**IMPROVING COVERAGE OF HARD-TO-REACH POPULATION GROUPS
IN HOUSEHOLD TRAVEL SURVEYS**

Sudeshna Sen, Ph.D.
NuStats
206 Wild Basin Road
Building A, Suite 300
Austin, TX 78746
Phone: (512) 279-4147
Fax: (512) 306-9077
Email: ssen@nustats.com

Sally Amen
NuStats
206 Wild Basin Road
Building A, Suite 300
Austin, TX 78746
Phone: (512) 279-4137
Fax: (512) 306-9077
Email: samen@nustats.com

Innovations in Travel Modeling Conference

May 9-12, 2010
Tempe, AZ

INTRODUCTION

Effective development and continual refinement of activity- and tour-based models depends upon the robustness of the household travel survey data collected to support the models. It is important that the data reflects the full diversity of the behavioral determinants of travel and provides for a reliable and statistically valid model. In order to collect such a rich source of information for use in conventional or new generation models, efficient sample designs are needed that ensure adequate coverage of key demographic segments and travel behaviors of interest. In this paper, we examine and evaluate the efficiency of multi-frame sample designs used in recent household travel studies.

In a perfect world, household travel behavior information would be obtained as needed from the appropriate population groups or subgroups with simple appeals or requests for information. Unfortunately, today's society is one of limited funds and imperfection in surveys. As a result, the data collected lacks information from specific population subgroups (undercoverage) and is often incomplete with regards to the information collected (item non-response).

Undercoverage or inability to adequately represent all population groups and travel behaviors of interest is one of the key concerns in household travel behavior surveys and is primarily associated with two factors. First, key population groups such as large households, low-income households, minority households, and households comprising of young adults have lower-than-average participation levels. Since the travel behavior characteristics of these population groups are likely to vary from the general population, under-representation of these groups can result in biased inferences. Second, use of imperfect sampling frames can result in significant under-representation of the target population (see Sen *et al.*, 2009 for review of sampling frames). For instance, Random Digit Dial (RDD) landline frame, that is the most commonly used sampling frame for over three decades in household travel surveys, provides coverage of only 81% to 85% of all households in the United States due to lack of coverage of no-phone households or the growing number of cell-only households (Link *et al.*, 2008; Kulp *et al.,* 2008).

Clearly, coverage bias in household travel surveys can lead to inaccurate estimates of the determinants of travel behavior which can have serious implications on the development of short- and long-range transportation plans. As a result, efficient and effective sample design is important to ensure adequate representation of these hard-to-reach population groups.

SAMPLE DESIGN

Background

The selection of an appropriate sampling frame(s) and sampling method is critical for an effective sample design that guards against unplanned selectiveness and produces a robust data set that is representative of the population.

Sampling Frame

A sampling frame can be defined as a body of information about the population being investigated that is used as the basis for selecting samples and in subsequent estimation procedures. In the context of household travel surveys, a sampling frame is an up-to-date listing of every household in the population,

with identification information such as telephone numbers or addresses. The common sampling frames considered in household travel surveys are as follows:

- An Address-based sampling frame includes all residential addresses that receive U.S. mail delivery. Its main advantage is its reach into population groups that typically participate at lower-than-average levels, largely due to coverage bias (such as households with no phones or cell-only households).

- The Listed Residential landline frame includes listed telephone numbers from working blocks of numbers in the United States for which the name and address associated with the telephone numbers are known. This frame can be used to strengthen the coverage of households with listed landlines. The advantage of this frame is its efficiency in conducting the survey effort—being able to reach households and secure their participation in the survey in a direct and active approach.

- The Random Digital Dial (RDD) landline frame includes both listed and unlisted landline telephone numbers from working blocks of numbers in the United States. Technically, RDD landline frame provides a near 100% coverage of all households with landlines. However, RDD landline frames do not provide coverage of no-phone households or the growing number of cell-only households.

- The RDD cell phone sampling frame is a list of all possible cell phone numbers that provide coverage of the ever-increasing cell-only households. According to the 2008 National Health Interview Survey (NHIS), adults from cell-only households are more likely to be younger (18 to 29 years), unmarried, Hispanic, renters, and have low-income status. Hence, inclusion of a cell phone frame to supplement other frames provides a comprehensive coverage of the study area. However, cell phone frame is associated with low response rate, high refusal rate, and low refusal conversion rate.

Sampling Method

Stratified probability sampling is a common technique used in household travel surveys as it ensures high levels of coverage, accuracy, and efficiency compared to non-probability samples. A strictly random sample from the study area would result in under-representation of households with specific travel characteristics, thereby reducing the anticipated model validity. By stratifying the sample, survey goals can be allocated to specific portions of the region or specific population groups in order to maximize the inclusion of different travel characteristics. The stratified sampling method thus results in over-samples for some strata to ensure that we capture the diversity of the population according to specific geographic and behavioral factors affecting travel activity in the study area. Thus, within strata and frame, households will be selected with equal probabilities but the combined sample (across strata and frames) will comprise an unequal probability sample of households.

