

## **Using a Dual-Frame Sample Design to Increase the Efficiency of Reaching Population Subgroups in a Telephone Survey**

Douglas B. Currivan, Ph.D.  
David J. Roe, M.A.

*RTI International\**

May 6, 2004

*This paper was prepared for the 59th annual meeting of the American Association for Public Opinion Research, Phoenix, Arizona, May 13-16. The survey data collection used as the basis of this research was sponsored by the American Legacy Foundation. In addition to the American Legacy Foundation, the authors thank project director Matthew Farrelly of RTI International for supporting this research. We also thank sampling statistician Don Akin of RTI International and Steven Dintino of Marketing Systems Group for assistance with sample frame specifications and sample selection procedures. Discovery Research Group collected the survey data.*

\* *RTI International is a trade name of Research Triangle Institute.*

## **Abstract**

The effort and cost required to reach households and complete interviews in random-digit dialing (RDD) telephone surveys has increased over the past several years. Effort and cost are even greater in RDD surveys when the sample design specifies screening the population for specific subgroups, such as age or ethnic groups. When the probability of reaching respondents in a specific subgroup is sufficiently low, the cost of using a standard RDD approach can be prohibitive. An alternative strategy is to supplement RDD numbers with numbers selected from directory listings. Using listed numbers can significantly improve the probability of reaching eligible respondents and thereby lower the effort and cost of screening households and completing interviews. Still, directory-listed sample frames have the important shortcoming of excluding the growing number of households that do not currently have listed numbers. Furthermore, additional information on the listed sample frame on how likely household members belong to a particular subgroup may be of limited accuracy. The potential result of these shortcomings of listed sample frames is bias in survey estimates. The goal of this research is to better understand the costs and benefits of supplementing an RDD sample with a listed sample in a national survey that targets respondents in certain age and ethnic groups. The analysis compares the two sampling frames on both key outcomes of the data collection effort and on substantive results among completed interviews. Key data collection outcomes include rates of working and residential numbers, eligibility rates, and completion rates. Among completed interviews, we examine differences in the targeted demographics (age and ethnicity), as well as other demographic characteristics and substantive indicators. This research will provide some evidence on the potential of dual-frame designs to provide accurate data on population subgroups with less effort and cost compared to RDD methods.

## **Using a Dual-Frame Sample Design to Increase the Efficiency of Reaching Population Subgroups in a Telephone Survey**

The difficulty in screening households and completing interviews using random-digit dialing (RDD) survey methods has increased over the past several years. Due to the increasing use of technology that allows households to avoid answering the telephone (such as answering machines, called ID, and call management systems) and increasing reluctance of households to participate in surveys when contacted, the effort required to complete RDD surveys has increased greatly over the past 15 or so years. For example, Curtin, Presser, and Singer (2000) report that the average number of calls required to complete interviews and the number of cases requiring refusal conversion on the Survey of Consumer Attitudes doubled between 1979 and 1996.

The effort and cost to complete interviews are even greater in RDD surveys when the sample design focuses on specific subgroups within the population, such as particular age or ethnic groups. In addition to the existing challenges of RDD surveys, the effort required to screen households only increases the cost of conducting such surveys. When the probability of reaching respondents in a specific subgroup is sufficiently low, the cost of using an RDD approach can be prohibitive.

One alternative strategy to relying solely on an RDD sample for surveys focused on specific subgroups within the population is to supplement the RDD numbers with numbers selected from directory listings. Using listed numbers significantly improves the probability of reaching eligible respondents and thereby lowers the effort and cost of screening households and completing interviews. Of course, directory-listed sample frames have the important shortcoming of excluding the growing number of households that do not currently have listed numbers. This shortcoming could limit the advantages associated with using listed numbers.

In order to assess the impact of combining listed numbers with RDD numbers on data collection efficiency and survey results in a study that targets

multiple population subgroups, we use data from a nationally-representative survey of smoking attitudes and behaviors among teens age 12-17 and young adults age 18 to 24. This survey employed a dual-frame design with approximately 50% listed telephone numbers and 50% RDD telephone numbers. Our analysis compares the two sampling frames on both key outcomes of the data collection effort and on substantive results among completed interviews. Key data collection outcomes include rates of working and residential numbers, eligibility rates, and completion rates. Among completed interviews, we examine differences in the targeted demographics (especially age and ethnicity), as well as other demographic characteristics and substantive indicators. This research will provide some preliminary evidence on the potential of dual-frame designs to provide accurate data on population subgroups with less effort and cost compared to RDD methods.

