

# CHICAGO REGIONAL HOUSEHOLD TRAVEL INVENTORY

## *White Paper: Efficient Data Collection*

*Primary Authors:*

*Johanna Zmd, NuStats*

*Greg Dawson, NuStats*

*Jon Wivagg, DataSource*

*Billy Bachman, GeoStats*

*Jean Wolf, GeoStats*

*Secondary Authors:*

*Kelly Castleberry, DataSource*

*Stacey Bricka, NuStats*



NuStats

3006 Bee Caves Rd., Suite A-300. Austin, Texas 78746  
(512) 306-9065 . fax (512) 306-9077 . [www.nustats.com](http://www.nustats.com)

Contact: Stacey Bricka, Research Director



# TABLE OF CONTENTS

---

Introduction	1
Background	3
Data Collection Issues	3
Mixed-Mode Designs for Household Travel Surveys	5
Internet Data Collection	7
Global Positioning System (GPS)	9
Recommendations	11
References	15



# INTRODUCTION

---

The Chicago Regional Household Travel Inventory is a comprehensive study of the demographic and travel behavior characteristics of residents in the greater Chicago area. Sponsored by the Chicago Metropolitan Agency for Planning (CMAP) and the Illinois Department of Transportation (IDOT), the study universe is defined as households residing in the Illinois counties of Cook, DuPage, Grundy, Kane, Kendall, Lake, McHenry, and Will. The project has two phases: Design and Data Collection. The design phase took place in the fall of 2006. The full data collection effort will take place in 2007.

The purpose of the design phase of the study was to identify (through research and primary data collection) the most appropriate design and methodological aspects that maximize the quality and validity of the inventory data for modeling purposes. The three main objectives of the design phase were: (1) to validate existing budgetary assumptions regarding data collection efforts anticipated for the full study (and establish new assumptions as necessary), (2) to ensure that the inventory design elements and methods provide for a data set that supports the development of a valid model, and (3) to vet the inventory design recommendations through a series of white papers, supported by both primary and secondary research, using a peer review panel of both topical and regional experts. This document is one of the four white papers developed as part of the study's design phase.

The purpose of the white papers prepared under this design phase is to address specific issues pertaining to the design of the data inventory and supporting data collection effort. Because the data will be used to both update the current regional travel demand model as well as for developing new models, the actual elements contained in the inventory need to meet the needs of both efforts. These white papers serve to delineate those elements that are critical to both efforts. Ultimately, the cost trade-offs, respondent reactions, white paper recommendations, and input from the expert and local peer review panels will be used by CMAP staff to finalize the actual inventory contents.

Each white paper has a primary author team and a secondary author. The primary author team was responsible for ensuring that the document addressed the necessary elements and provided actionable recommendations for the data collection phase. To facilitate this, the primary authors provided the project manager with a list of key questions or design elements for the pilot test (these are discussed below). The secondary author's role was as reviewer, with the specific intent being to balance the paper, to ensure that it was well-rounded and practical in approach and recommendations.

The white papers combine secondary research with primary data collection (through the study pilot) in order to make recommendations on key issues that impact inventory design. These issues were identified at the project kick-off meeting<sup>1</sup>, held Tuesday, May 23, 2006 in Austin, Texas and include: (1) inventory content, (2) sampling considerations, (3) maximizing participation, and (4) efficient data collection. Each of these is discussed in a separate document. This paper focuses on the survey modes necessary for efficient data collection, including the following issues:

- Data collection modes: telephone, web, GPS, in-person, mail
- Structure and length of survey instruments
- GPS – vehicle, on-person – who?
- Timing of special population surveys
- Household surveys, visitor surveys, employer surveys, commercial travel?

---

<sup>1</sup> This kick-off meeting included the project team members as well as members of the project's expert and peer review panels.

Given the other pilot studies underway or recently completed by NuStats, the team charged with writing this paper was also charged with leveraging the results and lessons learned from those other studies and focusing here on new areas or questions to address. As a result, the team identified the following research questions:

- How do we optimize the efficiency and effectiveness of using multiple modes to obtain higher levels of respondent participation and response?
- Where might the use of multiple modes add or reduce bias?
- Which modes are most appropriate for each type of sample, each type of respondent, and each stage of data collection?

These translated into the following pilot elements:

- 1) Use of an Address-based sample. Technological improvements have made the application of address-based samples more economical in recent years. In addition, the use of address-based samples is expected to minimize coverage bias introduced due to the growth in cellular phone-only households. The pilot will be used to address the question of whether the types of households included in the study through an address-based sample are statistically different in composition and travel from those included using a more traditional telephone-based sample. The pilot will need to allow for testing of response rates for the different modes (telephone recruitment, mail recruitment, and web options for providing recruitment data), as well as the more important tests of significant differences between the two types of respondents.
- 2) Obtaining Transit Trip Segments. What is best method for capturing all trip segments? Current pilots in San Diego and Washington DC have extensive focus on transit users in their pilots, so those results can help to inform this study. For this pilot, the question is what enhancements can be made to improve the interviewer's ability to collect complete transit trip segments.
- 3) Internet usage/penetration. Web surveys are becoming increasingly common. However, the survey literature is candid about the respondent groups who utilize this option. If an internet option is offered here, perhaps in conjunction with the passive recruitment effort, will we include in the study travelers that might not otherwise participate?



