

*turning knowledge into practice*

# **Using Dual-Frame Sample Designs to Increase the Efficiency of Reaching General Populations and Population Subgroups in Telephone Surveys**

David J. Roe, M.A.  
Douglas B. Currivan, Ph.D.  
*January 12, 2006*



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

3040 Cornwallis Road  
Phone 919-316-3598

■ P.O. Box 12194

■ Research Triangle Park, North Carolina, USA 27709  
Fax 919-316-3866

■ e-mail droe@rti.org

[www.rti.org](http://www.rti.org)

# Overview

- Challenges Facing RDD Studies
- Advantages and Potential Problems Associated with Directory Listed Numbers
- Dual Frame Approach
- NY-ATS & LMTS-9 Background
- Analysis of Differences Between Listed and RDD Cases
  - Age, Race/Ethnicity, other Demographics & Smoking Indicators
  - General Population vs. Targeted Subgroups
  - Unweighted vs. Weighted
- Conclusions and Suggestions for Future Research

# Challenges Facing RDD Studies

- The difficulty in screening households and completing interviews using random-digit dialing (RDD) survey methods has increased over the past several years
- Increasing use of technology that allows households to avoid answering the telephone
  - Answering machines, caller ID, and call management systems
- Increasing reluctance of households to participate in surveys when contacted

# RDD Effort and Cost

- Effort and cost to complete interviews are even greater when the sample design focuses on specific subgroups within the population, such as particular age or ethnic groups
- 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

# Major Advantages of Directory Listed Numbers

- More likely working and residential
- More efficient for reaching subgroups
- Better information for mail contact

# Potential Problems with Directory Listed Numbers

- Exclude households without listed numbers
- Potential for adding bias to survey estimates
- Information about household demographics might not be accurate

# Dual Frame Approach

- Combine listed numbers with RDD numbers
- Use information on listed numbers to target subgroups
- Goal: Improve efficiency while minimizing potential bias

# Research Questions

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

# Research Questions, Continued

4. Are these differences more prevalent in a survey of a targeted population subgroup when compared to a more “general” population?
5. To what extent does weighting adjust for potential biases between listed and RDD sample?

# Analysis Plan

- We analyze data from two surveys that employed a dual-frame sample design
  - A statewide survey of smoking attitudes and behaviors among adults 18 and over
  - A nationally-representative survey of smoking attitudes and behaviors among teens age 12-17 and young adults age 18 to 24
- Provide preliminary evidence on the impact of adding listed numbers to an RDD frame on two different target populations:
  - Data collection efficiency and sample representativeness
  - Assess the relative impact on efficiency and bias of the dual-frame approach on these two different populations

# The New York Adult Tobacco Survey (NY-ATS)

- Collects data on adults' beliefs on and experiences with tobacco use, in order to inform statewide public health programs
- Target population is permanent residents of the state of New York age 18 and over
- Follows a quarterly data collection schedule, beginning in the third quarter of 2003, with a goal of 2,000 interviews completed each quarter
- For our analysis, we use data from the fourth quarter of 2003 of the NY-ATS because the field period aligns closely with the second survey we examine

## Legacy Media Tracking Survey (LMTS)

- Collects data about the role tobacco advertising plays in smoking attitudes and behaviors among teens and young adults
- Target population was young people age 12 to 24
- Nine waves of this national telephone survey have been completed, with each wave about six to eight months apart
- Field period of November 2003 through January 2004
- Sample design specified interviewing targets for multiple age and racial/ethnic groups

# RDD Frames

- Generated numbers using the list-assisted RDD sampling system created by Genesys
- 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
  - Updated quarterly
- Clusters then form sample frame for selection of final sample telephone numbers
- Advantage:
  - All possible clusters are used to create a sample frame, the final sample is not clustered as in traditional Mitofsky-Waksberg RDD samples

# Listed Frames

- The listed sample frames were built primarily from White Page telephone directories
  - Also provided by Marketing Systems Group
- Name (as listed in phone book), phone number, address (where listed), and phone book identification code (book from which data originated) are compiled
- Assignment of geographic codes to each record based on the zip code provided with the address
  - Allows assignment of the household to the appropriate county, which is the building block for all other geographic areas

# Listed Frames, Continued

- Advantages:
  - Basic information on each record can be “enhanced” to include demographic data about the household
  - Secondary sources such as Census data, state automobile registrations, drivers’ license data, voter registrations, birth records, and proprietary data sources are used to supplement the records