### **Major Advantages of Using Directory-Listed Sample Frames**

The most important advantage of using a sample frame of directory listed telephone numbers is that the sampled numbers are quite likely to be associated with residential households. For most RDD surveys, a major challenge is screening telephone numbers to determine whether they are connected to households. This is especially problematic in urban population centers where RDD sample frames typically produce large numbers of non-working and non-residential numbers. Directory listed frames, on the other hand, have the potential to greatly reduce the initial screening effort relative to RDD frames by increasing the proportion of working, residential numbers in the sample.

The efficiency advantages of list frames over RDD frames are likely to be even greater when the sample design focuses on particular subgroups within the population. List frames can greatly increase the incidence rate of targeted subgroups in two ways. First, the elimination of a greater proportion of non-working and non-residential numbers will generally increase the likelihood of reaching a household with members of the subgroup compared to RDD surveys. Second, listed numbers can be matched against secondary databases to provide

information on demographic characteristics of household members. Information such as age and ethnicity of household members can be used to pre-screen sampled numbers to improve the likelihood that the sample will reach household members with desired characteristics.

A final advantage of listed numbers is that they are more likely to provide accurate names and addresses of household members, or at least the head of the household. Such information can be useful for advance mailings and limit the disadvantages of “cold calling” households. When using an RDD sample frame, many sampled numbers do not result in a name and address match, or the name and address match obtained is not accurate. This problem severely limits the effectiveness of lead letters in RDD surveys. With listed samples, lead letters are likely to be considerably more effective in reaching potential respondents. Overall, listed numbers are more likely to facilitate mail contact with potential respondents at any stage of the study, compared to RDD numbers.

### **Potential Problems in Using Directory-Listed Sample Frames**

The greatest limitation of directory-listed sample frames is that such frames obviously omit households without listed numbers. Some households with telephone service choose not to list their number, while others do not currently have a listed number because they have recently moved or otherwise recently begun telephone service. This omission has the potential to introduce the most serious source of error in sample frames, excluding elements that are actually part of the target population (Currivan, 2003; Edwards, Brick, and Flores-Cervantes, 2003). For most surveys, households without listed numbers should be just as likely to contain eligible members as households with listed numbers. The potential for introducing bias by excluding listed households in surveys is likely to be considerable, especially since the number of households without listed telephone numbers has steadily increased in recent years (Tucker, Lepkowski, and Piekarski, 2002).

Differences between households with listed versus unlisted numbers have been demonstrated through a number of analyses. In an extensive comparison

between about 33,000 households with directory-listed numbers and over 21,000 households without listed numbers, Piekarski (1989) found several important differences:

- Listed numbers over-represented established households and under-represented recent movers
- Unlisted households tended to include a disproportionate number of unmarried householders
- Younger females in one-person households were over-represented among unlisted numbers
- Retired householders appeared to be over-represented in the listed sample and employed householders over-represented in the unlisted sample
- Residents in unlisted households were significantly younger than those in listed households

Genesys has recently compared listed and non-listed numbers and found similar results. Households with listed numbers tend towards higher income, older, and better-educated homeowners (Genesys, 2003). These kinds of demographic contrasts led Piekarski (1989) to conclude that as rates of unlisted telephone numbers rise, differences between listed and unlisted households are significant enough to produce quantifiable coverage bias with directory-listed sample frames.

One further potential limitation of listed sample frames is that additional information about household members that might be used for sampling purposes may be of limited accuracy. When information on whether household members might belong to a particular subpopulation is used to inform sample selection procedures, the assumption is that this information increases the probability of reaching eligible members of the subgroup(s). If this information is inaccurate, using listed sample numbers may not improve data collection efficiency and may possibly introduce unanticipated biases into the sample. As Edwards, Brick, and Flores-Cervantes (2003) point out, lists must actually contain members of targeted subgroups, or substantial screening efforts will be incurred and

efficiency gains compared to RDD surveys may therefore be lost. Furthermore, inaccurate information might result in a sample frame that includes an inappropriate number of households with members that are not part of the subpopulation of interest, such as those who are older or younger than the targeted age group. The accuracy of information used in selecting listed numbers is another potential source of sampling error.