# BACKGROUND

---

Most travel survey research over the past decade has focused on improving the details obtained through these travel studies. This includes whether the focus should be on the trip, the activity, or the places visited; whether the respondent is recording all travel; whether non-motorized travel should be collected, etc. While these details have strengthened the reliability of the results, the increased burden on the respondents has led to declining response rates. As response rates decline, concerns about those who are not participating in these studies have increased (Zmud and Arce, 1997; Zmud, 2004). Non-response is a growing concern for practitioners who rely on the data collected from household travel surveys in the U.S. – at regional, state and national levels. In fact, a National Academy of Sciences review of the 2001 National Household Travel Survey (NHTS) concluded that its low response rate (41%) gives “cause for concern because of the likelihood of significant non-response bias because there is reason to believe that the travel behavior of non-respondents differs significantly from that of respondents”.<sup>2</sup>

## DATA COLLECTION ISSUES

A key issue in the Chicago Regional Household Travel Inventory is how to collect the data. The typical method of data collection was the use of an RDD sampling frame, the mailing of an advance letter to matched addresses, recruitment via the phone, mailing of a travel log, and retrieval of travel information via the phone (Zmud, 2003). This method of data collection was used in the CMAP pilot. It has also been the focus on much discussion in recent years due to the growing number of Americans who rely solely on a cell phone for their telephone service. A new study by the Pew Research Center estimates 7-9% of the general public is cell-only. These people are younger, less affluent, less likely to be married, or to own their own home. This trend of cell-only persons presents a challenge to household travel surveys and compounds the challenges associated with non-response in household travel surveys. As the cell-only population has grown, household travel surveys that rely on landline samples, such as RDD, have experienced a sharp decline in the percentage of younger respondents.

To attempt to address the cell phone-only issue, the CMAP pilot included a “passive mailing” condition for recruitment. In this condition, the utility of a mail-only recruitment package (i.e., cover letter, study brochure, household questionnaire (HQ), and postage-paid envelope for HQ return) was tested. Of the 1,000 packets mailed, 77 were returned as non-deliverable. Of the remaining 923, 53 HQ questionnaires were received for a 6% response rate. Those that did participate from this mode did not exhibit statistically different travel patterns from the data obtained from the RDD sample. While the issue of whether or not this mode could have captured cell phone only households was not addressed in the CMAP pilot, it will be addressed later in this paper.

Overall, the CMAP pilot resulted in a 9% response rate, with differential response by geographic location and by survey type (i.e., 1-day versus 2-day; travel versus activity). The sample dispositions for the CMAP pilot indicated a high proportion of call attempts in the recruitment phase were non-contacts and refusals, meaning we never even had the chance to convince them to participate in the survey. Likewise from prior survey experiences, we know that a common reason for non-response is that sampled households do not “recall” receiving the advance mailing. In the passive mailing

---

<sup>2</sup> Measuring Personal Travel and Good Movement (2003). Special Report 277, Transportation Research Board, Washington, DC.

condition, most debriefed respondents could not recall receiving the package. Sampled households can not participate if they are unaware of the survey experience.

The pilot sample also appeared to under-represent larger, low income, young adults, apartment dwellers, and non-white households. This failure to match Census parameters on these characteristics is not unique to the CMAP pilot. It is a consistent issue in household travel surveys. So, a question that this white paper attempts to address is whether modifying the method of data collection would / could have a significant effect on increasing response rates and improving data quality. Given the low response rate in the pilot, the authors agreed that the most significant issue revolves around the use (or not) of multiple-mode surveying to increase response and data quality (Zmud et al., 2003). "Modes" refer to approaches used either to contact or to obtain data from survey respondents. Some possible technologies that could be used in the data collection process are mail, phone, touchtone data entry, FAX, the Internet, and global position system (GPS). The mixing of modes and approaches seems to be limited at times only by the creativity of survey managers. For example, certain survey activities, such as prenotification and reminder messages, can be administered using one mode, but data collection might rely on a different, but single, mode. Or, the survey data can be collected using more than one mode (mixed-mode). For example, prenotification, reminder messages, and the initial questionnaire might be sent via mail and respondents given the option of reporting using mail, phone, the Internet, or some other approach.

The different response rates and response patterns by survey mode type are important considerations. Each mode has significant advantages and disadvantages. Self-administered modes (mail and Internet) generally cost less and may yield more objective data, while interviewer-administered modes (like phone) allow active probing and often yield higher response rates and more accurate data. For example, in a recent pilot test of the household travel survey in Washington, DC metro area, we found that simply using a phone number for an address-base sampled household nearly tripled the response as compared to offering a \$100 after-completion incentive. Because of the different advantages and disadvantages of various survey modes, a basic assumption has developed that offering multiple modes of reporting makes the reporting task easier for respondents, which will lead to higher response and better quality data. In addition, if respondents can be encouraged to use the more cost-effective modes, the costs of data collection can be significantly reduced, or at least better controlled.