# Listed Frames, Continued

- Result:
  - 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.
  - Focused on numbers with a higher probability of containing a 12-17 year old

## Age Group and Racial/Ethnic Group Proportions for Completed Interviews across RDD and Listed Numbers

| Subgroups of Interest |                         | Subgroup Target | All Interviews<br>(n = 4,993) | RDD Interviews<br>(n = 1,023) | Listed Interviews<br>(n = 3,970) |
|-----------------------|-------------------------|-----------------|-------------------------------|-------------------------------|----------------------------------|
| Age:                  | 12-14 years old*        | 35%             | 36%                           | 28%                           | 38%                              |
|                       | 15-17 years old*        | 35%             | 34%                           | 22%                           | 37%                              |
|                       | 18-24 years old*        | 30%             | 30%                           | 50%                           | 25%                              |
| Race/<br>Ethnicity:   | 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  $p < .05$

Percentages for race/ethnicity subgroup do not add up 100% due to missing data.

# Final Sample Dispositions for All Sampled RDD and Listed Numbers in The NY-ATS and LMTS-9 Samples

| Final Sample Disposition                     | NY-ATS      |                | LMTS-9      |                |
|--|-------------|----------------|-------------|----------------|
|  | RDD Numbers | Listed Numbers | RDD Numbers | Listed Numbers |
| Working numbers <sup>1,2</sup>               | 86.8%       | 89.3%          | 72.1%       | 83.7%          |
| Residential numbers <sup>1, 2</sup>          | 74.8%       | 93.8%          | 61.5%       | 77.6%          |
| Eligible person(s) in household <sup>2</sup> | 94.7%       | 95.1%          | 21.7%       | 51.4%          |
| Completed interviews <sup>1,2</sup>          | 21.3%       | 24.4%          | 44.6%       | 47.1%          |
| Final refusals                               | 51.1%       | 52.7%          | 37.5%       | 37.4%          |
| Other non-interviews <sup>1,2</sup>          | 27.7%       | 23%            | 17.9%       | 15.4%          |

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

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

# Demographic Characteristics for Completed Interviews across RDD and Listed Numbers in the NY-ATS Sample

| Demographic Characteristic                      | Unweighted                |                                | Weighted       |                   |
|---|---------------------------|--------------------------------|----------------|-------------------|
|   | RDD Interviews<br>(n=777) | Listed Interviews<br>(n=1,286) | RDD Interviews | Listed Interviews |
| One adult in household <sup>1</sup>             | 31.4%                     | 36.0%                          | 21.8%          | 22.9%             |
| Age < 45  | 47.8%                     | 43.4%                          | 53.7%          | 50.9%             |
| Race/ethnicity other than White <sup>1</sup>    | 24.1%                     | 20.1%                          | 36.8%          | 33.2%             |
| High school education or less <sup>1</sup>      | 35.5%                     | 31.2%                          | 33.5%          | 29.1%             |
| Respondent currently employed for pay           | 51.3%                     | 52.0%                          | 53.4%          | 53.5%             |
| Respondent has more than one residential number | 7.2%                      | 6.8%                           | 6.2%           | 5.8%              |

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

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

## Demographic Characteristics for Completed Interviews across RDD and Listed Numbers in the LMTS-9 Sample

| Demographic Characteristic                                    | Unweighted                  |                                | Weighted       |                   |
|---|-----------------------------|--------------------------------|----------------|-------------------|
|   | RDD Interviews<br>(n=1,023) | Listed Interviews<br>(n=3,970) | RDD Interviews | Listed Interviews |
| Age 12-17 <sup>1,2</sup>                                      | 49.9%                       | 75.4%                          | 32.8%          | 46.8%             |
| Race/ethnicity other than White <sup>1,2</sup>                | 53.6%                       | 42.0%                          | 41.3%          | 33.7%             |
| Lives with both parents (12-17 year olds only) <sup>1,2</sup> | 60.0%                       | 80.9%                          | 64.9%          | 82.3%             |
| Lives in own home (18-24 year olds only) <sup>1,2</sup>       | 42.9%                       | 18.4%                          | 45.6%          | 14.3%             |
| Respondent currently employed for pay <sup>1,2</sup>          | 42.3%                       | 32.6%                          | 53.0%          | 42.1%             |
| Respondent has cell phone                                     | 43.2%                       | 40.7%                          | 45.4%          | 44.4%             |

