SOUTHERN CALIFORNIA ASSOCIATION OF GOVERNMENTS

YEAR 2000 POST-CENSUS REGIONAL TRAVEL SURVEY

GPS Study Final Report

February 3, 2004

NuStats

3006 Bee Caves Rd., Suite A-300 ▪ Austin, Texas 78746
(512) 306-9065 ▪ fax (512) 306-9077 ▪ *www.nustats.com*

Contact: Johanna Zmud, *Project Director*

Battelle
The Business of Innovation
505 King Avenue
Columbus, OH 43201
www.battelle.org/transportation

Contact: Mark Lepkowski, 202-646-7786

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1.0 – INTRODUCTION AND OBJECTIVES

The Southern California Association of Governments (SCAG) is conducting a regional household travel survey. NuStats Partners (NuStats) is serving as the data collection agent for this study. Battelle provided data collection support to NuStats on this project to collect information using global positioning system (GPS) technology from a sub-sample of households. By collecting both self-reported computer-assisted telephone interview (CATI) information and in-vehicle GPS information on trips, data from the GPS study enhances the understanding of where and why people in the study area travel. The overall goals of the GPS portion of the study were:

- Primary Purpose: To detect under-reporting of trips in respondent self-reports for the CATI method of capturing travel information. GPS-derived data will be compared to respondent self-reported data to identify missed trips.
- Secondary Purpose: To determine the feasibility of using GPS to develop household trip rate correction factors.
- Tertiary Purpose: To evaluate the accuracy of reported trip elements (e.g., trip start and finish times, locations, durations, lengths).

The GPS portion of the survey effort was conducted in three waves: June 2001, September 2001 though December 2001, and January 2002 through March 2002. Four hundred households were targeted for participation in the GPS subsample. Eligible travel days included all weekdays, but excluded holidays. As part of the survey effort, across all three phases, 198 GPS Leader™ devices were shipped to a total of 820 households to be installed in 1,217 vehicles. Roughly 470 households participated in the study and installed the GPS devices in their household vehicles. Therefore, among those households to which equipment was shipped, roughly 60 percent participated in the study. Of these 470 households, data from 293 households was available for analysis of trip under-reporting behavior. The number of households that were available for analysis decreased for several reasons. First, not all 470 households had both self-reported and GPS data. Second, with contract delays, the final edit checks of self-reported CATI data were completed nearly one year after Battelle delivered the GPS data file. Several potential GPS study cases did not pass the final edit checks.

This document summarizes the results of the GPS portion of the overall study, henceforth referred to as the SCAG GPS Study. It was prepared both by Battelle and by NuStats. Battelle documented the GPS study methods and outcomes of the field experience, and NuStats analyzed the resultant data, using the final delivered data files. Appendix A contains a data dictionary, compiled by Battelle, for the accompanying GPS datasets.