The use of multiple modes of data collection is a trend that appears to be increasing in government-sponsored surveys. For example, the Current Expenditures Quarterly Interview Survey (CEIS) was designed to be administered by personal visit, through the use of computer-aided personal interview (CAPI). The survey procedures were initially established so that a telephone interview was supposed to be a rarity, to be used only in unusual situations, for example, when a respondent demanded it. A face-to-face interview was chosen because the length and complexity of the questionnaire seemed to require the skills and in-person presence of an interviewer in order to encourage response and the collection of high quality data. However, a recent paper by McGrath (2005) revealed that about 42 percent of the interviews are currently being done by phone (decentralized CAPI), with unknown effects on data quality.

Dillman (2002) points out that the "future of surveying is far more likely to evolve toward the use of different survey modes for different studies than it is to be the disappearance of older modes in favor of only one or two new ones." For example, in the 1970s, telephone surveys were expected to replace face-to-face methods because of their greater cost- and sampling efficiencies (Massey, 1988). For the last five years, some people have been talking about Internet surveys replacing all other modes because of the greater cost-efficiencies and respondent reporting improvements (Adler, et al., 2002). GPS is also highly appreciated for its passive, real-time data capture capabilities, leading to improved

data quality (Wolf, et al., 2001). Leveraging technology is important for dealing with the conflicting pressures related to response rate and respondent burden. However, technological solutions run the risk of fitting the survey situation to the survey mode instead of the reverse – fitting the survey mode to the survey situation.

From a research perspective, however, the use of multiple modes in a survey such as the Chicago Regional Household Travel Inventory requires further examination. Just because this represents a trend in contemporary surveying, does this mean it is the right approach for the CMAP survey? Effective use of a mixed-mode approach can reduce overall costs versus the use of a single high-effort interviewer-administered mode (like phone or in-person) as it allows motivated respondents to respond to the generally cheaper self-administered modes and reserves more intensive (and expensive) modes only for more reluctant respondents. Generally, a mixed mode approach remains more expensive than the cheaper single modes and for each additional survey mode employed in a study, the data collection requires increasingly complex fieldwork and quality assurance strategies.

A key issue for this project is the question of whether offering concurrent, mixed modes of responding actually leads to higher response. There does not seem to be a clear answer (note the results of the CMAP pilot in terms of the “passive recruitment”). The introduction of more than one mode of data collection can add bias that results from the different response modes and the decision to use a mixed mode approach must consider whether any reduction in non-response bias – either unit nonresponse or item nonresponse – outweighs the possible bias resulting from the introduction of mode effects.

## MIXED-MODE DESIGNS FOR HOUSEHOLD TRAVEL SURVEYS

The use of multiple mode surveys can increase response rates and, for this and other reasons, mixed-mode surveys have long been the best practice approach to travel data collection despite awareness of mode effect bias. However, determining the optimal fielding strategy for a survey to collect data to support the development of a household travel inventory in a large and diverse metropolitan area requires a reconsideration of the significance of mode effects in multiple-mode surveys due to the following considerations:

- Travel data are minimally susceptible to mode effects
  - Travel data collection instruments are highly sophisticated and require multiple phases
  - Experience and best practices have shown that mode effects in travel studies are consistent across modes on key variable measures
  - Mixed mode response options can reduce survey costs
  - Travel data inventories often require geographic data that respondents cannot directly provide
1. **Travel data are minimally susceptible to mode effects.** The mode effects of self-administered (mail, web) versus interviewer-administered (CATI, in person) reported in the literature relate to response differences on primarily opinion-related items and reports of behavior subject to social desirability concerns. However, the primary data needed for travel inventories are facts rather than opinions. Reports of behavior required in travel inventories ask about relatively neutral behaviors such as the times, modes, and destinations of daily travel in a very limited time window rather than positive or deviant behaviors such as volunteering or drug use. Consequently, travel-related behavior has few clearly socially desirable or socially undesirable responses and almost no need for opinion-based data that would create significant mode effects due to the presence or

absence of an interviewer. **For this reason, we would recommend that multiple modes of data collection be employed.**