### **Research Questions**

The potential limitations of using a directory-listed sample frame suggest that relying solely on listed numbers would pose a serious threat to the validity of most survey estimates. On the other hand, adding a set of listed numbers to a sample of RDD numbers has the potential to significantly improve the efficiency of telephone data collection in reaching subpopulations of interest. This kind of dual-frame sample design offers the possibilities of both increasing data collection efficiency and minimizing sample bias.

To assess the viability of dual-frame designs for reaching particular subgroups in the population, we seek to answer three research questions:

1. Compared to RDD numbers, how much more accurate were directory-listed numbers in reaching households with members of targeted age and race/ethnic subgroups?
2. To what extent did adding listed numbers to RDD numbers improve data collection efficiency by increasing the rates of (1) working and residential numbers in the sample, (2) eligible respondents in sampled households, and (3) completing interviews with eligible respondents?
3. Are there any differences in either demographic characteristics or substantive indicators (such as smoking behaviors) between households sampled from listed versus RDD numbers that suggest bias in the survey estimates?

We use data from a nationally-representative survey of smoking attitudes and behaviors among teens age 12 to 17 and young adults age 18 to 24 that employs a dual-frame sample design to provide answers to these questions.

## **Research Methods**

### **1. Survey Design**

The Legacy Media Tracking Survey (LMTS) was designed to collect data about the role tobacco advertising plays in smoking attitudes and behaviors among teens and young adults. The target population was young people age 12 to 24, which included a nationally-representative sample and oversamples in four specific states (California, Florida, Minnesota, and Mississippi). Computer-assisted telephone interviewing (CATI) techniques were used to complete the data collection. Each sampled telephone number was screened to determine whether it was a residential household and whether any young people age 12-24 lived in the household. For respondents under age 18, parental consent was required before enlisting participation of the children in the survey. All interviews were completed in English only, although the introduction, screening, and parental consent text was translated into Spanish to facilitate communication with Spanish-speaking parents.

Nine waves of this national telephone survey have now been completed, with each wave about six to eight months apart. RTI International coordinated the ninth wave of this survey, which had a field period of November 2003 through January 2004. The LMTS-9 survey called for 5,000 completed interviews; 4,993 were actually completed. The overall response rate for the survey was 30%, using AAPOR response rate 4.

An important feature of the survey is that the sample design specifies interviewing targets for multiple age and racial/ethnic groups. Table 1 presents the sample targets and final results for each of the key subpopulations specified by the survey design. Most sample targets were reached (or exceeded), although the interviewing results were significantly short of the goals for Asian/Pacific Islander and White respondents.

**Table 1.** Subgroup Targets and Final Results for the LMTS-9 Survey

<b>Subgroups of Interest</b>	<b>Interviewing Targets</b>	<b>Completed Interviews</b>
Age: 12-14 years old	1,750	1,778
15-17 years old	1,750	1,685
18-24 years old	1,500	1,472
Race/Ethnicity: Hispanic	750	849
African-American	750	873
Asian/Pacific Islander	500	341
Native American	0	68
White	3,000	2,740

**Note:** Completed interviews for each subgroup do not add up to the total number of completed interviews due to missing data (Refused to answer).

## **2. Sample Frames**

The need to reach interviewing targets among multiple subpopulations defined by age and race/ethnicity led the LMTS researchers to seek an alternative to relying solely upon list-assisted RDD methods. Instead, the LMTS employed a dual-frame design in which approximately half of the telephone numbers were generated through a list-assisted RDD frame and half were selected from directory-listed telephone numbers. Both frames were stratified by the five target geographic areas – the states of California, Florida, Minnesota, and Mississippi and the remainder of the United States.

The RDD frame generated telephone numbers by using the list-assisted RDD sampling system created by Genesys, which is provided by Marketing Systems Group. The Genesys system identifies all residential clusters of 100 telephone numbers (area code + exchange + first two digits of phone number) that have at least one published residential number. These clusters are updated quarterly. The clusters then form the sample frame for selection of final sample

telephone numbers. Since all possible clusters are used to create a sample frame, the Genesys provides an advantage, as the final sample is not clustered as in traditional Mitofsky-Waksberg RDD samples.<sup>1</sup>

The listed sample frame was built primarily from White Page telephone directories and also provided by Marketing Systems Group. For each listing in the frame, name (as listed in phone book), phone number, address (where listed), and phone book identification code (book from which data originated) are compiled. The final component in this process is the assignment of geographic codes to each record based on the zip code provided with the address. This allows assignment of the household to its appropriate county, which is the building block for obtaining all other geographic information.