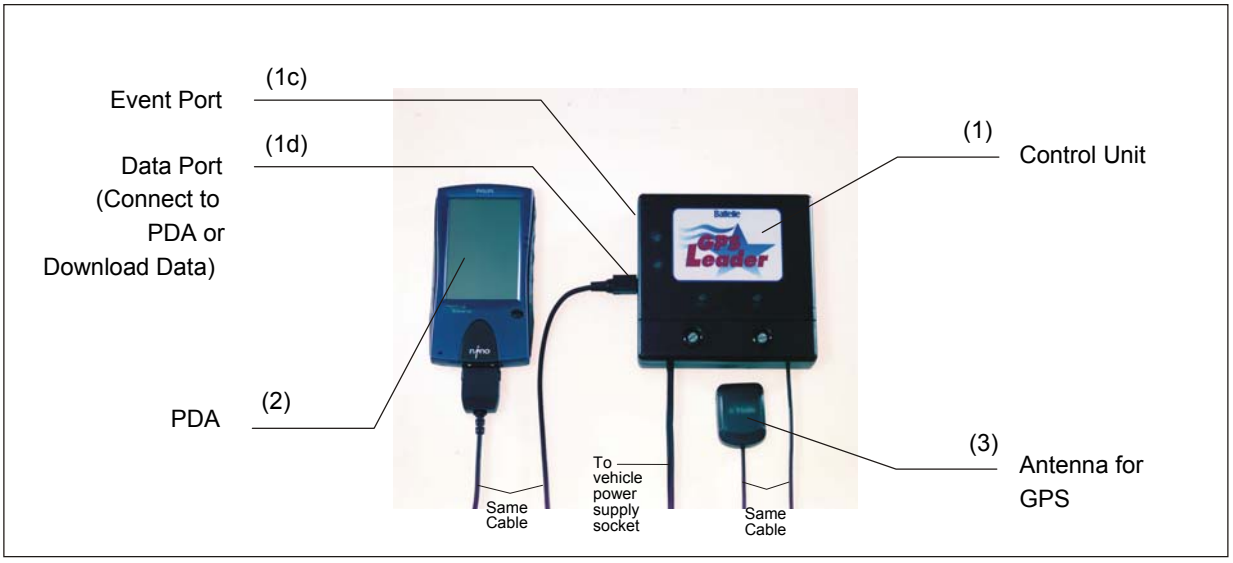
2.0 – GPS LEADER™ EQUIPMENT

This project used Battelle's GPS Leader™ devices (Figure 1) to capture the vehicle-based travel in household vehicles. This device consists of two basic components: a combined GPS receiver and storage component and a hand-held personal digital assistant (PDA). For this study, the GPS receiver was set to capture position (in latitude and longitude format) and speed information at 1-second intervals. To conserve storage space, if the measured speed of the vehicle fell below 3.75 miles per hour, the devices were set to sample at 30-second intervals. The output from the 8-channel, continuous tracking GPS receiver was not differentially corrected. There was a 12 MB capacity of memory, which was sufficient to store over 60 hours of collected GPS and user-generated data records.

The GPS Leader™ devices were powered through the vehicle's accessory power port/cigarette lighter. The power management circuitry sensed when the vehicle's engine was started or turned off. The device automatically activated and initiated the collection of GPS position fixes when the vehicle engine was turned on. In addition, respondents were asked to utilize the hand-held PDA at the start of each trip to identify the household members on the trip and to record a trip purpose for each occupant.

The equipment was designed to be self-installed by survey participants. Installation of the equipment consisted of mounting the GPS antenna on the roof of the household vehicle, plugging the GPS receiver into the accessory power port/cigarette lighter, and stowing all cables. Installation of the equipment typically took less than five minutes per household vehicle and was successfully performed by the vast majority of survey participants. In a handful of cases, the data showed a lack of satellite fix, implying that the GPS antenna had not been correctly mounted on the roof of the vehicle.

FIGURE 1:
DIAGRAM OF BASIC SYSTEM COMPONENTS



The GPS receiver is capable of reporting position (in latitude and longitude format) and speed information at up to one-second intervals. The output from the 8-channel, continuous tracking GPS receiver, housed within the control unit, is not differentially corrected, and can be expected to perform to the following specifications:

Accuracy:	Position:	25 meters circular error (50%) without selective availability	
	Velocity:	0.1 meters/second without selective availability	
Time to Fix:	Cold start:	< 2 minutes (90%)	– requires no initialization
	Warm start:	< 45 seconds (90%)	– with current position, time, almanac
	Hot Start:	< 20 seconds (90%)	– also with emphemeris
	Reacquisition:	< 2 seconds (90%)	– temporary loss of fix

3.0 – SURVEY METHODS

The GPS portion of the survey 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 a reminder telephone call attempt the night before the assigned travel day; and 5) collecting travel information for trips that occurred on the assigned travel day from household members via a CATI. Battelle coordinated the distribution and collection of the GPS Leader™ device while NuStats was responsible for identifying, recruiting, and screening households; providing successfully recruited households with travel diaries and other necessary information; and interviewing the households to collect information recorded on participants' diaries.