- 2. Travel data collection instruments are highly sophisticated and require multiple phases.** Unlike simpler opinion-based questionnaires where almost all respondents can understand the data the instrument intends to collect, travel data collection requires respondents to convert a natural, familiar, and infrequently examined process (their daily travel) into a large number of very precise data elements (specific times, addresses, etc.). While simpler surveys mainly require a consideration of respondent preference or how to present response instructions, travel data instruments must consider the respondent's ability to fully understand the data reporting task and the mode or modes that would most effectively communicate this. While some respondents would best learn from the probing and instructions of a trained interviewer, others might find it easier to understand printed or online instructions or examples. A mixed mode approach to travel data collection can maximize the training exposure the respondent receives and use mode effects to enhance the quality of this particular type of data. **This would suggest that having print materials or web materials available to respondents regardless of whether they actually complete via these modes or complete the survey via the phone would enhance the overall response.**
- 3. Experience and best practices have shown that mode effects in travel studies are consistent across modes on key variable measures.** NuStats and DataSource have conducted many household travel study data collection efforts that have directly compared mail with CATI response on the reporting of trips. Results of this comparison show those households who report their travel data in a phone interview without the benefit of a completed mail questionnaire/diary report fewer trips with more data problems than do households who use the diary as they respond to the phone interviewer. Results also show that households who only return the mail component and do not complete the phone interview also report fewer trips and have more data problems. **So, it is important to use the as the primary interviewing mode to enhance item response (i.e., trip reporting).**
- 4. Multiple mode survey options can reduce survey costs.** The multiple mode approach offers many advantages with the primary disadvantage being the risk of mode effects that might bias resulting data. However, mode effects are just one source of possible bias (and arguably a very small source in travel studies). **Mixed-mode approaches can yield response rates similar to the most direct and aggressive method used without the cost of using this method to approach all sample units by using less expensive modes to collect data from sample units with a higher response propensity.** This potential costs savings allows researchers to focus on reducing or measuring other sources of bias such as increasing response rates to offset potential non-response bias, cognitive interviewing and focus group testing of instruments to offset response bias, or even conducting experimental tests to assess the actual magnitude and direction of mode effect bias.
- 5. Travel data inventories often require geographic data that respondents cannot directly provide – Travel data inventories rely extensively on geographic coordinates of trip-end locations.** Respondents virtually are never able to provide geographic coordinates of trip-end locations directly. While travel inventories can determine geographic coordinates from respondent-provided addresses, respondents are sometimes imperfect sources of information about the location of their trip-ends and the timing of their trips. The need to validate respondent-reported data has led to the involvement of GPS technology as a new survey data collection mode. With this GPS mode, survey respondents use GPS loggers on their person and/or in their vehicles and these loggers provide objective coordinate data about where they are at each time of



day. GPS alone is subject to considerable mode effects, but can be a valuable data collection mode in the context of a mixed mode survey.

*Bottom line: Overall, a mixed-mode approach is appropriate and likely necessary for travel research and its benefits outweigh the risk of potential mode effects. However, the question of which mode or modes to apply to each of the key data collection tasks remains a key consideration as not all modes should be combined for each task.*

The team assumes that the current modes of computer-aided telephone interviewing (CATI) and mail will be options for this study. Thus, this paper takes a closer look at two newer modes of data collection (Internet and global positioning systems or GPS) in terms of their association with measurement error and/ or non-response bias. These two modes have been singled out because of the general opinion that these modes of data collection are both cost-effective and reduce non-response bias. Note we focus only on data collection issues in the sections below; not sampling issues.

## INTERNET DATA COLLECTION

Although there is a fairly large body of research relating to mode effects in survey work, relatively little research has been published that deals with the relative effects of internet and CATI administration of surveys. One study that does compare the Internet and telephone survey modes concludes that there are significant differences in the data collected by each mode. In addition, a number of studies have reported that response rates are significantly lower for Internet-based surveys than for other modes. Key issues in the use of internet data collection relative to the standard (i.e., phone) are noted below.

### *Presence or absence of interviewer*

In telephone surveys, interviewers fulfill a number of active and passive roles. They lead the respondent through the survey script, asking the questions and recording the answers. They clarify and explain questions and answers that the respondent does not understand. They also, potentially, act as a filter, because they must interpret responses and make those responses fit into the structure of the survey. In addition, their mere presence can act as either an inhibitor or an encouragement to respondents. Those respondents who might feel that a negative judgment would be attached to their true answer could be encouraged to alter that answer to one that is more acceptable. And respondents that might otherwise be reluctant to respond or participate in the survey could be inclined to divulge more than they otherwise would have simply because they are talking to a live, helpful human being. While participating in an Internet survey, on the other hand, a respondent has no interaction with the “interviewer.” There is no one there to clarify or rephrase questions, to explain answer choices, to act as a filter, or to be a potentially inhibiting or liberating presence. Respondents are completely on their own, and this could be reflected in the profile of the answers they give. There may be more non-responses to questions that are difficult to understand. There may also be a lack of precision in responses to questions that seek to make relatively fine distinctions.

### *Pacing (interviewer controlled/respondent controlled)*

The pace of Internet surveys is usually controlled entirely by the respondent; questions are either all presented on one page for the respondent to work through at their own rate, or displayed sequentially, keyed to the input of a response to the previous question. In a telephone survey, however, the pace is substantially controlled by the interviewer: reading speed determines how quickly the respondent can comprehend and answer each question. This can create bias in two ways. First, a respondent can

quickly click through an Internet survey, possibly without fully understanding the questions being asked, but also without necessarily investing significant thought in any of the responses. By contrast, in a telephone survey, the interviewer can control pacing in a way that guarantees that the respondent at least understands the questions being asked, and gives the respondent time to reflect on their answer. On the opposite extreme, an Internet survey allows respondents as much time as they like to ruminate on the questions being asked and (in some cases) allows them to return to an already-answered question and correct the response. While this practice might provide accurate responses, it also lengthens the survey administration time, adding burden. In a CATI environment, though, the interviewer is in control of the pace, potentially limiting the ability of a respondent to think through the questions being asked. However, in the case of *experienced* telephone interviewers, they work at the pace of the respondent.

### *Response rate issues*

One area where the research seems to agree is that Internet-based surveys have a lower response rate than other modes. In part, this may be due to the respondent having a “choice” to participate, in the sense that an invitation is delivered, but then the respondent must take positive action to participate. However, Internet participation is low even relative to self-administered paper-and-pencil surveys. The extent to which this response rate difference is a source of bias is less clear, however. In much of the research that has been done, respondents who participate via the Internet seem to be equally representative of the populations of interest to those who chose other modes.