An important advantage of listed numbers is that the basic information on each record can be “enhanced” to include demographic data about the household. Secondary data sources such as Census data, state automobile registrations, drivers’ license data, voter registrations, birth records, and proprietary data sources can be used to supplement the records. Most of the information on households that comes directly from these secondary data sources is fairly accurate. Other pieces of information, like income, are often modeled and therefore, represent estimates. The result is that the listed frame records contain a variety of other information about the household including age/gender of family members, income, dwelling unit size, etc. For the LMTS-9 sample, the listed numbers were drawn using enhanced information about the likely age and ethnicity of household members.

### **3. Analysis Plan**

To answer our first research question, we conducted analysis only among the 4,993 interviews completed. Of these, 1,023 interviews were completed among cases from the RDD sample frame and 3,970 were completed among cases from the directory-listed frame. This analysis crosstabulated the

---

<sup>1</sup> For more detailed discussion bias and efficiency associated with list-assisted RDD sampling methods compared to other RDD sampling techniques, see Brick, Waksberg, Kulp, and Starer (1995) and Tucker, Lepkowski, and Piekarski (2002).

proportion of respondents in each of the targeted age and racial/ethnic subgroups between RDD and listed numbers in order to compare how successful cases from each frame were in reaching targeted subgroups.

To answer our second research question, we started with the entire set of 64,584 sampled telephone numbers to analyze several indicators of data collection efficiency. Of all sampled numbers, 32,195 were drawn from an RDD frame and 32,389 were drawn from a directory listed frame. Our analysis focused on key sample dispositions for the RDD and listed numbers, including the proportion of numbers from each sample frame that resulted in:

- Working numbers: those numbers not determined to be non-working (disconnected) numbers among all sampled numbers
- Residential numbers: those numbers not determined to be non-residential (business, government, or unknown fax/data lines) among all working numbers
- Eligible person in household: those numbers among all residential numbers that resulted in at least one eligible household member being identified
- Completed interviews: those numbers that resulted in a completed interview among all numbers with eligible respondents
- Final refusals: those numbers that resulted in a final code of refusal (household or eligible respondent) among all numbers with eligible respondents
- Other non-interviews: those numbers that did not result in either a completed interview or final refusal among all numbers with eligible respondents

To address our second research question, we performed two sets of comparisons between the numbers from the RDD and listed frames among only the 4,993 completed interviews. The first set of comparisons involved the following demographic characteristics of youth participants or their household:

- age
- race/ethnicity

- respondent born in U.S. versus other country
- respondent lives with one or both parents (12-17 year olds only)
- type of residence (18-24 year olds only)
- respondent currently employed for pay
- respondent has cell phone

A second set of comparisons among completed interviews involved key survey indicators of smoking behavior for youth respondents and other household members. Smoking attitudes and behaviors were the substantive focus of the LMTS-9. The specific survey items included in this analysis are described in Table 2.

We tabulated all of the sample dispositions and survey indicators across the two sample frames and performed independent sample t-tests on each of the proportions or means compared. Since we were comparing variables across independent samples, we performed all analyses using unweighted data. The conventional alpha level of  $p < .05$  was used to determine statistical significance for all analytic procedures.

## **Results**

### ***1. Research Question 1***

Our first research question focused on the accuracy of directory-listed numbers versus RDD numbers in reaching households with members of the targeted age and race/ethnic subgroups. Table 3 provides the sample targets for the three age groups and four race/ethnic groups, the overall proportion of completed interviews in each of these subgroups, and the proportion of interviews in each subgroup crosstabulated by whether the numbers were RDD or listed. The total number of completed interviews was 4,993, 1,023 of which were completed with numbers from the RDD sample and 3,970 from the listed sample. AAPOR 4 response rate for the listed portion of the sample was 33%, while the response rate for the RDD portion of the sample was 26%.