In addition to all of the normal survey procedures, a few additional procedures focusing on the recruitment of GPS households and retrieval of the GPS equipment were employed that were unique to GPS households. These activities included selecting households from those initially recruited to receive the GPS Leader™ devices, processing the devices for shipment, retrieving the devices, and processing the data collected by the devices. Additionally, for efficiency, it was critical that the equipment be received back from the respondents in a timely fashion. Therefore, a nominal incentive of $10 was utilized to promote the prompt installation, record keeping, and return of the equipment. A $5 bill was initially shipped with each GPS Leader™ device, and after the device was returned, an additional $5 was sent to households for each device that was used in the study. These incentives were not used in non-GPS households.

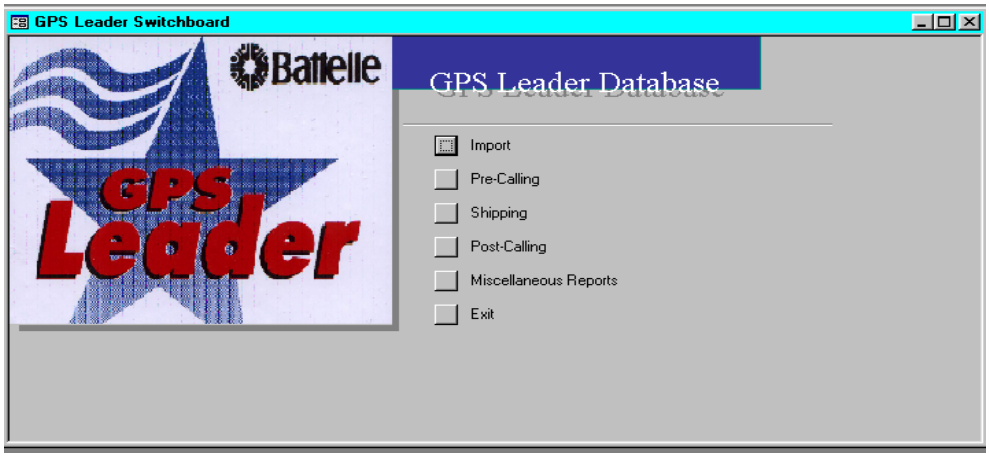
The following sections describe the additional activities that were conducted with households that were selected for participation in the GPS subsample.

3.1 INITIALIZATION OF GPS RECRUITS DATABASE

Households were initially recruited for participation in the GPS portion of the survey at the end of the recruitment interview for the main study. As part of the initial recruitment effort, information such as a contact name, address, and telephone number; the number of people in the household; the name, age, and driver status of each household member; the number of vehicles in the household; the specific year, make, and model of each household vehicle (up to three); and the assigned travel day were collected.

This information was loaded into an Access database specifically designed for this study. This database served as the central tracking and control database for the entire GPS portion of the study. Data on the status of every deployed GPS Leader™ device, recruitment information, pre-deployment household contacts, and post-deployment follow-ups were all available at a moment's notice from this centralized database. Data and reports on information from this master database were available through a built-in menu, which is illustrated in Figure 2.

FIGURE 2:
ILLUSTRATION OF SCAG GPS MASTER CONTROL DATABASE MENU



3.2 PRE-DEPLOYMENT TELEPHONE CONTACT

The GPS Leader™ database described in Section 3.1 allowed for the selection of a subset of the recruited households for shipment, based on the number of GPS devices available for that travel day. The night prior to shipping the GPS devices, the selected households were called to verify their shipping address, contact name, and their continued willingness to participate. Up to four calls were made to each household between 3:00 p.m. and 7:00 p.m. local time on Monday, Wednesday, and Thursday (devices were not shipped to households on Wednesday and therefore no calling was performed on Tuesday night). Standardized calling scripts that were generated each night from the SCAG GPS Leader™ Master Control Database were used for the pre-deployment telephone contact.

Answering machine messages were left, if possible, with those households that could not be reached directly. The message indicated that GPS devices would be delivered to the household as part of the study and provided a toll-free number for the project in case there were any questions by survey participants. Prior to the September 11th disaster, devices were shipped to households even if this preliminary contact was not made. However, following the disaster, feedback from survey respondents indicated that a large number of devices would be refused by participants who were fearful of receiving a package from an "unknown" person. Contact with at least one member of the household provided an opportunity to describe the box (i.e., "….it is about the size of a white shoebox and has the SCAG and Battelle GPS Leader™ stickers on the front…") and address questions and concerns by survey participants. Following September 11th, no GPS devices were shipped without at least leaving a message with the household. If a household could not be reached and no message could be left (e.g., busy signal on each call attempt or disconnected telephone line), the selected household was replaced with another household.