As part of their effort to make data collection more efficient, Census researchers offered CES panel respondents the option to switch from phone-based to Internet data collection. Although many respondents expressed interest in the Internet-based option, relatively few actually used that mode (Mockovak, 2006). Over 70 percent of panel respondents opted to switch to Internet reporting. These respondents were sent detailed instructions by mail for accessing their online accounts, and were provided passwords. However, only about 60 percent of those who originally opted to use the Internet actually provided any data via that mode.

In another survey of Texas agri-science teachers, researchers showed that participants were much more likely to respond via mail than Internet, even when all respondents had similar access to the Internet (Fraze, Hardin, Brashears, Smith, Lockaby, cite). They attribute this tendency to familiarity with paper-and-pencil surveys and less comfort with computer technology. However, the study also concluded that use of a web-based instrument can significantly speed up the period of data collection, since most Internet participants responded within a couple of days of receiving their invitation to participate. In a similar study of Florida high-school teachers, researchers found no significant difference between the participants who responded via the Internet and those who used the mail (Lang, Raver, White, Hogarty, Kromrey, 2000). Thus, these studies seem to indicate that while there is a trade-off between cost and speed on the one hand, and better coverage of the population of interest on the other, there is not necessarily any demographic bias that results from using one or the other mode.

*Bottom line: Assuming that Internet would be considered only in a mixed-mode context, it adds considerable cost, especially for higher end user interface, without known increases in quality, response, or representativeness.*

## GLOBAL POSITIONING SYSTEM (GPS)

Global Positioning System (GPS) devices have been used to augment travel surveys and improve final model validity for several years. There are several benefits to deploying GPS as part of the travel survey process: better interpretation of paper diary data; more accurate trip times and origin / destination locations; the collection of trip distances, link-level speeds, and delays; and the ability to perform route choice assessments. Additionally, deploying GPS as part of a pilot study can assist with the travel diary design and the finalization of CATI methods.

Typically, GPS devices are deployed to recruited households and assigned to drivers or vehicles. As trips are made, the device records the second-by-second travel activity (time, position, and speed). The resulting database can be used to generate trip tables, zonal travel times, travel routes, and link-level travel statistics (travel time, speed).

The potential value of using GPS data is tempered by sample size. More can be done with larger, balanced samples. In some cases, small sample sizes cannot provide a desirable level of confidence. The following sections describe the specific advantages and list background references for each potential concept:

### *Better interpretation of paper diary data*

Travel survey respondents frequently fail to report a percentage of trips during their assigned travel day. This happens for a number of reasons that cannot be fully mitigated. A GPS sub-sample assists in determining the percentages and characteristics of missed trips because the GPS records actual travel without interaction from the respondent. The paper diary and the GPS can be matched to determine the percentage and types of missed trips. This results in correction factors that can be applied to the entire survey, resulting in more accurate trip rates. NuStats and GeoStats have successfully conducted this approach in the following studies:

- 2001 California Statewide (San Diego, Alameda, and Sacramento Counties)
- 2001 Pittsburgh
- 2002 Laredo
- 
- 2003 St. Louis
- 2004 Kansas City
- 2005 Reno
- 2005 Oregon Statewide Pilot (on-person units)

### *Better trip locations, distances, and times*

GPS technology, by design, collects highly accurate location (within 10 meters typical) and time (synchronized to an atomic clock) data on a second-by-second basis. Used in travel surveys, GPS can provide highly accurate origin and destination location information. In addition, accurate trip distances and travel times between origin and destination zones can be used to calibrate and validate zonal trip time estimate. Wolf, Oliveira and Thompson (2003) presented a comparison of GPS-computed VMT and trip times with model computed VMT using GPS data from the California statewide survey and found consistent under-estimation of VMT in two of the analyzed MPO regions

(Alameda and Sacramento). This study also indicates the presence of consistent rounding problems in the reported start and end times as a possible cause for differences between modeled and measured trip times.

Further, intra-zonal trip VMT and VHT can be extracted and used for developing estimates of intra-zonal travel. Several other data analysis steps assist in validating model results: VMT by trip type, VMT by trip type, and VMT by demographics.

#### *Hourly link-level speeds and delays*

The GPS trace collected during respondent trips can be matched to the travel demand model network in a process usually called link-matching. This application has the benefit allowing model developers to generate average speeds by link class and area type. Several methods exist to match GPS points to a transportation network. Marchal, Hackney, and Axhausen (2004) present some of these methods and introduce a new one suitable for performing mass matching of large GPS datasets. GeoStats has developed similar link matching algorithms and successfully applied them to automatically link-match large GPS datasets.

#### *Route choice analysis*

The GPS trace collected during respondent trips can be used to evaluate route choice analysis. While this step can be used to validate some of the path-making assumptions, its true value is identifying “cut-through” paths and alternative routes between origins and destination. GeoStats has worked with researchers at the “Ecole Polytechnique Federale de Lausanne” (Bierlaire and Frejinger, 2005) to generate a route dataset based on a multi-day GPS dataset. The routes in this dataset were used to perform comparisons against shortest paths algorithm results.

