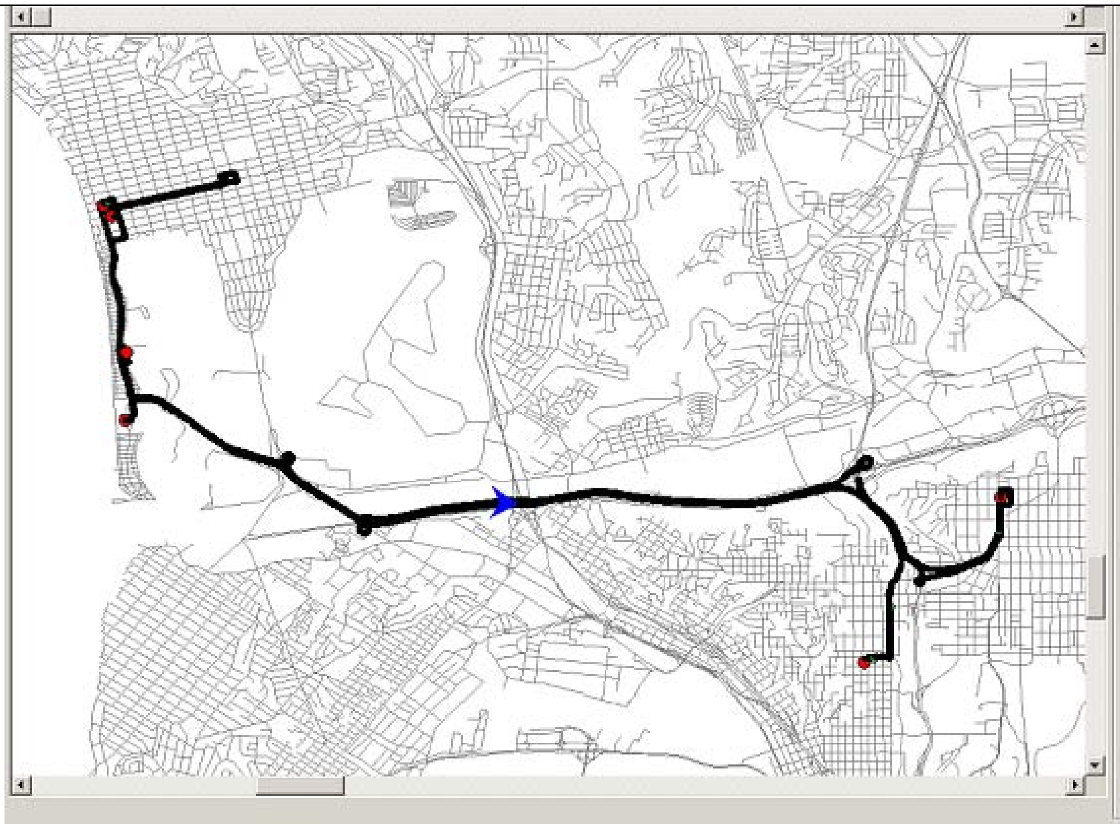


Identifying the Correlates of Trip Misreporting - Results from the California Statewide Household Travel Survey GPS Study

Johanna Zmud, NuStats
Jean Wolf, GeoStats

Conference paper
Session XXX

Moving through nets:
The physical and social dimensions of travel
10th International Conference on Travel Behaviour Research

Lucerne, 10-15. August 2003

Identifying the Correlates of Trip Misreporting - Results from the California Statewide Household Travel Survey GPS Study

Johanna Zmud
NuStats
3006 Bee Caves Road, Suite A300
Austin, Texas 78746
Phone: 512.306.9065
Fax: 512.306.9077
eMail:jzmud@nustats.com

Abstract

Automated trip recording technologies, such as Global Positioning System (GPS) receivers and data loggers, can provide great insight into the problem of travel survey misreporting. Recent studies have confirmed the feasibility of applying GPS technology to improve both the accuracy and completeness of travel data by using it as an audit tool for traditional reporting methods. This paper uses GPS and diary data collected in the California Statewide Household Travel Survey to draw conclusions about the nature of trips that might have been missed and the characteristics of persons or households that are positively associated with trip underreporting. The paper presents a measurement approach for conducting a comparison of diary and GPS data that specifies the units of analysis and the appropriate variables. Then, a statistical examination of underreporting trip behaviors will be conducted based on household, person, and trip characteristics. Once these correlates of underreporting are identified, a methodology for generating a set of weights (or trip rate correction factors) for the statewide household travel survey is presented.