Looking first at age, the proportion of completed interviews in each age group was significantly different for the interviews completed from RDD numbers versus listed numbers. Compared to the listed numbers, the RDD numbers were

**Table 2.** Key Smoking Indicators from the LMTS-9 Survey

Smoking Indicator	Survey Item	Analytic Variable
Respondent has ever tried cigarettes	Have you ever tried cigarette smoking, even 1 or 2 puffs? 1. Yes 2. No	Percent of respondents answering 1 (“yes”)
Respondent has smoked 1 pack or more of cigarettes in lifetime	About how many cigarettes have you smoked in your entire life? 1. 1 or more puffs, but never a whole cigarette 2. 1 cigarette 3. 2 to 5 cigarettes 4. 6 to 15 cigarettes or about half a pack 5. 16 to 25 cigarettes or about a pack 6. 26 to 99 cigarettes or more than a pack but less than 5 packs 7. 5 packs or more	Percent of respondents answering 5, 6, or 7 (about 1 pack or more)
Respondent was ever a regular smoker	Have you ever smoked at least one cigarette every day for 30 days? 1. Yes 2. No	Percent of respondents answering 1 (“yes”)
Respondent will likely smoke in next year	Do you think you will smoke a cigarette at anytime during the next year? 1. Definitely yes 2. Probably yes 3. Probably not 4. Definitely not 5. No opinion	Percent of respondents answering 1 or 2 (“definitely” or “probably” yes)
Smoking in respondent’s peer group	How many of your four closest friends smoke cigarettes? 1. None 2. One 3. Two 4. Three 5. Four 6. Not sure	Percent of respondents answering 2, 3, 4, or 5 (one or more smokers)
Respondent exposure to others’ smoke in past week	During the past 7 days, on how many days were you in the same room with someone who was smoking cigarettes?	Average number of days for all respondents (1 to 7)
Presence of other smoker(s) in household	Other than yourself, does anyone who lives in your home smoke cigarettes now? 1. Yes 2. No	Percent of respondents answering 1 (“yes”)

**Table 3.** Age Group and Racial/Ethnic Group Proportions for Completed Interviews across RDD and Listed Numbers in the LMTS-9 Sample

<b>Subgroups of Interest</b>	<b>Subgroup Target</b>	<b>All Interviews (n = 4,993)</b>	<b>RDD Interviews (n = 1,023)</b>	<b>Listed Interviews (n = 3,970)</b>
<i>Age Groups:</i>				
12-14*	35%	36%	28%	38%
15-17*	35%	34%	22%	37%
18-24*	30%	30%	50%	25%
<i>Race/Ethnic Groups:</i>				
Hispanic*	15%	17%	10%	19%
African-American*	15%	18%	39%	12%
Asian/Pacific Islander*	10%	7%	2%	8%
White*	60%	56%	46%	58%

\* Difference between RDD sample and listed sample interviews is statistically significant at  $p < .05$  based on independent samples t-tests

much less likely to reach households with youths age 12 to 17 and much more likely to reach households with 18 to 24 year olds. Overall, the distribution of listed numbers across age categories more closely mirrored the age subgroup targets than the RDD numbers did.

Under race/ethnic groups, all differences between completed interviews from RDD versus listed numbers were statistically significant. Compared to the listed numbers, the RDD numbers produced fewer interviews with Hispanic, Asian/Pacific Islander, and White respondents and considerably more interviews with African-American respondents. Again, the overall distribution of listed numbers across race/ethnic categories more closely mirrored the race/ethnic subgroup targets than the RDD numbers did.

## **2. Research Question 2**

In order to assess the data collection efficiency of RDD versus listed numbers, we went back to all sampled telephone numbers used in data collection. Table 4 presents key sample dispositions crosstabulated by numbers

from the RDD and listed sample frames. The first indicator is the proportion of working numbers, defined as those numbers not determined to be non-working (disconnected) numbers among all sampled numbers. Although a majority of both RDD and listed numbers were not classified as non-working, the proportion of working numbers from the listed sample (84%) was over 10% higher than the proportion of RDD numbers (72%). Similarly, the proportion of residential numbers (those numbers not determined to be non-residential among all working numbers) was significantly higher among listed numbers (78%) than among RDD numbers (62%).