*Bottom line: If cost is not a barrier, GPS provides information value, not otherwise available, as an augment to more traditional modes.*



# RECOMMENDATIONS

---

An effective combination of all data collection tools will maximize efficiency and response for the Chicago Regional Household Travel Inventory and the results of the CMAP pilot. Our recommendations for doing so are presented below.

- Use telephone as the primary interviewing method
- Ensure the widest population coverage with address-frame and telephone match
- Be creative in advance mailing packaging
- Employ a “household questionnaire” to capture cell-only persons -- Be Patient!
- Plan for and conduct follow-up activities for hard to reach populations
- Consider GPS for information value.

1. **Use telephone as the primary interviewing method.** Commonly, household travel surveys in the U.S. have been executed using telephone recruitment of households to complete a travel diary and telephone retrieval of travel data recorded in the diary. There are good experiential reasons for this. In the 1970s, recruiting households by telephone became easier than mail or face-to-face due to the completeness and lower cost of the sampling frames. Related to this latter point is the fact that there is a dynamic relationship between contact method (survey mode), sample composition (who responds), and sample attitude (how they feel about being involved in the survey. Groves *et al* (1992) suggested that most potential respondents take a heuristic approach to the decision to participate in a survey. By heuristic, they meant that the act of survey participation was typically based on one or two highly prominent messages in the survey introduction that appealed to the respondent’s individual motivation for survey participation. The decision to participate was rarely of sufficient personal relevance to cause respondents to systematically process all available information into their decision. They make the decision without thinking; without processing. A good recruitment script, delivered well, will be very effective in recruiting households to participate in the survey.

In addition, retrieving travel data by telephone produced higher quality data because trained interviewers were available to help the respondents provide complete information. The sophistication of the data required for this travel inventory make an interviewer-administered data collection mode preferable. The recommended data collection mode must consider the respondent’s ability to fully understand the data reporting task and decide upon the mode or modes that would most effectively communicate this. With telephone, the capability of surveyors to clarify or rephrase questions, to explain answer choices, and to act as a filter are significantly related to accurate and complete trip data (Schultz, et al, 1990).

While telephone should be the primary data collection mode, it should not be the only mode. Phone works well for many respondents, but mail retrieval may be appropriate for households with specific characteristics of interest that either do not have phones or are not inclined to respond to phone retrieval efforts. The average length of the Chicago retrieval interview for 3 person households was 39 minutes, and it was 41 minutes for 4+ person households. These types of households may be reluctant to complete the retrieval interview (most of which asks for data they have already recorded in written form) by phone. The typical drop-off in the percentage of larger households between the recruitment and retrieval phases suggests that this is the case. In a recent household travel survey in the San Diego region, the drop off among 4+

person households between recruitment and retrieval was nearly 10 percentage points. Thus, the full study should allow retrievals from returned logs. The study design should plan for and conduct follow-up telephone calls for any travel logs submitted via mail.

- 2. Ensure the widest population coverage with address-frame and telephone match.** RDD samples are biased. According to a recent study by the Pew Research Center, telephone surveys that rely on landline samples have experienced a sharp decline in the percentage of younger respondents interviewed. In Pew Research Center surveys over the past five years, the average percentage 18-34 year olds declined from 31% in 2000 to 20% through March 2006 (the population parameter was essentially unchanged through this period). This decline directly related to the fact that cell-only population is heavily skewed toward young people. Seniors stick with landlines, which is one reason why RDD surveys have a difficult time tempering the percentage of respondents over the age of 65. Also according to the Pew Center research, the landline-only group includes a greater proportion of whites than the general public (82% versus 73%). Even among dual phone service households (i.e., landline and cell), people who can be reached on a cell phone are different. People who were interviewed on a cell phone are somewhat younger and more likely to be Hispanic.

In the pilot test for the household travel survey for the metropolitan Washington area, we found that the sample resulting from an address-based frame was significantly different on the key parameters noted above from an RDD-based sample. The address-based frame, an incentive, and use of unlisted numbers did appear to be successful in capturing younger, single persons, apartment dwellers and persons with zero vehicles relative to RDD. The address-based frame, whether or not there was a telephone match, was effective in reaching cell-phone users.

- 3. Be creative in advance mailing packaging.** Chicago Pilot results indicate that many people who agreed to participate did so because of a concern about travel (47%), the information provided during the recruitment interview (25%), or the advance mailing (14%). The high number of hang-ups and initial refusals during the recruitment call attempts mean that many respondents refuse before the interviewer has a chance to fully communicate the study goals.

Research has indicated that advance mailings boost response rates. The Survey Research Center at the University of Maryland conducts ongoing research in survey participation by asking the question at the end of interviews: What was the main reason you agreed to begin this survey? The most frequently provided response has been “letter sent in advance” (Triplett, 1998). Advance letters have been known to increase response rates by 5 to 13 percentage points (Dillman 1978, Traugott 1987, Collins *et al* 1988). Advance letters are a relatively inexpensive way to communicate the surveys intent, provide advance notice of a possible phone attempt, and for a few respondents, sufficient motivation to make an active response (as shown by the 6% in the Chicago Pilot who took the initiative to respond to the passive mailing).