Application

In this section, we examine and evaluate the efficiency of multi-frame sample designs used in two on-going household travel surveys conducted in the State of Oregon - Oregon Household Activity Survey (OHAS) and Lane Council of Governments Activity Survey (LCOGAS). A multiple sampling frame approach was employed in both the studies which combined the strengths of Address-based, Listed

Residential, and New Movers sampling frames to ensure a broader and in-depth coverage of the study area. Note that only addresses from the Address-based frame that could be matched to telephone numbers were used. New Movers frame includes addresses and listed phone numbers of residents who recently moved into the study area such as trade-up home buyers, apartment renters, and retirees.

A stratified sampling approach was utilized to target hard-to-reach population groups. Specifically, targeted samples were drawn from the Listed Residential sampling frame. This includes (1) Sample of large households i.e., households with four or more members, (2) Sample of low-income households i.e., households earning less than $25,000, and (3) Sample of one-person worker households that earn more than $50,000. An income constraint was placed to control for the expected over-representation of elderly households.

Note that the sample was ordered in phases and the performance of each sample was monitored. Based on performance of prior samples, specifications of subsequent sample orders were adjusted. Table 1 presents the distribution of the total sample drawn by sample type. As shown in the table, 85% of the sample for OHAS was drawn from Listed Residential frame, 9% from Address-based frame, and 6% from New Movers frame. In comparison, 74% of sample for LCOGAS was drawn from Listed Residential frame, 21% from Address-based frame, and 5% from New Movers frame. Overall, Address-based sample was found to be more productive for LCOGAS compared to OHAS.

TABLE 1: DISTRIBUTION OF SAMPLE TYPE

Study	Address-Based Matched	Listed Residential				New Movers	Total
		General	Targeted				
			Large Households	Low Income Households	One-person Worker Households		
2009 Oregon Household Activity Survey [OHAS]	9%	46%	19%	8%	12%	6%	100%
2009 Lane Council of Governments Activity Survey [LCOGAS]	21%	41%	19%	6%	8%	5%	100%

Generally, in household travel studies, several demographic segments are under-represented due to low contact rates and high refusal rates. These groups include the following:
- Large households
- Hispanic or African-American households
- Low-income households
- Young households
- One-person worker households.

Table 2 presents the performance of different types of sample used to capture population groups with lower-than-average participation levels. The sample type performance is assessed by comparing the representativeness of population groups of interest in the census with those recruited from the sample. To illustrate, 29% of the targeted sample of low-income households were large households recruited for OHAS. In comparison, 24% of the survey universe for OHAS i.e., all households in the study area, are large households. The higher representation of large households in the data collected from targeted sample compared to the census is reflective of good sample performance.

As shown in the table, Address-based or general Listed Residential sample (i.e., Listed Residential sample without oversampling) moderately under-represented large households, young households, and one-person worker households while significantly under-represented low-income households. The targeted samples, on the other hand, make up for the under-representation of these population groups. The targeted samples of large households, low-income households, and one-person worker households successfully captured the respective population groups. Since low-income households are more likely to be large households, the targeted sample of low-income households was also effective in capturing the large households. Similarly, the New Movers sample effectively captured the large households and young households.

TABLE 2: SAMPLE PERFORMANCE

Hard-to-Reach Population Groups	Study	Address-Based Matched	Listed Residential					New Movers	Census
			General	Targeted				
				Large Households	Low Income Households	One-person Worker Households		
Large Households	OHAS	18%	20%	39%	29%	8%	36%	24%
	LCOGAS	21%	21%	54%	13%	12%	27%	24%
Young Households	OHAS	22%	20%	28%	27%	14%	38%	30%
	LCOGAS	22%	22%	34%	21%	23%	30%	30%
Low-Income Households	OHAS	10%	14%	5%	37%	9%	27%	30%
	LCOGAS	16%	12%	2%	40%	10%	26%	30%
One-person worker households	OHAS	9%	8%	3%	6%	19%	10%	11%
	LCOGAS	14%	10%	1%	21%	23%	11%	11%

CONCLUSION

Undercoverage or inability to adequately represent all population groups and travel behaviors of interest is one of the key concerns in household travel surveys. Inadequate representation of key population groups can lead to inaccurate estimates of the determinants of travel behavior in travel demand models. Consequently, efficient and effective sample designs are needed that ensure adequate coverage of the demographic segments and travel behaviors of interest. An evaluation of the multi-frame sample designs used in recent household travel studies show that appropriate selection of sampling frame(s) and sampling method is critical to adequately represent the hard-to-reach demographic segments (such as large households, low-income households, young households, and one-person worker households). The use of multiple sampling frames – Address based, Listed Residential, and New Movers sampling frame, with appropriate oversampling of specific hard-to-reach population groups, was found to be effective in ensuring collection of household travel survey data that is representative of the population.

REFERENCES

Kulp, D.W., J. M. Brick, and M. Fahimi. (2008) Biases in RDD Sampling: A 21st Century Digital World Reassessment. *American Association for Public Opinion Research Conference*, New Orleans.

Link, M. W., M. P. Battaglia, M. R. Frankel, L. Osborn and Ali H. Mokdad (2008). A Comparison of Address-Based Sampling (ABS) Versus Random-Digit Dialing (RDD) For General Population Surveys. *Public Opinion Quarterly*, Vol. 72, No. 1, pp 6–27

Sen, S., J. Zmud, and C. Arce (2009) Evaluation of Efficiency and Effectiveness of Cell Phone Samples in Transportation Surveys, forthcoming, *Transportation Research Record, Journal of the Transportation Research Board*, Issue 2105, pp. 44-50.