The results of this pre-telephone contact were entered into the SCAG GPS Leader™ Master Control Database. A shipping report was generated and used to process the GPS equipment for the next day's deployments.

3.3 PRE-DEPLOYMENT EQUIPMENT PROCESSING

Two primary activities were conducted to prepare each GPS Leader™ device for shipment. First, each device was programmed specifically for the household to which it was shipped, including an identification of the vehicle to which it was assigned. Each device was programmed to include the correct date and time for southern California, the household and vehicle identification numbers, and the names of all of the household members. If the recruit information indicated that the household member was a licensed driver, that person was entered in the list of drivers. All other household members were listed as potential passengers. This allowed the traveler to simply choose from a list of names when entering his/her trip information. To complete the preparation of the GPS Leader™ device, the battery level for each device was checked and new batteries were installed if necessary.

The next set of activities involved preparing the shipping and field materials. The GPS Leader™ devices were shipped individually in boxes containing return shipping material (box tape, return shipping label, cigarette lighter y-splitter), along with a list of the Federal Express drop locations nearest to the household. Field materials such as a laminated Quick Installation Guide, a Use Guide, a detailed Installation Manual, a detailed Use Manual, answers to frequently asked questions (FAQ), an installation and use video tape, and a personalized letter also were included in each box along with the shipping materials and GPS Leader™ device.

The appropriate number of devices was then shipped to the contacted households using overnight Federal Express. Each device sent to a household was shipped in a separate box (i.e., a three-vehicle household would receive three boxes). The use of separate boxes facilitated the return of the equipment, as these boxes would fit inside the "box" opening on a Federal Express drop-box. Larger boxes would require that they be dropped-off inside of a Federal Express office.

Shipments of GPS Leader™ devices were planned so that the survey participants would receive the devices at least one full day before their travel day. That is, devices were shipped via overnight Federal Express on Monday for a Thursday travel day, Tuesday for a Friday travel day, Thursday for a Monday or Tuesday travel day, and Friday for a Wednesday travel day. There were no shipments on Wednesdays.

3.4 POST-SHIPMENT EQUIPMENT PROCESSING

Two sets of activities were conducted once the GPS Leader™ devices were returned to Battelle including those needed to update the SCAG GPS Leader™ Master Control Database (i.e., for device tracking) and those performed to upload the information collected by the GPS equipment.

Immediately upon their receipt, each returned box was examined for wear and tear and the device identification number recorded on the outside of the box was logged and compared to the identification numbers recorded on the equipment. Next, the household identification number, device identification number, and the apparent status of the equipment were logged and entered into the database. Generally, four types of returns were encountered: 1) devices where the informed consent envelope (the protective envelope containing the equipment) was opened and the device was used, 2) devices where the informed consent envelope was opened and the device was not used, 3) devices where the informed consent envelope was not opened, and 4) devices where the boxes were not opened (i.e., they were undeliverable or refused delivery by the survey participant). The use status, missing equipment or field materials, and any notes from the survey participants were logged and then entered into the SCAG GPS Leader™ Master Control Database.

Using software written explicitly for the GPS Leader™ equipment, the trip information was uploaded from each device into a centralized database, maintained separately from the SCAG GPS Leader™ Master Control Database. The two databases are linked by the unique device and household identification

codes, but were maintained separately to preserve the confidentiality of the data. Staff responsible for programming the household information in the GPS Leader™ devices, and for preparing the boxes for shipment did not have access to the data collected by the equipment.

3.5 POST-CALLING/LETTERS

Telephone follow-up was initiated for each household that did not return all of the devices within two weeks following the assigned travel day. A Battelle team member attempted to contact each household every weeknight between 3:00 p.m. and 7:00 p.m. local time until contact with a household member was made. Once contact with a household member was made, the Battelle team member requested that the overdue devices be returned, provided assistance locating nearby Federal Express locations, arranged for Federal Express pick-up, etc. The results of each conversation were recorded and logged into the SCAG GPS Leader™ Master Control Database. If the respondent indicated that he/she had already shipped the devices, no further telephone contact was attempted for at least three days, at which point the process was repeated again if the devices still had not been received.