Surveyors in the pilot test for the Washington COG household travel survey indicated that respondents who recalled the advance mailing were easier to recruit and thus, served also to boost surveyor morale. It was more pleasant for surveyors to conduct recruitment interviews with motivated and willing respondents. Yet, in a non-respondent follow-up study (NRFS) for that same pilot, we found that the number 1 reason that respondents did not participate in the pilot was that they did not recall receiving information about the survey. This was particularly true among Hispanic and African American NRFS respondents. The annual household diary survey conducted by NuStats for the U.S. Postal Service for the past 7 years has found that the percent of mail going to U.S. households that is non-first class has consistently increased. This means that “junk mail” is more prevalent. The direct mail industry spends much time and money trying to decipher how to make their pieces stand out. This combination of factors

strongly indicates that we have to do a better job making our advance mailing stand out to sampled respondents.

4. **Employ a “household questionnaire” to capture cell-only persons.** According to data collected by the National Center for Health Statistics, 53% of Americans use both a landline and a cell phone; 37% have only a landline; and 8% rely only on a cell phone. Nearly half of the cell-only respondents (48%) are under age 30, compared to 21% in the population as a whole. The cell-only population also includes a higher proportion of minorities, especially Hispanics, according to the Pew Center Research. Cell-only persons have been systematically excluded from RDD frames leading to coverage issues for certain subgroups. The household questionnaire with a no-landline condition is a passive data collection mode (e.g., reliance on mail). Passive modes are not as productive as proactive modes. In the household travel survey pilot for the metropolitan Washington area, household questionnaires are being returned more than one month after the estimated receipt at the household. The pace is controlled by the respondent versus being controlled by the surveying organizations. Yet, this method should not be abandoned. It reaches respondent types that would not otherwise participate in the survey.

The Chicago Pilot indicated that the passive mail condition had only a 6% response rate, compared to 17% in the pilot sample as a whole. Household questionnaires in the Washington metro area pilot were returned at a rate between 5% and 8%. But, this does not mean that the household questionnaire is not functional. Whereas the pilot data for the Chicago pilot indicated that the passive sample did not differ significantly from the RDD on variables of interest, the pilot data for Washington revealed that they did. The difference is that the Chicago passive condition included address-based sample for which a telephone match was achieved. In Washington, we were able to disentangle the telephone matched sample from non-telephone matched. When this was done, we found that the non-telephone matched address-based sample that relied solely on the household questionnaire for participation reached very different types of households than did the RDD sample – cell-only persons in their 20s and 30s. These persons also tend to be zero-vehicle households.

5. **Plan for and conduct follow-up activities for hard to reach populations.** Household travel surveys, like any survey, are concerned about unit non-response—that is, entire sample units (households) that may be under-represented in the final survey because of lack of participation. Non-respondent follow-up studies in the Portland area (2005) and in Washington DC area (2006) have shown that respondents in household travel surveys differ in their characteristics and travel behavior from non-respondents, indicating a potential for bias in the estimates of travel that result from household travel surveys. These studies have shown that there are differences in persons who are contacted via telephone or in-person during the follow-up activities as well as differences between non-respondents who are “non-contacts” versus “refusers”. These studies have indicated that this can be sufficient return on investment for intense efforts to garner the participation of harder-to-reach groups if that is an important goal of the travel survey effort, such as with the Chicago survey.

The results of the MWCOC telephone NRFS, like the Portland non-response research, underlined the importance of including refusers in the re-contact sample. Thirty-six percent of the NRFS samples were coded as first refusals in the main survey, but 53 percent of the NRFS telephone completes came from this group. Perhaps refusal to participate is primarily a product of timing, and at another time (or with a little incentive) refusers can be convinced to become participants. We are not so lucky with non-contacts. Although 41 percent of the NRFS sample were non-contacts in the main survey (answering machine and no answer), only 26 percent of

the completes came from this group. Since this group had a lower mobility measure (2.6 places visited compared to 3.5 for refusers), we can assume that call screening is taking place, which might be a persistent habit rather than a temporal effect. This indicates that intense efforts (i.e., more call attempts) are necessary to reach these persons. The characteristics of the people reached through follow-up activities include minorities, younger persons, smaller households, and cell phone users. The characteristics of people reached through the in-person effort, compared to the telephone follow-up, importantly includes even more African-American and Hispanic households, cell-phone only households, younger apartment renters, with fewer vehicles available and overall lower mobility.

6. **Use a GPS subsample for information value.** GPS provides a rich database that has value for trip auditing, model validation and other related analyses. Technology challenges related to GPS satellite reception in urban canyons limit the ability of GPS to play a major role in this survey. Surveys across the country are using advanced technology to assess travel behavior and CMAP should become more experienced with the benefits and limitations. The following recommendations are provided regarding the GPS sample:

- Up to 5% of diary households should be provided GPS devices for the purpose of auditing reporting accuracy with respect to trip rates, trip times, and trip origin and destination locations.
- GPS households should be area-selected to households located outside of the urban core.
- GPS devices should be vehicle-based to ensure quality. Wearable devices have power and size limitations that impact their usefulness.
- GPS data should be collected for one full week. There are no technology limitations to this approach, and the ability to evaluate the number of household trips, VHT, VMT, and route choice over an entire week provides a wide range of data valuable for calibrating and validating model results.
- The data processing should include link-matching to the model network to evaluate speeds, delays, and route choices.

