A Comparative Analysis of GPS-Based and Travel Survey-Based Data

Stacey Bricka
Research Director and Principal, NuStats
PhD Student, The University of Texas at Austin
3006 Bee Caves Rd, Ste A300, Austin, TX 78746
Phone: 512-306-9065, ext 2240
Fax: 512-306-9077
Email: sbricka@nustats.com
(corresponding author)

and

Chandra R. Bhat
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: bhat@mail.utexas.edu

TRB 2006: For Presentation and Consideration for Publication
Paper # 06-0459
Re-Submitted on: March 31, 2006
Word Count: 7,324 + 4 tables = 8,324

ABSTRACT

This paper examines the driver demographics, driver travel characteristics, and driver adherence to survey protocol considerations that impact the likelihood of under-reporting in a household travel survey. The research considers both the likelihood of vehicle driver trip under-reporting as well as the level of vehicle driver trip under-reporting using a joint binary choice-ordered response discrete model. The empirical analysis uses the Global Positioning System (GPS)-equipped sample of households from the 2004 Kansas City Household Travel Survey who also provided travel diary information.

The empirical results provide important insights regarding under-reporting tendencies in household travel surveys. In particular, 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 land-uses, individuals who make many trips, travel long distances and trip-chain, and respondents who fail to use a travel diary to log their travel before telephone retrieval of their patterns are associated with higher under-reporting. Also, the underlying factors influencing whether an individual under-reports or not are different from the factors impacting the level of under-reporting.

1. INTRODUCTION

1.1 Background

An analysis of regional travel behavior characteristics is instrumental in developing travel demand models, guiding long-range transportation planning, and answering region-specific mobility questions. For more than fifty years, such regional travel behavior characteristics have been documented and analyzed through the design and administration of household travel surveys. The methods used to undertake household travel surveys have progressed from large-scale in-person interviews conducted with clipboards and pencils to smaller-scale random computer-aided telephone interviews (CATI), and from a simple recall of travel "yesterday" to the advance provision of diaries for recording travel throughout the day.

The transition from large-scale person interviews to smaller-scale computer-aided interviews has been accompanied by a greater emphasis on collecting comprehensive and accurate travel information from respondents. This has increased respondent burden, which is reflected in lower participation rates and higher refusal rates. These, in turn, impact the cost and quality of the survey data. Increased respondent burden is also reflected in the levels of completeness and accuracy of the data obtained from participating households, the very areas that the survey method improvements originally sought to strengthen. Thus, practitioners question whether the increased burden associated with efforts to obtain detailed information on travel-related activities for a 24-hour period has resulted in respondents purposefully or inadvertently not reporting all travel.

In the context of non-reported trips due to respondent burden, a missed trip may initially appear to be a minor aberration for travel modeling. However, one missed trip of a single individual can be magnified to the order of between 200 and 500 trips when the survey sample is expanded to reflect the survey universe. These missed trips can lead to an underestimation of the regional levels of vehicle miles traveled, particularly if the missed trips are complete round trips or multi-stop tours. For instance, Wolf *et al.* (*1*) modeled the impact of missing trips in Sacramento, Alameda County, and San Diego using the regional travel demand models. They found that missed trips resulted in up to 40% under-reporting of VMT estimates (calculated as the differences in modeled VMT when using the survey data trips vs. using Global Positioning System (GPS) detected trips). In addition, in the context of activity-based travel modeling, missed trips can result in the incorrect depiction of a household's overall activity-travel pattern over the day, resulting in mis-estimated activity-travel models.

In order to better understand vehicle driver trip under-reporting in household travel surveys, some studies have relied on GPS technology to track the vehicular travel of participating households. Basically, a subset of households participating in the travel survey is provided a GPS unit for each household vehicle. The unit stays in the vehicle throughout the assigned 24-hour travel period, recording all vehicle movement. At the same time, household members record their travel in conventional logs. The GPS navigational data streams are downloaded and processed into trips, while the household-recorded travel is retrieved using CATI. Differences between the GPS-detected and CATI-reported trips are examined, and the trips detected in the GPS data but not in the CATI data are used to estimate the level of trip under-reporting in a given dataset.

To date, and as just discussed, the main application of the GPS data has been for the purpose of detecting vehicle driver trip under-reporting levels in household travel survey datasets. These trip under-reporting levels are used to create adjustment factors that serve to account for the missed trips in the travel surveys [see Zmud and Wolf (*2*) for an example of how these adjustment factors are created]. However, there has been little effort to examine trip under-reporting from the vantage point of improving the travel survey methods.

1.2 Paper Objective

The primary objective of this paper is to determine whether driver demographics, travel characteristics, and driver adherence to survey protocol (*i.e.*, how well drivers adhere to the spirit of the survey protocols) correlate with missed vehicle driver trips. (In the rest of this paper, we will use the term "missed trips" to refer to missed vehicle driver trips). In addition to determining the correlates of missed trips, we identify ways in which survey instructions and materials can be improved such that respondents better understand the survey task and more accurately report their travel.

The analysis of the factors affecting trip under-reporting is accomplished through the formulation of a joint model for the presence of trip under-reporting and the level of trip under-reporting. The joint model is estimated using the GPS-equipped sample of households in the 2004 Kansas City Household Travel Survey (who also provided travel diary information).

The rest of this paper is structured in five sections. The next section provides a summary of GPS-related findings to date, while Section 3 presents an overview of the Kansas City GPS effort and its descriptive sample characteristics. Section 4 discusses the model structure and estimation procedure. Section 5 focuses on the empirical results. The final section summarizes the important findings from the results, and recommends specific improvements in travel survey methods to alleviate the trip under-reporting problem.

