Using GPS Data to Inform Travel Survey Methods

Stacey Bricka
Research Director and Principal, NuStats
PhD Student, The University of Texas at Austin
Phone: 512-306-9065, ext 2240
Email: sbricka@nustats.com
(corresponding author)

Chandra R. Bhat
The University of Texas at Austin
Dept. of Civil, Architectural & Environmental Engineering
Phone: 512-471-4535
E-mail: bhat@mail.utexas.edu

While we continue to work towards the long-term goal of using GPS technology to produce higher quality trip files, the reality is that the current method of random samples, telephone surveys, and travel logs continue to be used. Thus, for any given regional travel survey, trip underreporting will occur at some level. The research question that forms the focus of this paper is whether an analysis of GPS data collected as part of a regional travel survey can be used to minimize trip under-reporting through improved survey methods. The focus is on demographic characteristics, travel behavior characteristics, and indicators of adherence to survey protocol that potentially impact trip under-reporting. The results suggest that while more research into this subject is warranted, there are specific, low-cost changes to the survey materials as well as the interviewing process that can be made immediately to reduce trip under-reporting.

Introduction

Ten years ago, the transportation community began in earnest an investigation into the application of global positioning system (GPS) technology to travel survey data collection efforts. The immediate focus of this technology application has been to improve the quality of travel survey data, with a long-term goal of eventually replacing respondent-reported data with travel details collected passively through these devices. The main application of GPS in regional travel surveys to date has been for auditing trip reporting, to determine the level of trip under-reporting by vehicle drivers and develop appropriate correction factors for the data. Specifically, GPS has been used in 12 regional travel surveys: Lexington (1996), Austin (1997), California (2001), Los Angeles (2001), Pittsburgh (2001), St. Louis (2002), Ohio (2002), Laredo (2002), Tyler/Longview (2003), Kansas City (2004), Reno (2005), and in a pilot test for the upcoming Oregon statewide travel survey (2005). In addition, other GPS studies not directly linked to regional travel surveys have employed GPS for speed studies and in testing the development of trip tables solely from GPS data. For purposes of this paper, references to "GPS studies" refer to those conducted as part of regional travel surveys only. In conducting these studies, we have gleaned several important facts:

- Respondents who self-select to participate in GPS travel studies are different from those who do not elect to participate. As documented in several travel survey reports, GPS participants tend to report higher incomes and own their own homes as compared to those who elect not to participate.[1] Thus, most of the findings to date and conclusions regarding trip under-reporting are based on a select group of respondents and not general populations of entire regions.
- The methods used to process the GPS data streams vary across the GPS studies conducted to date, and influence the degree of trip reporting detected. Some studies, such as the Los Angeles study, used in-vehicle devices to capture trip details for both drivers and passengers, while others focused only on vehicle drivers. In addition, as shown in an early analysis of the Austin data, the time thresholds used in vehicle movement detection can cause the trip under-reporting rate to vary greatly (in Austin the rate is 12% or 31%, depending on the time threshold).
- The actual data collection methods and instructions to respondents can also influence the calculation of trip under-reporting rates. In most studies, respondents are instructed to not report trips out of the

[1] See, for example, the Kansas City Regional Household Travel Survey GPS Final Report (written by NuStats for the Mid-America Regional Council, 2004).

geographically defined study area and trips for commercial purposes. However, most early trip-detection algorithms did not distinguish between these types of trips and those reported by respondents, resulting in over-reported trip underreporting rates. In addition, as was determined in Laredo, the survey process does not directly collect information about trips made by non-household members driving the GPS-equipped vehicles.

Based on a review of literature on trip underreporting in regional household travel surveys and the development of associated correction factors, most trip under-reporting is associated with households that own 3 or more vehicles, households with incomes of less than $50,000, and respondents under the age of 25. From a travel behavior perspective, respondents who travel substantially make several short trips (less than 5 minutes), and make trips of a discretionary nature are most likely to "forget" to record this travel (as has been suggested on parallel literature regarding trip-chaining).

