical Association, 73, 40-46.

Worhen, B. R., & Valcarce, R. W. (1985). Relative effectiveness of personalized and

   form covering letters in initial and follow-up mail surveys. *Psychological Reports, 57,*

   735-44.

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	I
Partial (1.2)	1	P
2. Eligible, Non-Interview (2.0)	198	
Refusal and break-off (2.10)	107	R
Refusal (2.11)		
Household-level refusal (2.111)		
Known respondent refusal (2.112)		
Break-off (2.12)		
Non-contact (2.20)	75	NC
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	O
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	UH
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)		UO
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

	Population	Sample Drawn
Voter status - Active	92.4%	91.9%
Zip Code - Watkinsville	52.3%	53.5%
Percinct - Rural Southern	19.1%	16.3%
Municipal name - Incorporated Area	13.5%	14.5%
Race - White	92.1%	91.0%
Gender - Male	47.6%	46.4%
Registration Year 2001 through 2008	54.4%	54.0%
Last Vote Year - new voters/never voted since registration	16.7%	18.3%
Voted Feb 5 2008 Presidential Primary Election- Yes	45.5%	45.0%
Voted July 15 2008 Local/State Primary Election - Yes	32.7%	32.3%
Voted Aug 5 2008 Local/State Primary Runoff - Yes	11.4%	11.1%
Voted Nov 4 2008 General Election - Yes	77.0%	75.8%
Voted Dec 2 2008 Senatorial Runoff - Yes	49.0%	47.0%
Pct. Voters using Republican Ballot - Feb 5 2008	66.4%	69.8%

Pct. Voters using Republican Ballot - July 15 2008	90.8%	91.9%
Pct. Voters using Republican Ballot - Aug 5 2008	86.7%	87.5%
Voted Absentee Feb 5 2008 - Yes	5.9%	4.1%
Voted Absentee July 15 2008 - Yes	8.4%	7.4%
Voted Absentee Aug 5 2008 - Yes	2.5%	2.5%
Voted Absentee Nov 4 2008 - Yes	52.6%	49.1%
Voted Absentee Dec 2 2008 - Yes	18.3%	16.1%
N	21352	800
Sample error		3.5%

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	I
Partial (1.2)		P
2. Eligible, "Non-Interview" (2.0)	2	
Refusal & Break-off (2.10)	1	R
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	NC
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	O
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	UH
Not mailed (3.11)		
Nothing ever returned (3.19)	331	
Unknown if eligible respondent in unit (3.20)	9	UO
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
Voter status - Active	94.2%	93.6%
Zip Code - Watkinsville	52.0%	51.8%
Percinct - Rural Southern	19.1%	20.2%
Municipal name - Incorporated Area	13.4%	13.4%
Race - White	91.5%	93.6%
Gender - Male	47.5%	48.2%
Registration Year 2001 through 2008	56.5%	57.2%
Last Vote Year - new voters/never voted since registration	10.1%	11.4%
Voted Feb 5 2008 Presidential Primary Election- Yes	43.5%	42.4%
Voted July 15 2008 Local/State Primary Election - Yes	31.5%	30.2%
Voted Aug 5 2008 Local/State Primary Runoff - Yes	11.0%	12.6%
Voted Nov 4 2008 General Election - Yes	77.6%	76.4%
Voted Dec 2 2008 Senatorial Runoff - Yes	48.8%	48.2%
Voted Mar 17 2008 SPLOST	6.1%	5.4%
Pct. Voters using Republican Ballot - Feb 5 2008	66.7%	70.8%
Pct. Voters using Republican Ballot - July 15 2008	90.8%	91.4%
Pct. Voters using Republican Ballot - Aug 5 2008	86.5%	87.3%
Voted Absentee Feb 5 2008 - Yes	5.6%	7.2%
Voted Absentee July 15 2008 - Yes	8.0%	7.4%
Voted Absentee Aug 5 2008 - Yes	2.4%	2.4%
Voted Absentee Nov 4 2008 - Yes	53.0%	48.8%
Voted Absentee Dec 2 2008 - Yes	18.1%	17.8%
Voted Absentee Mar 17 2008 - Yes	1.9%	1.8%
N	22094	500
Sample error		4.4%

Table 5. Telephone numbers found distributed by mailing outcome

Telephone numbers	Return Received	Returned Refused
Found	92	1
Not found	4	0
Address matched voter but "unlisted" in white pages	8	0
Address did not match voter	20	0
Total	124	1

Sample of	Telephone Sample Inte	Telephone Sample Plus	Confirmed Improperly Registered
92.6%	95.3%	98.5%	95.0%
53.1%	53.4%	52.0%	53.7%
16.1%	16.7%	17.0%	19.0%
14.1%	13.2%	11.0%	11.1%
91.5%	93.1%	94.0%	92.6%
46.2%	48.0%	43.0%	46.7%
53.2%	50.9%	52.0%	56.2%
17.5%	13.6%	11.5%	14.5%
46.9%	52.2%	60.0%	54.5%
33.6%	38.0%	45.0%	39.7%
11.5%	12.5%	15.0%	13.2%
76.7%	82.4%	91.5%	86.0%
48.1%	53.0%	58.5%	55.0%
70.9%	72.9%	69.2%	68.1%

93.2%	94.8%	96.7%	95.7%
88.7%	91.2%	93.3%	90.9%
4.1%	4.3%	5.0%	5.4%
7.4%	7.4%	7.5%	6.6%
2.6%	2.4%	1.5%	1.2%
49.6%	54.2%	61.5%	58.3%
16.5%	18.5%	18.0%	17.8%
703	552	200	242
3.7%	4.2%	6.9%	6.3%

Eligible Re	Telephone	Mail Sample	Mail Sample Plus Confirmed	Improperly Registered
96.5%	99.4%	100.0%	90.2%	
50.3%	53.7%	47.6%	52.8%	
19.7%	20.8%	16.1%	18.4%	
13.2%	12.8%	15.3%	15.3%	
93.1%	94.6%	97.6%	98.2%	
48.2%	50.2%	48.4%	48.5%	
56.8%	48.9%	44.4%	48.5%	
10.4%	6.4%	2.4%	7.4%	
44.9%	53.7%	54.8%	44.8%	
32.1%	39.0%	41.9%	33.7%	
13.4%	17.6%	21.8%	17.2%	
80.7%	87.9%	94.4%	77.9%	
51.6%	61.3%	69.4%	54.6%	
5.9%	7.3%	12.9%	9.8%	
70.0%	70.2%	66.2%	68.5%	
91.2%	90.2%	94.2%	94.5%	
87.1%	87.3%	92.6%	92.9%	
7.8%	9.9%	14.5%	11.0%	
7.8%	9.6%	13.7%	11.0%	
2.6%	3.5%	4.8%	3.7%	
52.3%	58.8%	66.1%	52.1%	
19.1%	22.7%	30.6%	23.9%	
2.0%	2.9%	4.8%	3.7%	
461	313	124	163	
4.6%	5.5%	8.8%	7.7%	

Returned to Sender	No Response	Total
8	220	321
7	15	26
1	29	38
23	72	115
39	336	500