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

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

## Smoking Indicators for Completed Interviews across RDD and Listed Numbers in the NY ATS Sample

| Smoking Indicator  | Unweighted                |                                | Weighted       |                   |
|--|---------------------------|--------------------------------|----------------|-------------------|
|  | RDD Interviews<br>(n=777) | Listed Interviews<br>(n=1,286) | RDD Interviews | Listed Interviews |
| Respondent has ever tried cigarettes                           | 80.6%                     | 81.2%                          | 76.1%          | 76.5%             |
| Respondent has smoked 100 cigarettes or more in lifetime       | 54.6%                     | 53.4%                          | 49.1%          | 48.2%             |
| Respondent is current smoker                                   | 27.6%                     | 24.9%                          | 19.8%          | 20.9%             |
| Other smoker(s) in household                                   | 8.5%                      | 7.7%                           | 8.0%           | 9.4%              |
| Respondent tried to quit smoking in past year                  | 38.8%                     | 47.0%                          | 37.4%          | 48.4%             |
| Respondent exposure to others' smoke in past week (Mean Hours) | 4.99                      | 4.71                           | 3.99           | 4.47              |

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

## Smoking Indicators for Completed Interviews across RDD and Listed Numbers in the LMTS-9 Sample

| Smoking Indicator   | Unweighted                  |                                | Weighted       |                   |
|---|-----------------------------|--------------------------------|----------------|-------------------|
|   | RDD Interviews<br>(n=1,023) | Listed Interviews<br>(n=3,970) | RDD Interviews | Listed Interviews |
| Respondent has ever tried cigarettes <sup>1,2</sup>                           | 41.8%                       | 27.6%                          | 52.1%          | 35.6%             |
| Respondent has smoked 1 pack or more of cigarettes in lifetime <sup>1,2</sup> | 54.1%                       | 44.1%                          | 60.9%          | 46.7%             |
| Respondent was ever a regular smoker <sup>1,2</sup>                           | 42.1%                       | 28.1%                          | 48.1%          | 27.3%             |
| Respondent will likely smoke in the next year <sup>1</sup>                    | 4.6%                        | 2.5%                           | 5.2%           | 3.3%              |

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

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

## Smoking Indicators for Completed Interviews across RDD and Listed Numbers in the LMTS-9 Sample, Continued

| Smoking Indicator  | Unweighted                  |                                | Weighted       |                   |
|--|-----------------------------|--------------------------------|----------------|-------------------|
|  | RDD Interviews<br>(n=1,023) | Listed Interviews<br>(n=3,970) | RDD Interviews | Listed Interviews |
| Respondent tried to quit smoking in past year  | 70.7%                       | 68.2%                          | 69.5%          | 67.8%             |
| Smoking in respondent's peer group <sup>1,2</sup>                                      | 48.8%                       | 36.6%                          | 56.8%          | 41.4%             |
| Respondent exposure to others' smoke in past week (mean number of days) <sup>1,2</sup> | 4.12                        | 3.51                           | 4.24           | 3.65              |
| Other smoker(s) in household <sup>1,2</sup>  | 35.0%                       | 26.5%                          | 35.8%          | 26.6%             |

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

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

# Conclusions

- Listed numbers were much more effective in meeting sample targets for respondents by age groups and race/ethnic groups in LMTS-9
- In both studies, analysis of sample dispositions confirmed that the listed numbers were significantly more efficient than RDD numbers in reaching working numbers, residential numbers and eventually, completed interviews
- There were very few significant differences between respondents from listed sample and respondents from the RDD sample in the general population NY-ATS

# Conclusions, Continued

- RDD respondents in the subgroup targeted LMTS-9 study were more likely to differ when it comes to key demographics, and more likely to report all types of smoking behaviors than those from the listed sample, both for themselves and others in their household.
- The significant differences observed on all fronts in this research suggest the potential for bias. This is especially true in the case of LMTS-9, where weighting does not have much of a reducing effect on the potential biases.

# Future Research

- Future efforts should be made to continue to develop a clear understanding of the biases that could be involved in using listed sample
- Potential biases could affect estimates that researchers may hope to make as they normally would with a 100% RDD sample
- Investigate how using listed sample affects the effort and cost involved in conducting surveys
  - Serious consideration to the potential trade-offs more efficient data collection via dual-frame sampling may have to endure in the form of increased potential for bias, and vice versa.

# More Information

- Paper available at: <http://www.rti.org/tsmII>

Email

[droe@rti.org](mailto:droe@rti.org)

-or-

[dcurrivan@rti.org](mailto:dcurrivan@rti.org)