Keywords

Trip Underreporting, Misreporting, GPS Travel Surveys, International Conference on Travel Behaviour Research, IATBR

Preferred citation

Zmud, Johanna and Jean Wolf (2003). Identifying the Correlates of Trip Misreporting - Results from the California Statewide Household Travel Survey GPS Study, paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.

1. Introduction

Understanding traveler behavior is critical to urban transportation planning and modeling. For this reason, household travel surveys are conducted periodically. Dependent variables of great interest in a household travel survey relate to the respondent's reported trip making behavior, including the number of trips made by each respondent and the associated information about each trip. An important determinant of data quality is the accuracy of the reported trips. To enhance reporting accuracy, most household travel surveys rely on diary instruments in which respondents are asked to record each trip for a specific time period (e.g., 24-hours, 48-hours). These travel diary details then reported by the household members during a computer-assisted telephone interview (CATI). However, even with the use of diary and CATI methodologies, misreporting of trip information occurs. Evidence that misreporting is a problem in household travel surveys can be found in previous research (Wilmot and Adler, 2003; Adler, 2003, Wolf, *et al.*, 2003a, Zmud and Arce, 2000).

Respondent misreporting of travel behavior may be an honest mistake or may be intentional due to privacy concerns, nonchalance, or other factors. For purposes of this paper, the cause of the misreporting matters little. What does matter is how to identify and deal with misreporting. One option would be to use only trip data that has been validated by some ancillary source. This an appealing option if one can assume that the validation data are correct; then, both estimation and inference are fairly straightforward.

Collecting the validation data was once prohibitively difficult and costly. However, new advancements in Global Positioning System (GPS) technology have made GPS data a reliable source with which trip data from a particular sample could be compared. In the late 1990's, several pilot studies were conducted to investigate the use of GPS technology for trip data collection. These pilot studies confirmed the feasibility of applying GPS technology to improve both the accuracy and the completeness of collected trip data. In 1997, the first real deployment of GPS equipment in a household travel survey occurred in Austin; however, challenges with GPS data accuracy related to the U.S. government's intentional GPS signal degradation (known as Selective Availability) at the time made it difficult to assess the benefit of collecting GPS data concurrently with travel diary data. On May 1, 2000, the U.S. government announced the immediate termination of Selective Availability – which improved, literally overnight, the positional accuracy of raw GPS data from a 30-100 meter range down to 5-10 meters. This dramatic improvement in GPS positional accuracy made the use of GPS technology in household travel surveys more desirable, while the continuously declining costs associated with GPS equipment made the application of this technology more feasible.

Although GPS methods are now practical for some travel survey validations, the cost for collecting trip data using GPS from a typical household is at least triple of that for using CATI methods. Thus, GPS data have been used to validate only a small subsample of the larger survey effort. Recently, U.S. attempts to validate reported trip data were conducted with subsamples of household travel surveys in the States of California and Ohio, Southern California, Pittsburgh, St. Louis, and Tyler / Longview and Laredo, Texas. These studies have attempted to quantify the percent of misreported trips within the subsample and to cautiously evaluate the impact of these percentages to the overall household survey trip rates.

This paper takes trip-misreporting analysis to the next level by identifying the correlates of underreporting in GPS-enhanced household travel surveys. Data from the 2001 California Statewide Household Travel Survey GPS Study were used to test a proposed methodology for quantifying the amount of trip under-reporting that occurred, for identifying the factors which might contribute to misreporting, and for generating improved trip rate estimates for the statewide household travel survey sample.

2. Equipment Description

For this GPS-enhanced travel study, the GeoStats GeoLogger™ was the GPS data logger used. The GeoLogger is a rugged and simple GPS data-logging device (see Figure 1) that has been deployed in household travel surveys and travel time studies within the US, Canada, Europe, and Australia. The GeoLogger is easy to install – the respondent only needs to plug the power connector into the cigarette lighter socket within the vehicle and to place the combination GPS receiver/antenna on the roof of the vehicle via a magnetic mount.

Figure 1. The GeoStats GeoLogger™

This device can log at either one-second or five-second frequencies, all valid GPS points or only those valid points for which the speed is greater than 1 MPH (to screen out non-movement events). It is available in 1 MB, 2 MB, and 4 MB versions. For the purpose of this study, the 1 MB units were used, the logging logic was set at one second logging frequencies, and points were not to be logged when speed measured was less than or equal to 1 MPH.