The studies to date have clearly aided in identifying factors associated with trip under-reporting in regional travel surveys. In this paper we contribute to this existing literature and continuing discussion regarding GPS technology in travel surveys in several ways. First, in the current study (and unlike earlier studies), we model both the presence of trip under-reporting by an individual as well as the level of trip under-reporting by the individual. The separation of the presence of trip under-reporting from the level of trip under-reporting recognizes that different explanatory variables may affect these outcomes and/or that the same explanatory variable may affect these outcomes differently. Second, the joint model also recognizes that the likelihood of trip under-reporting and the level of trip under-reporting may be related to one another. For example, it is conceivable (if not very likely) that individuals who are, by nature, less likely to be responsive to surveys are the ones who under-report *and* under-report substantially. Similarly, individuals who are, by nature, very interested in the survey would be the ones less likely to under-report at all, and even if they did under-report, will do so only marginally. Third, in addition to jointly modeling trip under-reporting and the level of trip under-reporting, the empirical analysis in the current study considers a comprehensive set of variables related to driver demographics, driver travel characteristics, and driver adherence to survey protocol. Finally, we translate our empirical analysis results to recommendations regarding household travel survey procedures to reduce the magnitude of trip under-reporting in future travel surveys conducting using the traditional approach.

GPS and Travel Survey Data

The empirical analysis in the current paper uses data extracted from the Kansas City Regional Household Travel Survey that was conducted in Spring 2004, under the sponsorship of the Mid-America Regional Council and the Kansas and Missouri Departments of Transportation. As part of the Kansas City survey, complete demographic and travel behavior characteristics of 3,049 randomly sampled households were obtained, including details about 32,011 trips for 7,570 household members. The GPS component of the study involved equipping the vehicles of 294 households with GPS equipment to record all vehicle travel during the assigned travel period. Of the 294 households, both CATI and GPS data are available for 228 households. All subsequent analyses in the current paper focus on these 228 households, corresponding to 377 drivers and 2,359 vehicle trips. (For more details regarding the characteristics of these GPS households as compared to the general survey participants as a whole, see NuStats 2004).

Of the 377 drivers, 269 (or 71 percent) accurately reported all travel in their CATI survey, while 108 (or 29 percent) had at least one instance of a trip that was not reported.[2] Among the 108 respondents

[2] A subtle, but important, point needs to be noted here. For our under-reporting analysis, we focused on the CATI-reported vehicle trips across all individuals in the household who drove each GPS-equipped vehicle. This allows a fair comparison between the CATI-reported vehicle trips and the GPS-detected vehicle trips. However, rather than confine the analysis of the determinants of under-reporting to household-level characteristics, we also included person-level characteristics to accommodate person-specific tendencies to under-report. To accomplish this, we identified a primary driver for each GPS-equipped vehicle based on the information provided by respondents, and used these primary driver characteristics as explanatory variables in the analysis (along with household

who under-reported, 53 (49%) missed one trip, 22 (20%) missed two trips, 11 (10%) missed three trips, 6 (5.5%) missed 4 trips, and 16 (14.5%) missed 5 or more trips. There was a narrow, long, tail in the ≥ 5 missed trips category with one individual under-reporting 17 trips.

A comparison of the CATI-reported trips with the GPS-detected trips identified 280 GPS-detected trips that were not reported by the drivers in the CATI travel survey.[3] A descriptive analysis of trip under-reporting by driver demographics, driver travel characteristics, and driver adherence to survey protocols was undertaken. The results related to demographic characteristics suggest that drivers between the ages of 50 and 69, who are male, with low education levels, who are not employed or employed in sales/clerical occupations, working at locations characterized as "residential," from single-adult or retired households, from 1- or 3-person households, and from 3+ vehicle households are the most likely to under-report trips. The driver travel characteristics indicate that drivers who make a relatively large number of total trips during the survey day, pursue long distance trips, and undertake trip-chaining on the survey day are over-represented in the pool of those who under-report. Finally, in the category of driver adherence to survey protocols, the results suggest that drivers who do not use their travel diaries for recording travel and who have their travel details reported by proxy are more likely to under-report.