2. GPS IN HOUSEHOLD TRAVEL SURVEYS

To date, there has been ten U.S. travel surveys that have included a GPS component for the express purpose of identifying levels of trip under-reporting. This includes the "proof of concept" study in Lexington, two statewide travel surveys (Ohio and California), and regional travel surveys in Austin, Pittsburgh, St. Louis, Los Angeles, Laredo, Tyler/Longview, and Kansas City. The Lexington and Austin studies were conducted in the mid-1990s, while the remaining studies were conducted between 2000 and 2004 (see Table 1 for further details of, and reference sources for, each GPS study).

As can be observed from Table 1, and excluding the "proof of concept" Lexington study, the number of households that participated in the GPS studies varies from a low of one percent of total CATI surveyed households (Los Angeles) to a high of 11 percent (Tyler/Longview). The average size of the GPS sample across these studies was 5% of the CATI surveyed households. In six of the ten GPS studies, the GeoStats GeoLogger was used to collect and record data on vehicle movements (*13*). For three others, the Battelle GPS Leader was used (*13*). In the 1997 Austin study, NuStats developed the GPS equipment. In addition to using different equipments, the processing of the GPS data streams has varied across the studies, which limits cross-study comparisons.

The levels of trip under-reporting estimates range from a low of 10% in Kansas City to a high of 81% in Laredo. Obviously, the thresholds and assumptions used to process the GPS navigational streams have a substantial impact on the final trip under-reporting rate, as does the availability of variables to help detect whether the vehicle was driven by someone other than a household member and the screening of the GPS data to exclude (from the trip detection process) any travel that was not recorded as per respondent instructions in the CATI survey (for instance, several surveys ask respondents only to record travel in the study area and not to record commercial travel). Documentation is not consistently available to provide a clear understanding of how the data were processed and the trips detected in the studies listed in Table 1. Thus, a direct comparison of results across studies is not appropriate.

Several of the GPS studies listed in Table 1 were conducted with the express objective of detecting levels of trip under-reporting, as indicated in Section 1.1. As a result, the final reports of these studies focus on the methods used to obtain and process the GPS data. However, a few of the reports also include some discussion regarding the determinants/correlates of trip under-reporting. Of the ten studies listed in Table 1, five are of direct interest to the current study in the context of understanding the factors that influence trip under-reporting. These are the California Statewide, Los Angeles, St. Louis, Kansas City, and Ohio Statewide studies. The results from these studies are briefly discussed in Sections 2.1 through 2.5

2.1 The California Statewide Household Travel Study

In the California Statewide study, a binary logit model was developed to identify the contribution of key household demographics to trip under-reporting. The demographic variables found to significantly associate with trip under-reporting included households with 3+ vehicles, households with annual income less than $50,000, households with 3+ workers, and adults less than 25 years of age (*2*). In addition, a separate analysis of the GPS data found that the greatest "offenders" in terms of the magnitude of trip under-reporting were the heaviest travelers, consistent with prior research on the impact of respondent burden on survey data completeness (*15*).

2.2 The Los Angeles Travel Study

In this study, a binary logistic regression was developed to identify the variables associated with trip under-reporting. The results indicated, as in the California study, that individuals in households with an annual income less than $50,000 and adults less than 25 years of age were more likely to under-report. Also, the study found that short trips (of duration less than 5 minutes) were more likely to be missed than other trips (*5*).

2.3 The St. Louis Household Travel Study

The development of a trip correction factor for the St. Louis study also utilized a binary logit model. The results were similar to the earlier two studies in the effect of household vehicle ownership, household income, and age of respondent. As in the Los Angeles study, the results also indicated higher under-reporting of short duration trips (*8*).

2.4 The Kansas City Household Travel Study

In Kansas City, a binary logit regression model was again employed to investigate the demographic variables correlated with trip under-reporting. The key characteristics associated with trip under-reporting were: household size (1 and 3 person households in particular), households with 3+ vehicles, households with incomes less than $50,000 or greater than $100,000, and respondents under age 25 (*13*).

2.5 The Ohio Statewide Travel Study

The Ohio Statewide GPS results were analyzed differently than the four studies summarized above. In this study, the GPS equipment was the Battelle GPS Leader, which comprised both a GPS receiver and a PDA for the vehicle operator to enter trip details. As a result, the approach to determine trip under-reporting differed from that used in other studies (*9*). Specifically, the sample was categorized into three groups: (1) All Households with GPS Data, (2) Households with both GPS and Diary Data (this group was a subset of the first group), and (3) Households with no GPS data. Because the demographics varied across the three groups, the results were weighted to 2000 census parameters prior to comparisons. Household level trip rates were calculated and compared based on demographics, day of week, and trip purpose. The estimates of trip under-reporting were made "by comparing the average vehicle and person trip rates" (*9*). The study found that trip under-reporting "was more prevalent in one- and two-person households, households with fewer vehicles, and low-income households." In addition, discretionary trips were found to be more likely to be under-reported than non-discretionary trips.

2.6 Summary of Earlier GPS-based Under-reporting Studies and the Current Paper

The emphasis of the current study is on examining the influence of driver demographics, driver trip characteristics, and driver adherence to survey protocols on trip under-reporting. Accordingly, we summarize the results of earlier studies of trip under-reporting by each of these three variable categories below.

A relatively consistent finding among the studies discussed above is that trip under-reporting is most closely associated with the following demographic variables: households that own more vehicles (3+), households with incomes of less than $50,000, and respondents under the age of 25.