The standard GPS data stream elements recorded by the GeoLogger include date, time, latitude, longitude, speed, heading, altitude, number of satellites, and horizontal dilution of precision (HDOP, a measure of positional accuracy). These elements are stored in the logger in standard NMEA units and are converted into user-specified units and formats upon download. Output from the GPS receiver can be expected to perform to the following specifications:

- Accuracy Position: Within 5 to 10 meters typical, within 15 meters RMS
- Velocity: Within 0.1 meters per second RMS steady state (or 0.23 MPH or 0.36 KPH)
- Time to First Fix, Cold Start: 45 seconds
- Time to First Fix, Warm Start: 15 seconds
- Reacquisition: 2 seconds

3. GPS Validation Study Methodology

The equipment deployment goal for the GPS validation study was 500 households that were to be recruited from within the 16,990 households participating in the statewide household travel survey. It was determined that 500 deployments would most likely produce between 300 and 400 complete households. Given the small sample size (as compared to the statewide sample) and to allow for efficient deployment, a sampling plan was developed for three geographic regions within the state – San Diego, Sacramento, and Alameda counties.

The validation component followed the design and data methodology of the main survey, which included five basic survey procedures: 1) sending a pre-contact notification letter to provide information about the study to potential participants; 2) conducting a household-level recruitment interview to collect information on household demographics and solicit participation in the study; 3) distributing person-based travel diary booklets to each household that agreed to participate; 4) conducting reminder telephone call attempt the night before the assigned travel day; and 5) collection travel information for trips that occurred on the assigned travel day from household members via CATI. NuStats (Austin, Texas) was responsible for identifying, recruiting, and screening households; providing successfully recruited households with travel diaries and other necessary information; and interviewing the households to collection information recorded on participants' diaries. GeoStats (Atlanta, Georgia) was responsible for re-contacting households who agreed to receive GPS equipment to re-recruit them, processing the devices for deployment, managing the deployment firm that was contracted to distribute and retrieve the devices, and processing the data collected by the devices.

The validation study was conducted over a 20-week period, beginning on February 12, 2001, stopping during the summer months, and finishing up on October 3, 2001. A total of 112 GeoLoggers were used during this study – none of which were lost or damaged. A breakdown of the equipment deployment results can be seen in Table 1.

Table 1. GPS Field Results

Result	Alameda	San Diego	Sacramento	Total
Number of HH Deployed	148	195	174	517
Number of Completed HH (completed GPS and CATI)	88	111	93	292
Number of Completed Units	152	200	171	523
Completion Rate by HH (Completed / Deployed)	59%	57%	53%	56%

Of the 517 households that were originally recruited for the GPS portion of the study, 225 did not complete some or all parts of the study, or the data collected by them included data anomalies that could not be resolved. These households were categorized as partials (43 households) or refusals (182 households). A partial household is classified as a household that has one or more vehicles with useable CATI and GPS data, but other vehicles in the household that either refused the GPS unit, forgot to use it, had a broken cigarette lighter, experienced GPS equipment malfunction or traveled with their GPS unit on a day different then their scheduled travel day. A refusal household is one for which no comparative analysis can be made due to either a lack of useable GPS data or useable CATI data for all vehicles in the household. The remaining 292 households were considered to be completes and were used for all subsequent analysis. The breakdown by county and level of participation for all 517 households in the GPS study can be seen in Table 2.

Table 2. GPS Outcomes for All Households Deployed

Location	Completes	Partials	Refusals	Totals	% Total
Alameda	88	10	50	148	28.6
Sacramento	93	21	60	174	33.7
San Diego	111	12	72	195	37.7
Totals	292	43	182	517	100.0
% Total	**56.5**	**8.3**	**35.2**	**100.0**	

Post-processing of the trip data consisted of data editing and cleaning and then matching the GPS-recorded vehicle driver trips to diary reported vehicle driver trips. *Vehicle driver trips* were the basis for the comparative analysis because this is the unit of measurement captured by the in-vehicle GPS loggers.

Both the data editing and vehicle driver trip matching were accomplished through a blend of computer and manual edits. The GPS second-by-second data, once received, were first converted into GIS-compatible formats and then reviewed for potentially bad or poor data points. Then, potential vehicle driver trip ends were identified based on time intervals between consecutively logged points. For this study, 120 seconds was defined as the appropriate initial dwell time between GPS-recorded trips. Next, each potential trip was manually evaluated within an interactive GIS-based application to identify both missing and false trip ends. Once this step was complete, the updated GPS-based vehicle driver trip file for a given household vehicle was ready for comparison. A total of 2,566 GPS trips were identified based on the initial minimum 2-minute stop or dwell time. Further processing of the GPS trip data within the interactive GIS application revealed another 45 stops, with a duration of less than two minutes that occurred off of the vehicle's main travel path.