These descriptive statistics provide suggestive evidence of the effect of various driver attributes on the propensity to under-report trips. However, these are uni-dimensional statistics in that they do not control for the influence of other variables when examining the impact of any single variable. For instance, the gender difference in under-reporting may be a manifestation of different travel patterns of men and women. Further, the descriptive analysis does not focus on the characteristics impacting the level of trip under-reporting. To obtain a comprehensive picture of the factors affecting whether an individual under-reports and the level of under-reporting, it is necessary to pursue a multidimensional and comprehensive analysis that examines the effects of all potential determinants of both under-reporting propensity and the level of under-reporting propensity. In the next section, we present the model structure and empirical analysis for such a methodology.

Model summary

The approach adopted in this study utilized two equations – a binary model for whether an individual under-reports or not and an ordered-response model for the number of trips under-reported if there is under-reporting at all. The methodology it accounts for the correlation in error terms between the two equations. That is, it accounts for the potential presence of unobserved individual factors (such as, say, an overall disinclination to respond to surveys or substantial time constraints) that influences both whether an individual under-reports as well as the level of under-reporting. For a more detailed discussion of the modeling portion of this research, see Bricka and Bhat (2005).

The fundamental hypothesis underlying our empirical analysis was that trip under-reporting is largely due to three areas of influence: who the driver is (driver demographics such as household type, age, number of household vehicles, employment status, *etc.*), the characteristics of trips made (total number of trips, average distance of trips, and level of trip-chaining), and how well the driver adhered to the survey protocol (whether driver uses the travel diary to record all travel and whether driver talked directly with interviewer). All exogenous inputs to the model were classified according to these broad categories. The final variable specifications for the binary model of under-reporting and the ordered-response model for level of under-reporting among under-reporting individuals were developed by adopting a systematic procedure of eliminating statistically insignificant variables. Of course, as indicated earlier, the entire specification effort was also informed by the results of earlier studies and intuitive considerations.

---

demographics). This is reasonable because each vehicle in this study was predominantly used by only one "primary" driver in the household (especially within a short period of time, such as a survey day). Specifically, in the sample used for our analysis, there was car-sharing of some form among household members in 6% of all households.

[3] The extraction of trips from the GPS traces was based on a multi-level trip detection algorithm developed by GeoStats, with several built-in checks to avoid "ghost trips" (such as due to starts and stops at street lights). The full details of the GPS processing are available in the study report by NuStats (2004).

Results and Implications for Survey Methods.

The results from our modeling effort provide important insights regarding under-reporting tendencies in traditional household travel surveys. First, the underlying mechanism that represents whether an individuals under-reports or not is different from the mechanism that determines the level of under-reporting. At the same time, there are common unobserved factors that influence both the under-reporting propensity and the propensity associated with the level of under-reporting. Consequently, it is important to use the joint binary-unordered response framework of the current study to analyze trip under-reporting and its magnitude. Second, the effect of driver demographics indicates that young adults (less than 30 years of age), men, individuals with less than high school education, unemployed individuals, individuals working in clerical and manufacturing professions, workers employed at residential, industrial, and medical land-uses, and individuals in nuclear families are all more likely to under-report trips in household travel surveys than other respondents. Third, driver travel characteristics that affect the tendency to under-report include making a high number of trips on the survey day, traveling long distances per trip, and trip chaining. Fourth, drivers who do not use the travel diary to record their travel are more likely to miss trips than those who use the travel diary, and proxy-reporting leads to more missed trips.

The model results can be used to identify specific improvements in the methods to conduct future travel surveys. These improvements may include (1) the use of special survey materials for respondents who travel more than usual or who are under the age of 30 and (2) developing better probes in telephone interviews when collecting information from unemployed individuals, proxy reporters, and individuals who travel longer than average distances. Each of these potential improvements is discussed below.

Use of Special Survey Materials. The empirical results from this study indicate that an important predictor of trip under-reporting is the extent to which a respondent travels. Those who travel more have a higher propensity to under-report trips. This empirically supports the findings of prior studies, most of which related the increased travel to heavier respondent burden (and thus suggested missed trips were the respondent's way of ending the survey interview early). While the relationship between respondent burden and trip under-reporting is well accepted, there is another component to this relationship that should be considered – the design of the travel log.