The trip characteristics found to impact trip under-reporting in the earlier studies are total trips, trips of short duration (less than 5 minutes) and trips of a discretionary nature. The effect of the first trip variable, total trips, is as expected and can be attributed to respondent burden. The effect of the second and third variables (short trips and discretionary trips) may be attributable to under-reporting associated with trip chaining. In particular, a growing body of literature has found that trip chaining is often associated with short trips for discretionary purposes [see McGuckin (*16*), Levinson (*17*), and Taylor (*18*)].

Finally, the effect of driver adherence to survey protocol on trip under-reporting was not addressed in the GPS studies. However, all non-GPS studies to date have relied on an interview status variable (proxy or in-person reporting) as an explanatory variable in studies of trip under-reporting. In all cases, proxy reporting was found to be associated with lower trip reporting as compared to that obtained from in-person interviews [see for example Badoe (*19*), Kostyniuk *et al.* (*20*), Wargelin and Kostyniuk (*21*)].

The studies to date have clearly aided in identifying the factors associated with trip under-reporting in household travel surveys. In this paper we contribute to this existing literature in several ways. First, in the current study (and unlike earlier studies), we model both the likelihood 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.

3. THE KANSAS CITY HOUSEHOLD TRAVEL SURVEY GPS COMPONENT

3.1 Background

The empirical analysis in the current paper uses data extracted from the Kansas City 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, the reader is referred to the study's final report by NuStats) (*13*).

3.2 Descriptive Analysis of the Sample

As indicated above, the Kansas City GPS data set contained details on 377 drivers and 2,359 vehicle trips. 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.[1] Among the 108 respondents 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.[2] A descriptive analysis of trip under-reporting by driver demographics, driver travel characteristics, and driver adherence to survey protocols is presented in Table 2. The results related to demographic

---

[1] 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 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.

[2] 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 starts and stops at street lights). The full details of the GPS processing are available in the study report by NuStats (*13*).

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 in Table 2 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 trips. Finally, in the category of driver adherence to survey protocols, the results in Table 2 suggest that drivers who do not use their diaries for recording travel and who have their travel details reported by proxy are more likely to under-report trips.

The descriptive statistics in Table 2 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 in Table 2 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.

4. MODEL STRUCTURE AND ESTIMATION

The approach adopted in this study uses two equations, one for whether an individual under-reports or not and the other for the number of trips under-reported when there is under-reporting. In addition, 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 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. The model system is as follows:

$$u_i^* = \gamma' X_i + \varepsilon_i,\ u_i = 1 \text{ if } u_i^* > 0 \text{ and } u_i = 0 \text{ if } u_i^* \leq 0 \tag{1}$$
$$N_i^* = \alpha' Z_i + \eta_i,\ N_i = j \text{ if } a_{j-1} < N_i^* \leq a_j,\ j = 1,2,\ldots,J,\ N_i \text{ observed only if } u_i^* > 0,$$

where i is an index for individuals, u_i is an observed binary variable indicating whether or not a person under-reports (u_i =1 if person under-reports, 0 otherwise), u_i^* is an underlying under-reporting propensity related to u_i as shown above, N_i is an observed ordinal variable representing the number of trips that individual i under-reports, N_i^* is a latent continuous variable representing the under-reporting propensity underlying the frequency of missed trips, the a_j 's represent thresholds that relate N_i^* to the observed variable N_i in the usual ordered-response structure, X_i and Z_i are vectors of explanatory variables, γ and α are corresponding vectors of parameters to be estimated, and ε_i and η_i are normal random error terms assumed to be identically distributed across observations with a mean of zero and variance of 1. As written in Equation (1), X_i includes a constant while Z_i does not. Also, $a_1 = -\infty$ and $a_J = +\infty$.[3]

The error terms ε_i and η_i are assumed to follow a bivariate normal distribution. The probability that a person under-reports and does so by j trips can then be written from Equation (1) as:

$$\text{Prob}\,(u_i = 1,\ N_i = j) = \Phi_2\left(a_j - \alpha' Z_i,\ \gamma' X_i,\ -\rho\right) - \Phi_2\left(a_{j-1} - \alpha' Z_i,\ \gamma' X_i,\ -\rho\right), \tag{2}$$

where ρ is the correlation between the error terms ε_i and η_i, and Φ_2 is the cumulative standard bivariate normal function. We now define a set of dummy variables M_{ij} as below:

$$\begin{aligned} M_{ij} &= 1, \quad \text{if } N_i = j \ (\textit{i.e.}, \ a_{j-1} < N_i^* \le a_j), \text{ and} \\ M_{ij} &= 0, \quad \text{otherwise.} \end{aligned} \tag{3}$$

The appropriate maximum likelihood function for estimation of the parameters in the model system is:

$$L = \prod_{i=1}^{I} \left[\left[1 - \Phi(\gamma' X_i)\right]^{1-u_i} \times \left\{ \prod_{j=1}^{J} \left[\Phi_2\left(a_j - \alpha' Z_i, \gamma' X_i, -\rho\right) - \Phi_2\left(a_{j-1} - \alpha' Z_i, \gamma' X_i, -\rho\right) \right]^{M_{ij}} \right\}^{u_i} \right] \tag{4}$$

The parameters α, γ, and ρ, and the a_j thresholds $(a_1 = -\infty, a_J = +\infty)$, are estimated by maximizing the likelihood function in Equation (4). If the correlation between the error terms (ρ) is zero, Equation (4) simplifies to two independent models, one for the binary model for under-reporting and the other for the number of under-reported trips. In general, ignoring ρ and estimating independent models for under-reporting and number of under-reported trips will lead to biased parameter estimates. The model estimation was pursued using the GAUSS software package. Analytical gradients were coded with respect to the parameters of interest.