If the household did not return all of their devices within three weeks following the assigned travel day (i.e., two weeks plus an additional week of telephone follow-up), then a letter was sent via Federal Express requesting that the devices be returned immediately and providing information on nearby drop box locations. Additionally, this letter reminded the households that they would receive and additional $5 for each *used* device that was returned.

Households that had not retuned their devices within four weeks following the assigned travel day were sent a second letter via Federal Express. This letter also requested that the devices be returned immediately. However, in this letter, households were offered $50 if they returned all of the devices assigned to their household (regardless of whether they were used or not).

Obviously, there were special cases that arose during the post-calling, such as the respondent indicating that they destroyed the devices. In some of these cases, attempts at retrieving the equipment were made by attempting to contact a different household member. However, in some cases it became clear that the devices were actually destroyed or otherwise unavailable. Telephone and hardcopy follow-up with these households was discontinued.

Federal Express was routinely contacted to track and retrieve overdue devices. Local post offices also were contacted in case the devices were inadvertently placed into a U.S. Priority Mail, rather than a Federal Express, box.

4.0 – DATA PROCESSING

As discussed in previous sections, the raw trip information collected by the GPS equipment was uploaded into a centralized database for further processing. Post-processing of the trip data consisted of data editing and cleaning and then matching the GPS-recorded trips to diary-reported trips. Both the data editing and trip matching were accomplished through a blend of computer and manual edits.

4.1 EDITING AND CLEANING PROCEDURES

As with any technology deployment, several different types of users of the equipment were observed ranging from those that utilized the equipment perfectly to those that chose not to utilize the Personal Data Assistant (PDA). The collected trip information was processed and edited to account for the various types of users and the resulting collected information. Summaries of the types of processing that occurred are provided below.

The Respondent Forgot to End Trip Using PDA: In some instances, the respondent initialized a trip using the PDA but did not end the trip using the PDA (i.e., by pressing the "End Trip" button) when he/she reached his/her destination. Typically, in these situations, the respondent would "end" his/her previous trip when they returned to his/her vehicle and were ready to begin the next trip. This type of data edit was identified through an examination of the trip profiles, time between trips, the number of GPS records collected below idle speed, and a discrepancy between the number of miles traveled and the duration of the trip. The GPS records associated with this excess time were removed from the trip.

The Respondent Was Practicing Entering Trips With the PDA: Mostly these instances occurred on the day that the respondent received and installed the equipment for the first time, though some cases were observed when the vehicle drivers changed. In these cases, several very short (i.e., 5-10 second) trips were observed consecutively without any travel. These trips were removed from the database.

The Respondent Did Not Indicate the Trip Start and Trip End of a Trip for All Stops: As with diary information, respondents often have a different interpretation of the definition of a trip than do researchers. Respondents utilizing the PDA equipment to start and end a trip did not always start and end the trip for every location. That is, they did not consider travel to intermediate stops to be trips. In these instances, the trip was separated into the appropriate number of trips. Because the number of vehicle occupants and the trip purpose are entered after the respondent selects the "End Trip" button, this information was associated with the last trip in the chain. Preceding trips in the same trip-chain were treated as non-PDA trips.

One difficulty in identifying and correctly splitting these trips that was particularly prevalent in the Southern California area was the ability to distinguish between trips with many seconds of idle speeds due to traffic congestion and trips that should have been appropriately split because the respondent stopped at a location. Because of the nature of travel in the study area, a set algorithm for splitting trips that have many seconds of no travel was not employed. Rather, the speed profiles for each and every trip were examined and used, along with the time where speed was less than idle and ignition and power records, to determine on a trip-by-trip basis, whether to split the trip because of a missed stop.