The 2-minute initial dwell time for stop determination was established based on previous GPS studies that revealed this threshold to be a good starting value. Establishing a stop threshold at values less than two minutes often tend to pick up travel delays associated with traffic signals and congestion, while stop thresholds set at values greater than two minutes often miss other typical short-duration stops, including drop-offs and pick-ups. Post-processing of the CATI data entailed converting the files to a vehicle-based format to provide a standard unit against which GPS trips could be compared. The CATI vehicle trip file generated a total of 2,128 trips for the same households and vehicles in which GPS data were successfully collected. Once this conversion was complete, both the GPS vehicle-based trip files and the CATI-retrieved vehicle-based trip files were ready for comparison.

4. Defining and Measuring the Misreporting Problem

Travel behavior researchers know that misreporting occurs. There is little reliable data on the extent to which it takes place within a household travel survey. In this section of the paper, descriptive information is provided about the nature of misreports in household travel survey samples. The GPS data captured in the California Statewide Travel Survey were used as the validation data for misreports. This analysis was conducted under the assumption that the validation data are correct, or at least are not subject to systematic bias.

The analysis involved comparing GPS and CATI data on a per-vehicle level. The data set for the present analysis was comprised of those households where all vehicles had both CATI and GPS data. A program was written to perform the initial GPS and CATI comparison based on a 12.5-minute departure time buffer as the match criteria. After the initial comparison was made, the files were reviewed manually and adjusted if necessary. (For more explanation of this matching process, see Wolf, et.al, 2003a). Matching results and discrepancies fell into the following categories:

- Perfect match between GPS and CATI vehicle driver trips
- Imperfect, but "match-able" GPS and CATI vehicle driver trips (manually matched)
- Non-match-able vehicle driver trips -- detected via GPS but not reported via CATI (under-reported)
- Non-match-able vehicle driver trips -- not detected via GPS but reported via CATI (over-reported)

The information presented in Table 3 uses vehicles as the unit of analysis, as vehicle trips were the basis for comparison in this validation study. For more than half of the vehicles in the validation sample (56%), comparable trip information was recorded via GPS and CATI. For about one-third of the instrumented vehicles (31.9%), some CATI trips were missing when compared against trips found in the GPS stream. Five percent of the vehicles had more CATI trips reported than GPS trips measured, and about 7% had a combination of missing data – missing trips within the GPS and CATI datasets.

Table 3. Results of GPS/ CATI Vehicle Driver Trip Matching Process

Classification of GPS/CATI match	Number of Vehicles	% Total Completes
Perfect match	227	43.4
Imperfect but "match-able"	66	12.6
Missing vehicle driver trips in CATI reports	167	31.9
Missing vehicle driver trips in GPS*	26	5.0
Missing trips in both CATI reports and GPS	37	7.1
TOTAL	**523**	**100.0**

* Most of these are attributable to late installation or early removal of the GPS equipment by the respondent

A summary of the results of the vehicle driver trip comparison for the 292 complete households, broken down by county, can be seen in Table 4. This table shows the number of complete households (# hh) for each county, the number of vehicles instrumented in these households (# veh), the total number of GPS-identified trips after the review process for all instrumented vehicles (# GPS trips); the total number of CATI-retrieved vehicle driver trips associated with all household vehicles (# CATI trips); the number of missed trips detected (# Missed Trips), for which a baseline measure has been calculated as simply the difference between the total number of GPS-detected trips and the total number of CATI-reported trips, and the percentage of missed CATI trips. According to Table 4, there were 2,611 GPS trips detected and 2,128 trips reported via CATI across the 523 vehicles instrumented in this study. These "missing" 483 vehicle driver trips are equivalent to 22.7% of the trips captured via GPS.

Table 4. Missed Trips in CATI Reports

County	Total HH	Total Vehicles	GPS Captured Trips	CATI Self-Reported Trips	# Missed Trips* (Baseline)	% Missed Trips (Baseline)
Alameda	88	152	711	605	106	17.5%
Sacramento	93	171	854	635	219	34.5%
San Diego	111	200	1,046	888	158	17.8%
Totals	292	523	2,611	2,128	483	22.7%

* The numbers in this column represent the net result of under- and over-reporting of trips in CATI data.

The baseline calculation of missed vehicle driver trips presented in Table 4 does not take into account that the total GPS trips (2,611) do not reflect the whole "truth." There were CATI reported trips that were not captured via GPS due to equipment mis-use (typically late installation or early removal of equipment). When these 101 missing GPS trips are added to the GPS baseline of 2,611, the total vehicle driver trips increases to

2,712 and the percent of missed trips increases to 27.4%, as shown in Table 5. (For more explanation of these findings, see Wolf, et.al, 2003a).