5. EMPIRICAL ANALYSIS

5.1 Variable Specification

The fundamental hypothesis underlying our empirical analysis is 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 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,

---

[3] This equation system is the typical joint binary-ordered response structure in the econometric literature [see Popuri and Bhat (*22*) and Misra and Bhat (*23*)].

as indicated earlier, the entire specification effort was also informed by the results of earlier studies and intuitive considerations.

There are two additional points to note here regarding our variable specification process. First, for several continuous variables (such as age, number of trips made, and average trip distance), we tested alternative functional forms other than the linear effect on the underlying propensities associated with under-reporting and the level of under-reporting. For example, we considered piecewise linear effects (using a spline approach) as well as dummy variables for different ranges of the continuous variables. The final functional form was based on intuitive and parsimony considerations. Second, in our specification analysis, we used a t-statistic value of 1.00 as the threshold for retaining variables due to the relatively small GPS sample size and the very small fraction of individuals who under-report. While the use of such a low t-statistic threshold increases the probability of incorrectly including variables, it also reduces the probability of incorrectly rejecting variables. Because of the limited data available in our empirical analysis, we believe a t-statistic of 1.00 is a good balance. Besides, including suggestive variables should aid researchers working with richer data sets in the future to identify the variables associated with the presence and level of under-reporting.

5.2 Estimation Results

The model results are presented in Table 3. The coefficients indicate the effects of variables on the propensity to under-report (under the “Trip Under-reporting” column) and on the propensity underlying the frequency of missed trips for individuals who under-report (under the “Magnitude of Trip Under-reporting” column). In the discussion below, we will not belabor the point that the ordered response model results pertain to the group of under-reporters. The reader will also note that we use the categories of 1, 2, 3, 4, and 5+ missed trips as the ordinal variable in the ordered response model, which leads to the four thresholds in the “Magnitude of Trip Under-reporting” column in Table 3 (see toward the bottom of the table).

5.2.1 Driver Demographic Characteristics

The impact of the driver demographic characteristics indicate that the propensity associated with the level of under-reporting among those who miss one or more trips decreases with age; those with the highest propensity to under-report several trips are those under the age of 30 years, which is consistent with the findings of earlier research (see the studies reviewed in Section 2). Further, individuals who are in the age group of 50-69 years are likely to be over-represented in the pool of individuals who under-report one or more trips.

The “male” variable in Table 3 shows that, as discussed in Section 3.2, men are indeed more likely to under-report than women, even after controlling for other demographic and travel related characteristics. The effect of the next variable, the education level of the respondent, is clear and highly statistically significant; individuals who have less than a high school education are likely to under-report trips, though this variable does not appear to affect the level of under-reporting. The strong effect of this variable suggests that individuals with a low education have difficulty comprehending and properly using the survey instrument.

The employment status variables point to a higher propensity in the level of trip under-reporting among unemployed individuals, which may be related to the higher number of discretionary trips and shorter duration trips pursued by unemployed respondents. The occupation variables suggest a lower propensity associated with the level of under-reporting among workers in clerical and manufacturing vocations compared to those in professional, sales, and other vocations. Also, the effects of the variables corresponding to land use at the work site indicate an overall higher propensity of under-reporting and level of under-reporting in the group of employed respondents who work in residential settings. In addition, employed individuals working in industrial/medical land-use sites are likely to under-report more trips. These results of the impacts of land-use at the work site are not immediately intuitive, and need further exploration in the future.

Finally, among the group of driver demographics, Table 3 suggests that respondents in households with children are less likely to under-report. Consistent with this result is the finding that individuals in nuclear family arrangements (two adults of opposite sex with one or more children) have a lower propensity associated with the level of under-reporting than individuals in other types of households.

5.2.2 Driver Travel Characteristics

Driver travel characteristics are very statistically significant in explaining both whether a driver under-reports and the level of under-reporting. As would be expected based on the general respondent burden theory, individuals making many trips are the ones likely to under-report and under-report substantially (the measure of total trips was obtained by adding all trips of an individual in the GPS data and the household survey data, and subtracting from this the number of common trips identified in both data sources).

The results also indicate that respondents who make long distance trips are disproportionately represented in the pool of under-reporters, and are among those who have a propensity to under-report several trips. This result needs additional exploration in future research, but perhaps is an indicator of time constraints among individuals who travel long distances. Finally, in the set of driver travel characteristics, the results also reveal that respondents who chain trips during the day have a higher propensity to under-report (though there is no effect on the level of under-reporting). This result is rather intuitive, since it is likely that individuals may forget to, or choose not to, report short and/or discretionary trips that are typically associated with trip chaining.

5.2.3 Driver Adherence to Survey Protocol

The final group of variables shows the lower likelihood of under-reporting and level of under-reporting among those who use a travel diary. This is reasonable, since individuals who do not use their diary are the ones who may find it difficult to recall their trips during the telephone retrieval. Also, as expected, travel information retrieved through a proxy person is likely to be associated with a higher level of under-reporting.

5.3 Constant Parameter, Threshold Parameters and Correlation

The constant term in the binary choice model is negative, indicating that a majority of individuals do not under-report at all. The threshold parameters determine the correspondence between the latent propensity associated with the level of under-reporting and the observed number of missed trips. We estimated four threshold parameters, since we did not include a separate constant term in the latent propensity associated with the level of under-reporting. As such, the threshold parameters do not have any substantive behavioral interpretation.