Trimming Trip Start Times and Trip End Times to Account for Idling: In some instances, PDA-initiated trips had significant idling time at the beginning or ending of a trip. For example, it appeared that some users took several minutes at the end of a trip to input the number of vehicle passengers and the trip purposes for each passenger. In other cases, the car ignition was activated, but the car did not move for many minutes (e.g., the driver was waiting for a passenger to arrive, but had already started the car). These situations were also identified by examining the speed profiles and time that the vehicle stood at

idle. If identified as having excessive idling, the trip start and end times were adjusted to reflect actual movement by the vehicle.

Removing the Impact of False Ignition Record: The GPS Leader™ devices automatically sense a vehicle ignition and begin to capture trip (GPS) information. This ignition sensing is based upon identifying a power surge that exceeds a threshold value for a set amount of time. Because all vehicles are different and a unique threshold value cannot be reliably determined for each vehicle remotely, a common threshold value was used. To be conservative, a relatively low threshold was employed so that non-PDA trips were not inadvertently missed. However, in some instances this creates false ignition records, resulting in subsequent trips that were later identified and removed from the data. These trips are easily identified through examination of the number of GPS records where speed is greater than idle and the profile of ignitions recorded by the GPS Leader™ device. If the pattern of ignition records suggested that the vehicle was one where the threshold was too low, and the non-PDA trip was of short duration with no travel, then this trip was removed.

False ignition records could also inadvertently split non-PDA trips into more than one trip. These were identified through an examination of the speed profiles. Trip combining was performed when one trip ended and the next began within 10 seconds and the speed indicated that travel was continuous.

Removal of Erroneous Non-PDA Trip Immediately Preceding a PDA-Trip: As mentioned previously, the GPS Leader™ device initialized a trip as soon as it sensed an ignition. This was designed to prevent the equipment from missing trips. In some instances, the vehicle occupants initialized a PDA trip within a few seconds of this non-PDA trip. That is, they got into their vehicle, turned the vehicle on (thus initializing a non-PDA trip), and then proceeded to initialize a PDA trip. Therefore, the non-PDA trip, for which there was no travel, was superceded by a PDA trip and was removed from the database.

4.2 TRIP MATCHING

Merging the trips recorded on a household's travel diary (interview trips) with the corresponding trips recorded on the GPS Leader™ device (GPS trips) was accomplished through the use of computer algorithms followed by a manual review and matching process. First, the diary information and the GPS trip information were brought together into a common database. A computer program then attempted to match corresponding trips. This program first sorted the trips in the two sets of data by time (and date) of day. It then compared the start-time, end-time, trip duration, trip distance, etc., for the first trip from both sets of data. If these trip parameters matched within reason (see below), the trips were considered a match. If it was determined that the trips did not match, the first interview trip was then compared with the second GPS trip. This iterative process continued until all interview trips were either matched against GPS trips or were determined not to have a valid computer match. The computer program utilized the following criteria for determining whether a match had been made:

- The trip start and end times both matched (rounded to the minute)
- The trip start and end times both matched within plus-or-minus ten minutes
- The trips had identical start times (rounded to the minute) and have identical trip purposes.
- The trips had nearly identical start times (plus-or-minus five minutes) and have identical trip purposes.
- Trip start and end times are within one hour and the GPS recorded latitude and longitude matched the geocoded latitude and longitude within a 1,000 feet of each other.

Following the processing of these computerized matches, a Windows-based interface was used to display the results of the trip matching. This interface allowed for a manual review of each match and adjustments as necessary. Manual review of the trip information sometimes revealed a pattern to the trips that was not captured by the computer algorithm. For example, cases occurred where the trip distance, duration, and vehicle occupants matched but the start and end times consistently differed by over an hour. These, and other such, cases were manually reviewed and matched on a case-by-case basis. See Table 1 for results of the matching process.

TABLE 1:
RESULTS OF MATCHING PROCESS

MATCH TYPE	FREQUENCY	PERCENT	CUMULATIVE PERCENT
Trip start and end times both matched (rounded to the minute)	94	2.3	2.3
Trip start and end times both matched within ± 10 minutes	805	19.7	22.0
Trips have nearly exact match (± 5 min) for start time and identical trip purposes	36	0.8	22.8
Start and end times within 1 hour, ending longitude and latitude within ± 1000 feet of each other	292	7.1	29.9
Manual Match	981	24.0	53.9
Not Matched	1877	45.9	99.8
Total	**4085**	**99.8**	**--**

Note: Base is vehicle trips captured in GPS and / or recorded in CATI diaries.
Total does not sum to 100.0 percent due to rounding.