The overall eligibility rate is another critical indicator of data collection efficiency in surveys with subgroup targets. We defined eligibility as those numbers that resulted in the identification, through screening, of at least one eligible household member, among all residential numbers in the sample. Among RDD numbers the eligibility rate was only 22%, but the rate among listed numbers was more than double at 51%.

The final step of the screening and interviewing process is completing interviews among all numbers with eligible respondents. Table 4 provides the proportions of completed interviews, final refusals, and other non-interviews

**Table 4.** Final Sample Dispositions for All Sampled RDD and Listed Numbers in the LMTS-9 Sample

Final Sample Disposition	RDD Numbers	Listed Numbers
Working numbers*	72.1%	83.7%
Residential numbers*	61.5%	77.6%
Eligible teen(s) in household*	21.7%	51.4%
Completed interviews*	44.6%	47.1%
Final refusals	37.5%	37.4%
Other non-interviews*	17.9%	15.4%

\* Difference between RDD sample and listed sample interviews is statistically significant at  $p < .05$  based on independent samples t-tests

crosstabulated by RDD versus listed numbers. At this stage of data collection, differences between RDD and listed numbers were much smaller. The proportion of refusals among eligible households was virtually identical. The listed numbers resulted in a greater proportion of completed interviews and a lower proportion of other non-interviews that were statistically significant. In practical terms, these differences seem small. The statistical significance is likely due to the relatively large sample size of 10,716 numbers with eligible household members.

### **3. *Research Question 3***

Our final research goal was to determine whether differences in specific survey responses collected from RDD versus listed numbers differ in ways that suggest overall estimates may be biased. Table 5 compares responses to various demographic items from interviews with RDD versus listed numbers, while Table 6 compares responses on smoking indicators from interviews with RDD versus listed numbers. Although all comparisons between the survey estimates from RDD versus listed numbers are statistically significant, a few differences do not appear to represent meaningful differences.

Looking at demographic characteristics, the listed numbers produced significantly more interviews with youths age 12 to 17 but significantly fewer interviews with youths whose racial/ethnic background was not White or was Hispanic. The results for both demographic characteristics are important, since age and race/ethnic groups were the two targets of the sampled listed numbers. The RDD numbers resulted in fewer respondents age 12 to 17 than needed and more non-White respondent than needed to meet the sample targets.

Household characteristics also differed greatly between the RDD and listed interviews. The RDD numbers produced significantly fewer 12 to 17 year olds who live with both parents and significantly more 18 to 24 year olds who have their own house or apartment. These findings follow expectations, since listed numbers tend to over-represent established households and married householders (Piekarski, 1989).

**Table 5.** Demographic Characteristics for Completed Interviews across RDD and Listed Numbers in the LMTS-9 Sample

<b>Demographic Characteristic</b>	<b>RDD Interviews (n = 1,023)</b>	<b>Listed Interviews (n = 3,970)</b>
Age 12-17 *	49.9%	75.4%
Race/ethnicity other than White*	53.6%	42.0%
Lives with both parents* (12-17 year olds only)	60.0%	80.9%
Lives in own home* (18-24 year olds only)	42.9%	18.4%
Respondent currently employed for pay*	42.3%	32.6%
Respondent has cell phone*	43.2%	40.7%

\* Difference between RDD sample and listed sample interviews is statistically significant at  $p < .05$  based on independent samples t-tests

Two final demographic items compared in Table 5 are having a paid job and having a cell phone. Respondents from the RDD sample frame were significantly more likely to be currently employed than those from the listed frame. This is an interesting difference, and not easy to understand. Although respondents from the RDD sample were also significantly more likely to have a cell phone than those from the listed sample, the overall difference was only 2.5% and possibly not a meaningful difference.

Table 6 presents comparisons between the RDD and listed interviews on several key indicators of smoking behavior among respondents or members of their households. Respondents from the RDD sample were more likely to report all types of smoking behaviors than those from the listed sample, both for themselves and others in their household. One significant difference that might not be meaningful is the average number of days in the past week that the respondent was exposed to others' smoke. Respondents from the RDD sample reported exposure of just over one-half day more than respondents from the listed sample. All other differences appear to be robust and meaningful.