These recommendations maximize useful results while minimizing the risks associated with deploying advanced technology in the field.





## REFERENCES

---

Adler, T., L. Rimmer, and D. Carpenter (2002), "Use of Internet-Based Household Travel Diary Survey Instrument," *Transportation Data and Information Technology Research, Journal of the Transportation Research Board*, No. 1804, pp. 134-143.

Bierlaire, M., Frejinger E., Route Choice Models with Subpath Components, Conference paper STRC 2005, Session Transport Modelling II, STRC 5 th Swiss Transport Research Conference, Monte Verità / Ascona, March 9-11, 2005

Collins, M., W. Sykes, P. Wilson, and N. Blackshaw (1988), "Non-response: The UK Experience," In *Telephone Survey Methodology*, John Wiley and Sons, New York.

Couper, M.P., Rowe, B. (1996). Evaluation of a Computer-Assisted Self-Interview Component in a Computer-Assisted Personal Interview Survey. *The Public Opinion Quarterly* 60 (1), 89-105.

de Leeuw, E. D. (2005). To Mix or Not to Mix Data Collection Modes in Surveys. *The Journal of Official Statistics*, 21(2), 233-255.

de Leeuw, E. D. (1992). *Data Quality in Mail, Telephone, and Face-to-face Surveys*. Amsterdam: TT-Publicaties.

Dillman, D. (2002), "Navigating the Rapids of Change: Some Observations on Survey Methodology in the Early Twenty-First Century," *Public Opinion Quarterly*, Volume 66, Number 3, pp. 473-494.

Dillman, DA (1978). *Mail and Telephone Surveys: The Total Design Method*, New York, John Wiley and Sons.

Fowler, F.J.Jr., Roman, A.M., and Di, Z.X. (1998). Mode Effects in a Survey of Medicare Prostate Surgery Patients. *The Public Opinion Quarterly* 62 (1), 29-46.

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., and Tourangeau, R. (2004). *Survey Methodology*. New Jersey: John Wiley and Sons.

Groves, R.M., Singer, E., and Corning, A. (2000). Leverage-Salience Theory of Survey Participation: Description and an Illustration. *The Public Opinion Quarterly* 64 (3), 299-308.

Groves, R., R. Cialdini, and M. Couper (1992), "Understanding the Decision to Participate in a Survey", *Public Opinion Quarterly*, Vol. 56, No. 4, Winter, pp. 475 -95.

Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley & Sons, Inc.

Marchal, F., Hackney, J. , Axhausen, K.W., Efficient map-matching of large GPS data sets - Tests on a speed monitoring experiment in Zurich, *Transportation Research Record*.

Massey, J. (1988), "An Overview of Telephone Coverage," *Telephone Survey Methodology*, John Wiley & Sons, New York, pp. 3-8.

Moon, Y (1998). Impression Management in Computer-Based Interviews: The Effects of Input Modality, Output Modality, and Distance. *The Public Opinion Quarterly* 62 (4), 610-622.

Presser, S. and Stinson, L. (1998). Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance. *American Sociological Review* 63, 137-145.

Schwarz, N., Strack, F, Hippler, H.J., and Bishop, G. (1991) The Impact of Administration Mode on Response Effects in Survey Measurement. *Applied Cognitive Psychology* 5, 193-212.

Tourangeau, R., Rips, L.J., and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Traugott, M. (1987). "The Importance of Persistence in Respondent Selection for Preelection Surveys." *Public Opinion Quarterly*. Vol. 51, pp. 48-57.

Triplett, T. (1998). "What is Gained from Additional Call Attempts and Refusal Conversion and What are the Cost Implications?" Unpublished paper, University of Maryland, Survey Research Center.

Wolf, J., M. Oliveira, and M. Thompson (2003). The Impact of Trip Underreporting on VMT and Travel Time Estimates: Preliminary Findings from the California Statewide Household Travel Survey GPS Study. *Transportation Research Record No. 1854*, pp. 189-198.

Wolf, J., R. Guensler, and W. Bachman (2001). "Elimination of the Travel Diary: Experiment to Derive Trip Purpose from Global Positioning System Travel Data," *Transportation Data and Information Technology, Journal of the Transportation Research Board*, No. 1768, pp. 125-134.

Zmud, J, and Arce, C. (1997) "*Item Non-response in Travel Surveys: Causes and Solutions.*" Published in Conference Proceedings, *Transport Surveys: Raising the Standard*, International Conference on Transport Survey Quality and Innovation, Grainau, Germany, May 1997.

Zmud, J., H. Contrino, and C. Frye (2003). "Mixed-mode Surveys: Utility of Web Options for Household Travel Surveys". Paper presented at the *Application of Transportation Planning Methods Conference*, Baton Rouge, Louisiana.

Zmud, J. (2003) "*Designing Instruments to Improve Response*" in *Transport Survey Quality and Innovation*. Edited by P. Stopher and D. Jones. London: Pergammon Press.

Zmud, J. (2004) "*Instrument Design: Decisions and Procedures*". Commissioned resource paper for the *International Conference on Transport Quality and Innovation*, Costa Rica.