Table 5. Adjusted Missed Trips in CATI Reports

County	Total HH	Total Vehicles	Self-Reported CATI Trips	GPS-Captured Trips	# Missed GPS Trips	Total "Missed" Trips*	Total % Missed Trips	Total Trips
Alameda	88	152	605	711	28	134	22.1	739
Sacramento	93	171	635	854	45	264	41.6	899
San Diego	111	200	888	1,046	28	186	20.9	1,074
Total	292	523	2,128	2,611	101	584	27.4	2,712

* The numbers in this column represent the net result of under- and over-reporting of trips in CATI data.

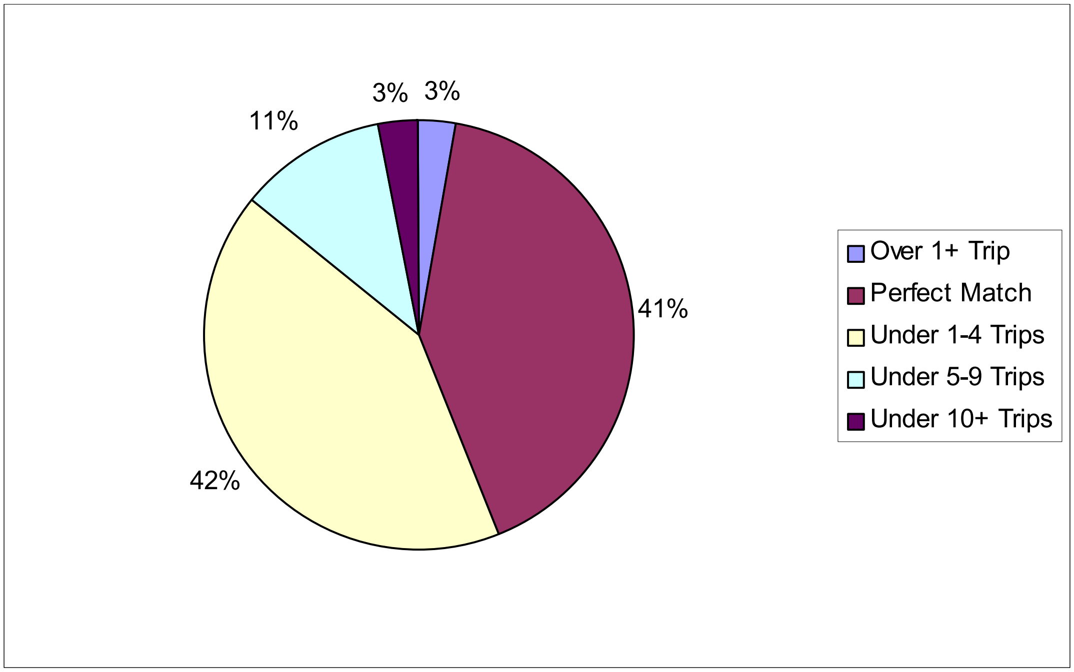
Based on the information in Table 5, a trip correction factor *could* be computed to weight the vehicle driver trips for households in the validation sample to adjust for the bias introduced from under-reporting. The overall adjustment weight would be 1.29 (2,712 GPS captured trips / 2,127 CATI reported trips). When this weight is applied to the 292 households in the validation sample, the trip rates increase as seen in Table 6.

Table 6. Household Trip Rates: Unadjusted and Adjusted for Under-Reporting

County	Number of Households	Total Vehicle Driver Trips	Total Trips	Unadjusted Trips per HH	Adjusted Trips per HH
Overall	292	2,127	2,340	8.0	10.7
Alameda	88	605	726	8.3	11.0
Sacramento	93	635	667	7.2	9.8
San Diego	111	887	1,014	9.1	11.2

However, the aggregate percent of under-reporting (27.4%) conceals the variation that actually exists in respondent reporting accuracy between counties and among individual households. At the net (or aggregate) household level, approximately 41% of the participating households reported the same total number of vehicle driver trips reported as were measured by GPS. About 3% of households over-reported vehicle driver trips; this percentage is calculated after the adjustment made for missed GPS trips due to equipment misuse. The remaining households, which account for approximately 56% of the validation same, failed to record or report at least one vehicle driver trip that was measured by the GPS devices.

Figure 2. Household-Level Variation in Vehicle Driver Trip Reporting among Validation Sample



5. Factors Associated with Under-Reporting

There was not uniform misreporting across all households in the validation sample. This is evident in the variation of under-reporting measured by county (as seen in Tables 4 and 5). When comparing the results for these three counties, the discrepancy in reporting accuracy is significant between Sacramento and the other two counties. About half of all households had at least one instance of a missing vehicle driver trip in the diaries collected from household members. The variation in under-reporting that occurs argues against the application of a constant trip correction factor to vehicle driver trips. There are obviously some types of households that do not require a correction factor, others that require a small adjustment factors, and others still that require a fairly large weight. This analysis attempted to identify the factors that significantly impact trip under-reporting so that the resulting information could be used to derive a set of weights (or correction factors) for more accurate adjustment of household trip rates.