**Table 6.** Smoking Indicators for Completed Interviews across RDD and Listed Numbers in the LMTS-9 Sample

<b>Smoking Indicator</b>	<b>RDD Interviews (n = 1,023)</b>	<b>Listed Interviews (n = 3,970)</b>
Respondent has ever tried cigarettes*	41.8%	27.6%
Respondent has smoked 1 pack or more of cigarettes in lifetime*	54.1%	44.1%
Respondent was ever a regular smoker*	42.1%	28.1%
Respondent will likely smoke in the next year*	4.6%	2.5%
Smoking in respondent's peer group*	48.8%	36.6%
Respondent exposure to others' smoke in past week (mean number of days)*	4.12	3.51
Other smoker(s) in household*	35.0%	26.5%

\* Difference between RDD sample and listed sample interviews is statistically significant at  $p < .05$  based on independent samples t-tests

## **Discussion**

Our goal was to assess the impact of combining listed numbers with RDD numbers on data collection efficiency and survey results in a study that targets multiple population subgroups. Using data from a nationally-representative survey of smoking attitudes and behaviors among youths that employed a dual-frame design, our analysis compared numbers from an RDD frame with those from a listed frame. We compared the two sampling frames on both key outcomes of the data collection effort and on substantive results among completed interviews, and found many significant differences.

First, the listed numbers were much more effective in meeting the sample targets for respondents by age groups and race/ethnic groups. Analysis of sample dispositions confirmed that the listed numbers were significantly more efficient than RDD numbers in reaching households with eligible members. This

resulted both because the listed numbers included more working, residential numbers and also because these numbers found more eligible members among the working, residential numbers.

Second, respondents from the RDD sample were more likely to report all types of smoking behaviors than those from the listed sample, both for themselves and others in their household.

Based on the results presented, there is little question that the listed numbers analyzed here are more effective than their RDD counterparts when it comes to reaching the sample targets for respondents by age and race/ethnic groups. With one exception, final refusals, all of the differences between the two groups presented herein were statistically significant. Despite the success in reaching these targets, this preliminary investigation into the differences between listed and RDD sample does suggest that researchers use listed sample with caution, as the significant differences observed on all fronts in this research suggest the potential for bias.

Future efforts should be made to continue to develop a clear understanding of the biases that could be involved in using listed sample. These potential biases could affect estimates that researchers may hope to make as they normally would with a 100% RDD sample. Researchers who desire large level estimates from survey data should perhaps use a smaller proportion of listed sample, or none at all. If efficiency and not estimates are what researchers desire, listed sample could be a useful tool.

In addition, future research should search investigate how using listed sample affects the effort and cost involved in conducting surveys. Comparisons between listed and RDD numbers in terms of call counts (number of calls required to finalize a case) and hours per complete, two points of information not available for this research, and factors related to the cost of sample, should be used in combination to provide a cost/effort analysis of the use of listed vs. RDD sample.

## References

- Brick, J. Michael, Joseph Waksberg, Dale Kulp, and Amy Starer. 1995. "Bias in List-Assisted Telephone Samples." *Public Opinion Quarterly*, 59: 218-235.
- Currivan, Douglas B. 2003. "Sampling Frame." In Michael Lewis-Beck, Alan Bryman, and Tim Futing Liao (Eds.), *The Sage Encyclopedia of Social Science Research Methods*. Thousand Oaks, CA: Sage.
- Curtin, Richard, Stanley Presser, and Eleanor Singer. 2000. "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly* 64:413-428.
- Edwards, W. Sherman, J. Michael Brick, and Ismael-Flores-Cervantes. 2003. "Sampling Race and Ethnic Groups in RDD Surveys." Paper presented at the Joint Statistical Meetings of the American Statistical Association, Section on Survey Research Methods.
- GENESYS. 2003. Nonresponse and Practical Sampling Issues. Stamford, CT: Marketing Systems Group.
- Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert M. Groves, and Stanley Presser. 2000. "Consequences of Reducing Nonresponse in a National Telephone Survey." *Public Opinion Quarterly* 64:125-148.
- Piekarski, Linda B. 1989. "Choosing between Directory Listed and Random Digit Sampling in Light of New Demographic Findings." Paper presented at the *American Association of Public Opinion Research* annual conference.
- Tucker, Clyde, James M. Lepkowksi, and Linda Piekarski. 2002. "The Current Efficiency of List-Assisted Telephone Sampling Designs." *Public Opinion Quarterly* 66:321-338.