The sign of the correlation term in the joint under-reporting/level of under-reporting model (see seventh row from bottom in Table 3) indicates a positive correlation in unobserved factors affecting whether or not an individual under-reports and the magnitude of under-reporting. As the reader will note, the correlation parameter is fixed at 0.999 in the estimation. This is because the correlation parameter rapidly converged toward 1 in the iterations, and this 'broke down" the maximum likelihood optimization procedure as the gradient function involves terms of the form of $\sqrt{1-\rho^2}$ in the denominator (ρ is the correlation parameter). However, we verified that the likelihood function indeed improves as the correlation parameter approaches 1 by fixing the correlation parameter at values between 0.0 and 0.999, and examining the profile of the optimized likelihood function. This profile indicated a monotonically decreasing likelihood function maximum from –262.4 at zero correlation to –262.2 at a correlation of 0.5 to –261.8 at a correlation of 0.75, and finally to the low of –258.9 at a correlation of 0.999. The almost unit correlation indicates that the same unobserved factors (no interest in the survey, lack of time, or other unobserved personality/travel characteristics) that increase the propensity to under-report also increase the propensity associated with the magnitude of under-reporting. Also, while we are unable to obtain a t-statistic on the correlation parameter (because it is fixed), we can compute a likelihood ratio test statistic to evaluate its statistical significance. In particular, the likelihood function value of the model with no correlation is –262.4, while the likelihood function value of the model in Table 3 is –258.9. The likelihood ratio test for testing the presence of significant correlation is 7.0, which is larger than the chi-square table value with one degree of freedom even at the 0.01 level of significance.

5.4 Measures of Fit

Several measures of fit are computed for the joint binary choice-ordered response model of under-reporting and level of under-reporting, as discussed in the footnotes of Table 3. The likelihood ratio test value for the null hypothesis that none of the independent variables and the correlation parameter help in explaining under-reporting and its level may be obtained from the log-likelihood value at convergence for the model $[L(\hat{\beta})]$ and the log-likelihood value at sample shares $[L(c)]$. This value is –2[–371.56 + 258.90] = 225.3, which is larger than the critical chi-squared value with 24 degrees of freedom even at the 0.00001 level of significance. This finding clearly highlights the value of the joint model estimated here.

The joint model can also be compared to the model that uses a single ordered response structure to estimate both whether an individual under-reports and the level of under-reporting. In this single ordered response structure, the dependent variable is the number of missed trips and takes the values of 0, 1, 2, 3, 4, and 5+. The restriction in this single ordered response model is that the underlying mechanism representing whether an individual under-reports or not also captures the propensity of the level of under-reporting. That is, the propensity to under-report and the propensity associated with the level of under-reporting become one and the same, and are tied tightly together, unlike the joint model estimated in the current paper. The joint model and the single ordered response model can be compared using a non-nested likelihood ratio test because both these models have the same value of log-likelihood at sample shares $[L(c)]$. For the comparison, we estimated a single ordered response model with all variables that appear in either the binary choice component or the ordered response component of the joint model. Thus, the single ordered response model has a total of 19 variables and five thresholds (since there are six under-reporting categories of 0, 1, 2, 3, 4, 5+ and we do not include a constant in the ordinal propensity). The log-likelihood at convergence of this single ordered response model was -279.17, yielding a $\bar{\rho}_c^2$ fit value of 0.197 (compared to 0.239 for the joint model). The probability that the differences in the $\overline{\rho}_c^2$ values between the two models (0.239 – 0.197 = 0.042) could have occurred by chance is less than $\Phi\left\{-[-2\times 0.042\times L(c)+(24-19)]^{0.5}\right\}$ [see Ben Akiva and Lerman, p. 172 (*24*)]. This value is 0.002, indicating that the difference in adjusted rho-bar squared values between the two models is highly statistically significant and that the joint model of the current paper is to be preferred.

5.5 Elasticity Effects of Explanatory Variables

The parameters on the exogenous variables in Table 3 do not directly provide the magnitude of the effects of variables on trip under-reporting. To address this issue, and also to estimate the effects of variables on the overall expected value of under-reported trips across all individuals, we compute the aggregate-level elasticity effects of variables. To do so, we first write the expected value of the number of under-reported trips for individual i using the notation in Section 3 as:

$$E(N_i)=\sum_{j=1}^{J} j\times\left[\Phi_2\left(a_j-\alpha' Z_i,\gamma' X_i,-\rho\right)-\Phi_2\left(a_{j-1}-\alpha' Z_i,\gamma' X_i,-\rho\right)\right]. \tag{5}$$

Next, we change the value of the exogenous variables in X_i and Z_i to evaluate changes in $\sum_i E(N_i)$ across all individuals. The reader will note that there are some common variables in Z_i and X_i (number of trips made, average trip distance, *etc.*), and an increase in these common variables will affect $E(N_i)$ through both the under-reporting binary model and the level of under-reporting ordinal model.

An important issue to consider when computing the "elasticity effects" of the exogenous variables is that the variables in Table 3 include a continuous variable (average trip distance), an ordinal variable (number of total trips made by individual), and dummy variables (all the remaining variables). For the continuous variable, we compute an arc elasticity by increasing the average trip distance by a uniform 10% across all individuals in the sample, estimating the new expected value of number of under-reported trips for each individual i using Equation (5), computing the new total value of under-reported trips across all individuals, and obtaining the proportional change from the baseline total value of under-reported trips. For the ordinal variable (number of total trips made by the individual), we increase the value of the variable by 1 unit for each individual and obtain the change in the expected total number of under-reported trips across all individuals. For the dummy variables, we change the value of the variable to one for the subsample of observations for which the variable takes a value of zero and to zero for the subsample of observations for which the variable takes a value of one. We then sum the shifts in expected total number of under-reported trips in the two subsamples after reversing the sign of the shifts in the second subsample and compute an effective proportional change in expected total number of under-reported trips in the entire sample due to a change in the dummy variable from 0 to 1.