The travel logs used in the Kansas City study allowed space for recording up to 10 trips and instructed respondents to record additional travel on paper. The limit of 10 trips was based on the fact that most people report an average of 5 person trips in a day. In addition, it allows for a portable-sized log when printed. It works well for "normal" or "light" travelers who typically have room in their travel diaries at the conclusion of the travel day. It is possible that the "heavy" travelers only record up to the space in the log and nothing more (while the GPS unit continues to detect trips for the remainder of the travel day). The problem may be further compounded if the data are then reported by proxy – the person reporting travel for the heavy traveler may read the 10 trips from the log, and, not knowing what other travel was made that day, end the travel day prematurely. Additional study is warranted to determine the characteristics of heavy travelers such that they can be pre-identified in the recruitment interview and provided a special log with either additional pages or a special insert for recording the additional trips (similar to how special instructions regarding transit trip recording are provided to 0-vehicle households presently). This is a relatively low-cost solution that would help to minimize trip under-reporting from the heavy traveler group of respondents.

A second important driver characteristic is age. This study reveals that the propensity to under-report travel decreases with age. Thus the worst trip-reporters are those respondents under the age of 30. We recommend that future travel surveys consider the funds to conduct cognitive interviews or focus groups targeted specifically toward younger drivers. The purpose of this qualitative research would be to identify specific methodological improvements to the survey instruments that would result in better capture of travel from this age group. It may be possible, for example, that this group is more impatient with the telephone interview format and more receptive to self-reporting their travel via an Internet based retrieval tool or simply being encouraged to return their logs by mail, with telephone follow-up as needed.

Finally, most travel survey materials are designed for persons with an 8th grade education. However, this study found that respondents with less than a high school education are very likely to under-report their travel. This finding is independent of the age effect (*i.e.*, a continued reflection of being under age 30). While most of the respondents reporting the lowest education level were under the age of 30 and still in high school (67%), one-third reported ages from 32 to 82. Further investigation is warranted to identify improvements in survey materials so that individuals with a low education level can understand what travel to report and how to record the travel as part of the survey. Different approaches may likely be needed based on whether the respondent is still in high school or in a later stage of life.

Developing Better Probes. Based on the findings of the earlier GPS studies, it has become standard procedure to probe workers about potential stops made during their commutes. In addition, as a form of validation, respondents who report no travel are subjected to a series of questions to confirm the legitimacy of the reporting. The results of this study suggest that additional probes as part of the travel retrieval interview may be warranted for all travelers, not just workers or those who report no travel.

Specifically, this study indicates that there is a high propensity to under-report travel if the driver is unemployed, has his/her travel data reported by proxy, or travels long distances. The finding that unemployed drivers have a higher tendency to under-report trips is a new correlate to be considered. In the past, the modeling focus on the work trip (and how discretionary travel may be incorporated into the work commute) has led to an emphasis on collecting travel/activities that occur during the lunch break or during the commute to/from the workplace. Drivers who are unemployed do not receive similar levels of scrutiny or probes, but should according to the findings of this study.

Unlike employment status, the finding that proxy-reported travel is associated with higher propensities of under-reported travel is well documented. While the most obvious solution is not to allow any proxy reporting, the cost implications of such a decision are tremendous and may introduce more bias into the survey data than that introduced by allowing proxy reporting. A second, but also costly, approach is to only allow proxy interviews if the travel log is used. The better solution here may be to strengthen the telephone interview in a manner similar to the recommendation above for strengthening the travel of unemployed persons.

In summary, this paper has examined the driver demographics, driver travel characteristics, and driver adherence to survey protocol considerations that impact the likelihood of under-reporting as well as the level of trip under-reporting. These results can be used to adjust for under-reporting in traditional household travel surveys and/or to improve travel survey data collection procedures. Although we do plan to replicate this analysis on future travel surveys with GPS components, we believe that the survey method improvements identified in this study will enhance the collection of complete trip information in any household travel survey.

REFERENCES

Bricka, S. and C.R. Bhat, "*A Comparative Analysis of GPS-Based and Travel Survey-based Data*," Technical paper, Department of Civil Engineering, The University of Texas at Austin, July 2005.

NuStats. *Kansas City Regional Household Travel Survey: GPS Study Final Report*. Mid-America Regional Council, Kansas City, 2004.