So, what are the factors that contribute to trip under-reporting? Unless one can identify these correlates, there is no way to quantify the impact of under-reporting on overall survey data quality. Misreporting could be quite random with respect to variables of analytic interest, or it could systematically bias analyses of travel behavior and trip rate estimates. For example, one variable that has been suspected as a correlate for underreporting is trip length (or duration. Trips of short duration are assumed to be missing from respondent diaries more frequently than trips of long durations. In this study, 71% of missed trips were less than 10 minutes in duration (see Table 7). Although this estimate may be slightly high given that GPS trips may not include travel times associated with cold start signal acquisition delays, which could last one to two minutes, it still represents a significant portion of the total missed trips identified.

Table 7: Missed Trips by Trip Duration

Trip Duration	Total Vehicle Driver Trips	Total Missing Vehicle Driver Trips*	% Missing Vehicle Driver Trips
0-10 minutes	1,016	441	70.9%
11-20 minutes	616	119	19.2%
21-30 minutes	281	29	4.6%
30+ minutes	214	33	5.3%
Total	2,127	622	100.0%

* The numbers in this column represent the total number of under-reported trips in CATI data. It does not "net out" the over-reported trips.

The impact of underreporting in this sub-sample on modeled VMT and travel times was evaluated with assistance from the transportation modelers in each of the three regions (Wolf *et al.*, 2003b). The results of this analysis revealed that the characteristics of the models used by each region influenced the impact of the missing trips on overall VMT and travel time estimates. This research also found that inaccurately reported trip start times had a significant impact in shifting trips between peak and non-peak periods, causing major changes to the distribution of peak and nonpeak VMT and travel times. Finally, the VMT analysis did reveal that even though a large percentage of the underreported trips were short duration trips, there was a significant impact on overall VMT due to the quantity of trips missed across all durations.

For this analysis of the correlates of underreporting, numerous socio-demographic variables available in the California Statewide Survey database were selected based upon prior experience with household travel survey databases. Including Trip Duration, which is a trip characteristic rather than a socio-demographic variable, ten variables were analyzed for their contribution to underreporting : Trip Duration; Household Size; Vehicle Ownership; Household Income; Respondent Age; Employment Status; Student Status; and Presence of Children under 18. Table 8 summarizes the percent of "missed" trips according to these characteristics.

Table 8: Missed Trips by Household Size, Number of Vehicles, Household Income, Employment Status, Student Status, Presence of Children Under Age 18

Household Type	Number of Households	% of Households with Missed Trips	Total Missed Vehicle Driver Trips	% of Missed Vehicle Driver Trips
Overall	292	52.1%	622	22.6%
Household Size				
1 person	74	34.2%	69	18.6%
2 person	133	54.5%	291	24.0%
3 person	40	62.5%	113	24.7%

4 or more person	45	64.4%	149	21.0%
Number of Vehicles				
1 vehicle	91	37.8%	99	19.1%
2 vehicle	142	53.9%	302	21.6%
3 vehicle	38	68.4%	119	23.8%
4 or more vehicle	21	71.4%	102	30.7%
Household Income (of those reporting income)				
Less than $50,000	84	44.6%	197	30.1%
$50,000 to $99,999	128	58.3%	259	20.4%
$100,000 or more	54	55.6%	134	21.8%
Respondent Age				
Less than 25 years	9	44.4%	25	31.3%
25-34 years	38	50.0%	95	25.1%
35-44 years	68	55.2%	167	24.4%
45-54 years	76	61.3%	164	21.2%
55 years or older	101	44.6%	171	20.6%
Employment Status				
0 workers	63	39.7%	95	19.9%
1 worker	119	45.8%	162	17.8%
2 worker	99	65.3%	302	26.6%
3 or more workers	11	72.7%	63	27.5%
Student Status				
0 students	17	2.0%	18	<1.0%
1 student	78	25.2%	140	7.3%
2 students	120	42.4%	238	14.9%
3 or more students	77	30.5%	226	17.8%
Presence of Children Under 18				
Children present	80	62.0%	245	22.4%
No children present	212	48.3%	377	22.9%

The general findings in this section agree with common expectations. Respondents who may not travel a lot (i.e., not employed) may report more accurately because they have less to report. Smaller size households and also those with fewer vehicles are more accurate reporters. However, there are some interesting differences within these results. The presence of children in the household, which is significantly tied to higher mobility, does not appear to be associated with misreports. In general, subgroups in the sample that represent likely misreporters are: Households with 3 or more vehicles, households with annual incomes of less than $50K, respondents younger than 25 years of age, and households with 3 or more workers.