The elasticity effects are presented in Table 4 by variable category. As can be observed from the table, the most important determinants of trip under-reporting are associated with individuals who work in residential land-use locations and who do not use a diary. This is because the "residential" and "use travel diary" variables affect both trip under-reporting propensity and the magnitude of trip under-reporting propensity in Table 3 in the same direction, and have relative large parameters compared to other variables in both the propensity equations. Among the remaining variables, the effects of being a male and traveling long distances per trip are comparatively less than the other variables.

6. CONCLUSIONS

This paper has examined the driver demographics, driver travel characteristics, and driver adherence to survey protocol considerations that impact the likelihood of an individual under-reporting trips in a household travel survey. The research models both the likelihood of trip under-reporting as well as the level of trip under-reporting (*i.e.*, the number of trips under-reported). This 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 variables may affect these outcomes differently. At the same time, the research recognizes that these outcomes may be related to one another. The model framework takes the form of a joint binary choice-ordered response structure and is estimated using classical maximum likelihood estimation techniques.

The empirical analysis in the paper uses data extracted from the Kansas City Household Travel Survey that was conducted in Spring 2004. Specifically, drivers who both reported their travel patterns during the CATI interview and whose vehicles were equipped with GPS technology were selected for the analysis.

The results 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 choice-ordered 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 diary to record their travel are more likely to miss trips than those who use the diary, and proxy reporting leads to more missed trips.

The model results in the paper can be used to determine the expected value of under-reported trips for each individual in a household travel survey (see Section 5.5). This estimate can then be used to create adjustment factors to control for under-reporting. However, as importantly, the 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.

6.1 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 Kansas City study travel logs allowed space for 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 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 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). As shown in Table 1, 40% of these respondents had missed trips, of which 67% were under age 30 and 33% between the ages of 32 and 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.

6.2 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, have 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, 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.

ACKNOWLEDGEMENTS
The authors acknowledge the helpful comments of four anonymous reviewers on an earlier version of the paper. A special thank you to Todd Ashby and Charles Gorugantula of the Mid-America Regional Council for permission to utilize the data for this analysis, and to Lisa Macias for her assistance in typesetting and formatting this document. Finally, the second author would like to dedicate his part of the research efforts to his Father, Dr. Ramalinga Bhat, who passed away in May 2005. The contents of this report reflect the opinion of the authors only.

REFERENCES

1. Wolf, J., M. Oliviera, and M. Thompson. The Impact of Trip Underreporting on VMT and Travel Time Estimates: Preliminary Findings from the California Statewide Household Travel Survey GPS Study. In *Transportation Research Board 82nd Annual Meeting*. Preprint CD-ROM. Transportation Research Board, National Research Council, Washington, D.C., 2003.

2. Zmud, J., and J. Wolf. Identifying the Correlates of Trip Misreporting - Results from the California Statewide Household Travel Survey GPS Study. Presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.

3. Battelle Memorial Institute. *Global Positioning Systems for Personal Travel Surveys. Lexington Area Travel Data Collection Test*. Final Report to Office of Highway Information Management, Federal Highway Administration, Washington, D.C., 1997. http://www.fhwa.dot.gov/ohim/lextrav.pdf

4. Casas, J., and C. H. Arce. Trip Reporting in Household Travel Diaries: A Comparison to GPS-Collected Data. In *Transportation Research Board 78th Annual Meeting*. Preprint CD-ROM. Transportation Research Board, National Research Council, Washington, D.C., 1999.

5. NuStats. *Year 2000 Post-Census Regional Travel Study, GPS Study Final Report*. Southern California Association of Governments, Los Angeles, 2004a.

6. NuStats. *Pittsburgh Travel View Household Travel Survey*. Southwestern Pennsylvania Corporation, Pittsburgh, 2002.

7. NuStats. *Household Travel Survey Final Report of Survey Methodology*. East-West Coordinating Council, St. Louis, 2003b.

8. NuStats. *Household Travel Survey Final Report of Survey Results*. East-West Coordinating Council, St. Louis, 2003c.

9. Pierce, B., J. Casas, and G. Giaimo. Estimating Trip Rate Under-Reporting: Preliminary Results from the Ohio Household Travel Survey. In *Transportation Research Board 82nd Annual Meeting*. Preprint CD-ROM. Transportation Research Board, National Research Council, Washington, D.C., 2003.

10. Pearson, D. Comparison of Trip Determination Methods in GPS-Enhanced Household Travel Surveys. In *Transportation Research Board 84th Annual Meeting*. Preprint CD-ROM. Transportation Research Board, National Research Council, Washington, D.C., 2005a.

11. Pearson, D. Household Survey Trip Underreporting: Case Study in Texas. In *Transportation Research Board 84th Annual Meeting*. Preprint CD-ROM. Transportation Research Board, National Research Council, Washington, D.C., 2005b.

12. Texas Department of Transportation. *Tyler/Longview Household Travel Survey Documentation*. Texas Department of Transportation, Austin, 2003.

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

14. Wolf, J. Applications of New Technologies in Travel Surveys. Presented at Seventh International Conference on Travel Survey Methods, Playa Herradura, Costa Rica, 2004.

15. Wolf, J., M. Loechl, J. Meyers, and C. Arce. Trip Rate Analysis in GPS-Enhanced Personal Travel Surveys. Presented at International Conference on Transport Survey Quality and Innovation, Kruger Park, South Africa, 2001.

16. McGuckin, N., and E. Murakami. Examining Trip Chaining: A Comparison of Travel by Men and Women. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1693*, TRB, National Research Council, Washington, D.C., 1999, pp. 79-85.

