

10th International Conference on Transport Survey Methods

## Workshop synthesis: Sampling issues, data quality & data protection

Jimmy Armoogum<sup>a\*</sup>, Jennifer Dill<sup>b</sup>

<sup>a</sup>*UPE-IFSTTAR-DEST, 14-20 Boulevard Newton - Cité Descartes - Champs sur Marne, F-77477 Marne la Vallée, France*

<sup>b</sup>*Nohad A. Toulon School of Urban Studies & Planning, Portland State University, Portland, Oregon, United States*

---

### Abstract

This workshop discussed various aspect of the mathematical part of survey methodology, as well as archiving and confidentiality issues aimed at improving data quality and its use through time. Participants identified ways to correct or minimize bias by dealing with incomplete sampling frames, using weighing and imputing procedures. We discussed methods to archive and share GPS-based survey data to preserve anonymity. Finally, we debated research needs on these topics for the next following years.

© 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of International Steering Committee for Transport Survey Conferences ISCTSC

*Keywords:* Sampling, weithing, imputing, nonresponse, bias, data quality, measurement errors, archiving, data privacy

---

### 1. Introduction

Issues related to sampling and data quality can affect the usefulness of any travel survey. This workshop addressed issues such as respondent bias, non-response, sampling issues, and declaration bias, along with models that can help reduce the impacts of these issues and preserve data quality for transport analysis. The workshop also addressed the challenges of data protection, including confidentiality and storing to ensure long-term usability.

### 2. Sampling Issues

A sampling frame provides the means to reach each member of the study population who would be eligible to be surveyed. Researchers have several options for sampling frames, including using household- or person-based sampling and phone or address sampling frames. A common approach for travel surveys has been to sample

---

\* Corresponding author. Tel.: + 33 1 81 66 86 07; fax: 33 1 81 66 80 01.

E-mail address: [jimmy.armoogum@ifsttar.fr](mailto:jimmy.armoogum@ifsttar.fr)

households and interview all members of a household. One advantage of this approach is that the researcher is able to analyze how household members interact in travel behavior and how they share the vehicles. The disadvantages of this approach are that it is not easy to contact everybody in a household and interview them about the same day (or week) and that the burden of the household may be very high. This can result in a low response rate if a household is considered non-responsive if only one member is not interviewed. This will also bias the response rate to smaller households. Often, the survey will allow proxy responses—one household member can report data for another household member. However, this process introduces other sources of error and bias (Badoe & Steuart, 2002).

One problem with any sampling frame is that they are often incomplete; they are missing some of the eligible sampling units. For example, some people have phone numbers that are not in a phone registry. The practices of phone number registries vary significantly between countries. Random digit dialing (RDD) can address the issue of incomplete sampling frames for phone samples, but poses other problems. For example, the researcher cannot send the household an introductory letter prior to the phone survey – a technique that has been shown to improve response rate.

The decision of how to address incomplete sampling frames would be made easier with more information about who is missing. In most cases, it is unlikely that the missing people or households are random. Knowing, for example, the demographics of the missing people can lead to better decisions, such as choosing another frame, sampling strategy, or weighting, and allows the researchers to most appropriately interpret their findings.

One increasingly common approach to dealing with sampling bias is to augment the sample with additional sources, use more complex sampling frames. For example, it is becoming more common for surveyors to join mobile and land-line phone samples. Researchers might also use two-stage sampling approaches, stratified sampling, and over-sampling – all to help ensure that harder-to-reach individuals (e.g. younger adults, people who use less-common travel modes, etc.) are included. These approaches also help address another issue—collecting data on relatively rare behavior. For example, long distance mobility is very unevenly distributed in the population (e.g. 47% of individuals made no trip over 100 km during at least 3 months before the interview, according to the last National Travel Survey conducted in France). Thus uniform sampling is especially inefficient for understanding long distance mobility.

These more complex sampling frames pose other difficulties. For example, when two or more data sources are combined, care must also be taken to minimize duplicate entries and address potential inconsistency that may appear in both sampling frames. In addition, calculating confidence intervals is more challenging.

### **3. Data Quality**

#### *3.1. Non-response*

Non-response is the inability to measure all the units of sample of all variables of interest. Two different types of non-response exist: (1) Total (or unit) non-response, where there is no information about the unit selected other than sample frame; and (2) Partial (or item) non-response, where the unit selected responded only to a part of the survey.

##### *3.1.1. Unit Non-response*

Falling response rates are a growing problem for travel survey researchers. The increased reliance on mobile phones is pointed to as a key contributing factor in this trend. While this trend is seen worldwide, there are still significant differences in response rates across countries. However, workshop participants noted that comparing response rates across surveys is not a simple task because of variations in what counts as a response.

While non-response results in a reduced sample size, a more important concern of researchers is the possible impact of non-response bias. Bias is introduced when those that do not respond to the survey are systematically different from those that do respond on key variables of interest. Researchers first must *understand* the unit non-response. For travel surveys aimed at representing the general population, this is commonly done by comparing the sample to a census. However, this will only identify the demographics of the underrepresented units, e.g. lower-income households or younger adults. It does not tell us how the travel behavior of the non-respondents might systematically differ from those of the respondents. For example, is it safe to assume that younger adults that do not respond to the survey travel the same as younger adults who did respond? Possibly not.

Download English Version:

<https://daneshyari.com/en/article/1106767>

Download Persian Version:

<https://daneshyari.com/article/1106767>

[Daneshyari.com](https://daneshyari.com)