6. Re-Estimating Trip Rates Accounting for Misreporting

A significant goal of this research was to quantify the amount of trip under-reporting that may occur in a household travel survey by using GPS data as validation information, to identify the conditions under which misreporting will be a problem, and to use the validation study data to improve trip rate estimates for a household travel survey sample. This section introduces a proposed methodology for estimating a set of adjustment weights (i.e., correction factors) for household trip rates and for applying those weights to improve trip rate estimates.

The database of vehicle trip records was used to test a model of trip misreporting. In this model, y_i is an indicator (dummy) variable that is 1 if a vehicle trip record was "missing" when compared to the GPS data and 0 otherwise and x_i is a vector of associated characteristics that influence whether a trip will be "reported" via CATI or not. The objective of this analysis is to estimate the conditional distribution of y_i given x_i, $\Pr[y_i | x_i]$. A multivariate regression technique (Logistic Regression) was used to determine which variables, of the ten that have formed the core of this analysis (i.e., household size, number of vehicles, household income, employment status, student status, presence of children, trip duration), have the most impact on trip under-reporting.

Table 9: Results of Logistic Regression

Variable	S.E.	Sig.	Exp(B)
Trip Duration	.006	.000	.948
Vehicle Ownership	.065	.000	1.561
Household Income	.030	.000	.860
Age	.004	.000	.984
Household Size	.077	.002	.785
Student	.195	.037	1.501
Multiple Activities on Trip	.140	.129	.809
Presence of Children	.171	.430	1.145
Employed	.122	.864	.979

As indicated in Table 9, the logit analysis identified four variables as being significantly associated with trip under-reporting. The remaining six variables were found to be insignificant. The variables that were identified as contributing most the trip under reporting were Trip Duration, Vehicle Ownership, Household Income, and Age of Respondent. Missing data for Household Income or Age were assigned a value based on other members of the household and maintains the same overall distribution as the known cases. The other variables did not have any missing data. Based on these results, a 54-cell matrix representing the 4-way cross-tab of the four significant variables was created, which was then used to derive the adjustment weight for specific households types. Some cells were collapsed since the sample sizes were too small, resulting in 46 cells (see Table 10). The following are the code labels for each variable shown in Table 10. Within each of the resulting 46 cells, the total sample count (TOT) was divided by the number found by GPS to give an adjustment factor (WEIGHT).

Trip Duration (minutes)

1 0-6

2 7-14

3 >14

Vehicle Ownership

1 1-2

2 3+

Household Income

1 1-4 (<$50,000)

2 5-6 ($50,000 - $99,999)

3 7-8 ($100,000+)

Respondent Age

1 0-39

2 40-49

3 >49

Table 10: Adjustment Weights based on Model of Misreporting

TRIPDUR	VEHOWN	INCOME	AGE	TOT	GPS	WEIGHT
1	1	1	1	75	46	1.63
1	1	1	2	40	15	2.67
1	1	1	3	45	25	1.80
1	1	2	1	107	66	1.62
1	1	2	2	97	71	1.37
1	1	2	3	103	74	1.39
1	1	3	1	48	22	2.18
1	1	3	2	15	11	1.36
1	1	3	3	41	32	1.28
1	2	1	9	52	18	2.89
1	2	2	1	34	24	1.42
1	2	2	2	31	16	1.94
1	2	2	3	34	18	1.89
1	2	3	1	35	28	1.25
1	2	3	2	36	22	1.64
1	2	3	3	18	17	1.06

2	1	1	2	34	21	1.62
2	1	1	3	61	50	1.22
2	1	2	1	84	68	1.24
2	1	2	2	104	87	1.20
2	1	2	3	114	100	1.14
2	1	3	1	34	19	1.79
2	1	3	2	27	21	1.29
2	1	3	3	39	39	1.00
2	2	1	9	35	18	1.94
2	2	2	9	73	48	1.52
2	2	3	1	22	18	1.22
2	2	3	2	34	28	1.21
2	2	3	3	18	15	1.20
3	1	1	1	74	69	1.07
3	1	1	2	40	31	1.29
3	1	1	3	83	79	1.05
3	1	2	1	98	87	1.13
3	1	2	2	120	110	1.09
3	1	2	3	121	116	1.04
3	1	3	1	50	46	1.09
3	1	3	2	20	19	1.05
3	1	3	3	53	47	1.13
3	2	1	9	46	38	1.21
3	2	2	1	52	47	1.11
3	2	2	2	40	32	1.25
3	2	2	3	43	40	1.08
3	2	3	1	38	29	1.31
3	2	3	2	52	47	1.11
3	2	3	3	24	19	1.26
				2481	**1922**	**1.29**

The adjustment factors for the 46 cells range from a low of 1.0 to a high of 2.89. The cell with the lowest weight consists of trips with the following characteristics. It can be assumed that these trips were reported with most accuracy.