17. Levinson, D. How Patterns Changed from ’68 to ’88. *ITS Review*, Vol. 20, No. 2, February 1997, pp. 2-3.

18. Taylor, B. Beyond the Gender Gap. *ITS Review*, Vol. 20, No. 2, February 1997, p. 4.

19. Badoe, D. A., and G. N. Steuart. Underreporting of Trips in Travel Survey Conducted by Telephone. In *Transportation Research Board 78th Annual Meeting*. Preprint CD-ROM. Transportation Research Board, National Research Council, Washington, D.C., 1999.

20. Kostyniuk, L. P., L. Wargelin, C. Purvis, and K. Vaughn. Improving the Accuracy of Trip-Chaining Information in Activity/Travel Surveys. Presented at International Conference on Transport Survey Quality and Innovation, Kruger Park, South Africa, 2001.

21. Wargelin, L., and L. P. Kostyniuk. Proxy Respondents in Household Travel Surveys. Presented at Seventh International Conference on Travel Survey Methods, Playa Herradura, Costa Rica, 2004.

22. Popuri, Y. D., and C. R. Bhat. On Modeling Choice and Frequency of Home-Based Telecommuting. In *Transportation Research Record: Journal of the Transportation Research Board, No.* 1858, TRB, National Research Council, Washington, D.C., 2003, pp. 55-60.

23. Misra, R., and C. R. Bhat. Activity Travel Patterns of Non-Workers in the San Francisco Bay Area: Exploratory Analysis. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1718*, TRB, National Research Council, Washington, D.C., 2000, pp. 43-51.

24. Ben-Akiva, M., and S. R. Lerman. *Discrete Choice Analysis*. MIT Press, Cambridge, 1985.

LIST OF TABLES

TABLE 1 Summary of GPS Surveys

Study	Year Conducted	GPS Firm †	# HH	# HH w/ GPS & CATI	% of total CATI surveyed HHs participating in GPS survey	Level of Trip Under-reporting	Reference
Lexington	1996	B	100	84	84.0%	NA	Battelle Memorial Institute (*3*)
Austin	1997	N	2,000	200	10.0%	31%/12%*	Casas & Arce (*4*)
California	2001	G	16,990	292	1.7%	23%	Zmud & Wolf (*2*)
Los Angeles	2001/2	B	23,302	293	1.3%	35%**	NuStats 2004a (*5*)
Pittsburgh	2001/2	G	2,554	46	1.8%	31%	NuStats 2002 (*6*)
St. Louis	2002	G	5,094	150	2.9%	11%	NuStats 2003b,c (*7,8*)
Ohio	2002	B	6,338	230	3.6%	30%**	Pierce *et al.* (*9*)
Laredo	2002	G	1,971	87	4.4%	81%***	Pearson 2005a,b (*10,11*)
Tyler/Longview	2003	G	2,336	249	10.7%	NA	Texas DOT (*12*)
Kansas City	2004	G	3,049	228	7.5%	10%	NuStats 2004b (*13*)

Source: NuStats Project Archives (*14*).

† B - Battelle, N - NuStats, G - GeoStats

*Two different dwell time thresholds were employed in the GPS trip detection algorithms for Austin: 45 seconds and the more widely accepted 120 seconds. At the 45-second dwell time threshold, the trip under-reporting rate was 31%. At the 120-second dwell time, it was a 12% rate (*15*).

**These rates reflect both driver and passenger trips that were under-reported. The other studies focus only on missed driver (vehicle) trips.

***Based on documentation in Pearson (*10,11*), this rate was determined by considering all linked travel captured through the GPS units vs. that reported by CATI. However, it does not screen the GPS data relating to trips outside the study area, commercial trips that respondents were told not to report, nor vehicle usage by non-household members.

TABLE 2 Descriptive Analysis of Trip Under-reporting by Driver Characteristics

Variables	% of Drivers with No Missed trips*	% of Drivers with Missed Trips*
Driver Demographic Characteristics		
Age-related variables		
Age < 30 years**	71.1	28.9
Age 30-39 years**	77.1	22.9
Age 40-49 years**	72.0	28.0
Age 50-59 years**	65.9	34.1
Age 60-69 years	65.9	34.1
Age ≥ 70 years**	78.6	21.4
Sex of individual		
Male **	66.7	33.3
Female **	76.1	23.9
Education level (Highest level obtained)		
Less than high school	59.1	40.9
High school graduate **	68.9	31.1
Some college credit, but no degree**	70.3	29.7
Associate or technical school degree**	69.8	30.2
Bachelor's or undergraduate degree**	74.8	25.2
Graduate degree**	74.0	26.0
Employment status		
Full-time employed**	73.5	26.5
Part-time employed**	69.7	30.3
Not employed**	66.2	33.8
Occupation (for employed individuals)		
Sales or service**	66.2	33.8
Clerical or administrative	68.2	31.8
Manufacturing, construction, maintenance, or farming**	75.0	25.0
Professional, managerial, or technical**	73.4	26.6
Other	80.0	20.0
Land-use at work site (for employed individuals)		
Office building**	68.3	31.7
Retail**	77.8	22.2
Industrial/manufacturing**	84.4	15.6
Medical**	75.0	25.0
Educational**	81.8	18.2
Residential	58.5	41.5
Other**	75.0	25.0

* The percentages sum to 100% for each row
** Indicates statistical significance at the 90% confidence interval

TABLE 2 Continued

Variables	% of Drivers with No Missed trips*	% of Drivers with Missed Trips*
Driver Demographic Characteristics (continued)		
Household structure		
Single adult household (not including retirees)	61.8	38.2
Couple family household**	74.1	25.9
Nuclear family household**	71.0	29.0
Single parent household**	80.0	20.0
Joint family household (>2 adults)	68.8	31.2
Retired household	50.0	50.0
Presence of children		
No children**	70.7	29.3
1 or more children**	72.2	27.8
Household size		
1	60.5	39.5
2**	74.4	25.6
3**	68.4	31.6
4+**	74.5	25.5
Household vehicles		
1**	68.8	31.3
2**	74.7	25.3
3	63.4	36.6
4+**	80.0	20.0
Driver Travel Characteristics		
Total number of vehicle driver trips during the survey day		
<5**	91.9	8.1
5-9**	69.1	30.9
10-14	30.4	69.6
≥ 15	12.5	87.5
Average vehicle driver trip distance across all trips		
<5 miles**	74.6	25.4
5-9 miles**	70.2	29.8
10-14 miles	65.2	34.8
15-19 miles	68.8	31.3
≥ 20 miles	60.0	40.0
Trip chaining		
Individual never trip chains**	95.8	4.2
Individual trip chains in some tours**	64.5	35.5
Individual always trip chains in each tour**	69.4	30.6
Driver Adherence to Survey Protocol		
Diary usage for recording travel		
Used diary**	72.0	28.0
Did not use diary	58.0	42.0
Proxy reporting status		
Reported by individual**	74.0	26.0
Reported by proxy**	70.0	30.0

* The percentages sum to 100% for each row
** Indicates statistical significance at the 90% confidence interval

TABLE 3 Estimated Model Results of Trip Under-Reporting and Level of Trip Under-Reporting

Explanatory Variables	**Trip Under-Reporting**		**Magnitude of Trip Under-Reporting**	
	Parameters	**t-stat**	**Parameters**	**t-stat**
Driver Demographic Characteristics				
Age-related variables				
Age 30-59 years			-1.1258	-3.855
Age 60-69 years			-1.6965	-3.061
Age $\geq$ 70 years			-2.5751	-3.789
Age 50-69 years	0.2958	1.561		
Male	0.2114	1.381		
Education level less than high school	1.0410	3.119		
Employment status				
Full-time			-0.6436	-2.296
Part-time			-0.9734	-3.233
Occupation				
Clerical			-1.0595	-2.159
Manufacturing			-0.8573	-1.724
Land-use at work site				
Residential	0.4653	1.745	1.4705	4.248
Industrial/medical			1.0579	2.339
Household structure				
Presence of children	-0.2612	-1.255		
Nuclear family			-0.8838	-3.725
Driver Travel Characteristics				
Total number of trips during the survey day	0.2281	6.561	0.3407	7.594
Average trip distance across all trips	0.0904	5.403	0.0988	3.943
Trip chaining	0.9204	2.696		
Driver Adherence to Survey Protocol				
Used travel diary	-0.9641	-1.870	-1.2573	-2.947
Travel reported via proxy			0.7240	3.239
Constant Term in Trip Under-Reporting	-1.8915	-3.272		
Threshold Parameters in Ordered-Response				
Threshold 1			1.5163	2.484
Threshold 2			2.1108	3.330
Threshold 3			2.5210	3.759
Threshold 4			2.8581	4.239
Correlation Parameter	0.999 (fixed)			
Number of Observations	377			
Log-likelihood at equal shares $L(o)$*	-435.14			
Log-likelihood at sample shares $L(c)$†	-371.56			
Log-likelihood at convergence $L(\hat{\beta})$	-258.90			
ρ_0^2 ‡	0.338			
$\bar{\rho}_c^2$ §	0.239			

* The log-likelihood value at equal shares corresponds to the case where (a) the probability of each individual under-reporting is 0.5, (b) the probability of each of the five trip under-reporting magnitude categories (1, 2, 3, 4, and 5+ missed trips) for each individual who under-reports in the sample is 0.5*1/5 = 0.1, and (c) the correlation parameter is 0.

† The log-likelihood value at sample shares corresponds to the case where (a) the probability that each individual under-reports is equal to the share of individuals under-reporting in the sample (=0.29), (b) the probability of each of the five trip under-reporting magnitude categories for each individual who under-reports is equal to the corresponding observed category sample shares, and (c) the correlation parameter is zero. This is equivalent to the model with only thresholds in the model.

‡ $\bar{\rho}_0^2 = 1 - \dfrac{[L(\hat{\beta}) - K]}{L(0)}$, where K is the total number of parameters in the joint model (29 in the current model).

§ $\bar{\rho}_c^2 = 1 - \dfrac{[L(\hat{\beta}) - (K - T)]}{L(c)}$, where T is the total number of thresholds (5 in the current model).

TABLE 4 Elasticity Effects of Explanatory Variables on Under-Reporting Trips

Explanatory Variable	Elasticity Effect
Driver Demographic Characteristics	
Age-related variables	
Age 30-49 years	-0.3694
Age 50-59 years	-0.2556
Age 60-69 years	-0.3413
Age ≥ 70 years	-0.5163
Male	0.0906
Education level less than high school	0.4471
Employment status	
Full-time	-0.2490
Part-time	-0.3369
Occupation	
Clerical	-0.3559
Manufacturing	-0.3089
Land-use at work site	
Residential	1.2000
Industrial/medical	0.5617
Household structure	
Presence of children	-0.1113
Nuclear family	-0.3155
Driver Travel Characteristics	
Total number of trips	0.2782
Average trip distance	0.0863
Trip chaining present	0.3922
Driver Adherence to Survey Protocol	
Used travel diary	-1.3588
Travel reported via proxy	0.3767