TRIPDUR = 7-14 minutes

VEHOWN = 1-2 vehicles

INCOME = 7-8 ($100,000+), and

AGE = 50-91.

The cell with the highest weight consists of trips with the following characteristics. These trips were reported with least accuracy.

TRIPDUR = 0-6 minutes

VEHOWN = 3-5 vehicles, and

INCOME = 2-4 (<$50,000).

Next, the weight was applied to all households in the statewide household travel survey dataset – not just households in the validation sample. A value was assigned for trips with missing data on any one of the four variables. A random value was assigned to these records in the same proportion as the valid data. Each weekday driver trip record was matched with the 46-cell matrix and the weight was applied. The updated weekday trip rates per household by region appear in Table 11.

Table 11: Weekday Total Trips per Household by Region

Region	Total Household	Mean
Western Slope/Sierra Nevada	681	9.0
AMBAG	867	10.4
MTC	1,643	10.9
SACOG	982	10.4
SCAG	3,384	9.9
Rural	2,437	11.1
Butte	546	11.6
Fresno	616	8.6
Kern	574	9.5
Merced	498	12.0
San Diego	1,187	9.7
San Joaquin	577	9.3
San Luis Obispo	648	11.3
Santa Barbara	817	11.6
Shasta	511	10.3
Stanislaus	536	8.8
Tulare	536	13.6
Statewide	**17,040**	**10.2**

The application of the weight was only applied to driver trip *rates* and not to *distributions* such as travel mode since the GPS units were only deployed in vehicles (i.e., all other modes were excluded). Also, the GPS sub-sample was not necessarily a representative sample of the overall sample. This means that the distribution across the 46 cells in the GPS sub-sample does not necessarily correspond to this distribution across all households in the full dataset. In other words, once the weight was applied, cell-by-cell, to the remaining households in the full dataset, the overall correction factor is virtually guaranteed to differ from the original level of 1.29. In this case, the overall correction factor among *all* households in the dataset is 1.33. A solution to avoid this outcome is to first weight the GPS sample so that its distribution patterns match exactly that of the entire sample. Then the weights are calculated using the above procedure but applied to the weighted data. It would also be desirable to have all missing value imputation completed before the weighting.

7. Conclusions and Next Steps

The contributions of this paper are twofold. First, the correlates of underreporting in a given household travel survey have been identified. This is crucial and has largely been ignored by the literature on travel survey data quality. These studies have been more concerned with the causes of misreports and solutions for preventing the problem, rather than how to adjust for its impact on important survey estimates. Second, this research introduces a model that corrects for this misreporting and a methodology for applying a set of adjustment weights to survey estimates. This approach, if accepted, could become a standard method for adjusting and correcting survey estimates on a wider scale. This methodology will next be applied to the St Louis Regional Household Travel Survey dataset collected in 2002.

8. References

T. Adler (2003). Reducing the Effects of Item Nonreponse in Travel Surveys. In *Transport Survey Quality and Innovation*, P. Stopher and P. Jones, eds., Elsevier: London, England

Willmot, C., and T. Adler (2003). Item Nonreponse. In *Transport Survey Quality and Innovation*, P. Stopher and P. Jones, eds., Elsevier: London, England, pp. 483-498.

Wolf, J., M. Loechl, M. Thompson, and C. Arce (2003a). Trip Rate Analysis in GPS-Enhanced Personal Travel Surveys. In *Transport Survey Quality and Innovation*,. Elsevier: London, England, pp. 483-498.

Wolf, J., M. Oliveira, and M. Thompson (2003b). The Impact of Trip Underreporting on VMT and Travel Time Estimates: Preliminary Findings from the California Statewide Household Travel Survey GPS Study. In press, Transportation Research Record for 2003 Annual Transportation Research Board Conference. Washington, DC.

Zmud, J., and C. Arce (2000). Item Nonreponse in Travel Surveys: Causes and Solutions. In *Transport Surveys: Raising the Standard*. Proceedings of an International Conference on Transport Survey Quality and Innovation, May 24-30, 1997. Grainau, Germany.