5.0 – SUMMARY OF FIELD EXPERIENCES

The GPS portion of the survey effort was conducted in three waves: June 2001, September 2001 though December 2001, and January 2002 through March 2002. Eligible travel days included all weekdays, but excluded holidays. As part of the survey effort, across all three phases, 198 GPS Leader™ devices were shipped to a total of 820 households to be installed in 1,217 vehicles. Roughly 470 households participated in the study and installed the GPS devices in their household vehicles. Therefore, among those households to which equipment was shipped, roughly 60 percent participated in the study. Table 2 summarizes the deployment participation.

TABLE 2:
DEPLOYMENT STATUS AND PARTICIPATION

Number of Households Deployed	820
Number of Unit Deployments	1,217
Number of Households Returned	794
Number of Units Returned	1,185
Number of Households Providing GPS Information	469
Number of Households Where All Devices Used	373
Number of Households Where Some of the Devices Were Used	96
Participation Rate (Returned/Deployed)	97%
Completion Rate (Completed/Returned)	59%
Response Rate = Participation * Completion	57%

Originally, the study was to be conducted using a total of 50 devices. However, delays in the initiation of the GPS field period, lower recruitment and participation rates, and a longer device turn-around time than expected required that the number of deployed devices be increased during the course of the project to meet schedule and completion requirements. Overall, 198 devices were utilized for this project. Table 3 summarizes the number of initially recruited households, number of usable households (based upon the results of the pre-deployment telephone contact), and the number deployed devices by week

TABLE 3: DEPLOYMENT SUMMARY BY WEEK

WEEK OF:	RECRUITED HOUSEHOLDS	USABLE HOUSEHOLDS	ELIGIBLE VEHICLES IN USABLE HOUSEHOLDS	GPS UNITS SENT	GPS UNITS RETURNED
June 4	36	36	67	68	8
June 11	48	48	85	41	49
June 18	117	117	171	36	47
June 25	23	23	42	14	55
September 10	44	36	63	0	0
September 17	22	9	20	27	0
September 24	46	32	67	65	14
October 1	6	2	4	9	34
October 8	24	19	39	41	34
October 15	133	100	202	69	40
October 22	188	162	319	61	56
October 29	323	262	479	50	57
November 5	262	214	373	69	39
November 12	32	30	62	48	55
November 19	322	262	447	34	21
November 26	284	244	445	49	50
December 3	153	146	255	67	59
December 10	2	2	4	3	61
December 17	0	0	0	0	28
December 24	0	0	0	0	6
December 31	18	16	27	19	9
January 7	89	86	161	22	5
January 14	301	295	556	79	29
January 21	349	337	622	88	44
January 28	408	402	724	76	57
February 4	291	282	527	57	76
February 11	338	331	579	58	62
February 18	496	490	853	67	58
February 25	429	429	740	0	67
March 4	301	301	538	0	27
March 11	176	176	307	0	15
March 18	0	0	0	0	11
March 25	0	0	0	0	5
April 1	0	0	0	0	3
April 8	0	0	0	0	1
April 15	0	0	0	0	0
April 22	0	0	0	0	1
April 29	0	0	0	0	0
May 6	0	0	0	0	0
May 13	0	0	0	0	1
May 20	0	0	0	0	0
May 27	0	0	0	0	0
June 3	0	0	0	0	0
June 10	0	0	0	0	0
June 17	0	0	0	0	0
June 24	0	0	0	0	1

There are 26 households (32 devices) that did not return the GPS equipment, which represents roughly three percent of all deployed households or devices. Extensive efforts were made to retrieve the devices from these households, but were discontinued after several months of exhaustive attempts. In a few cases, the respondents have indicated that they have either destroyed the equipment (2174841) or had it stolen (2258101). However, in the bulk of the cases, the devices were unrecoverable because the study participant moved and could not be found or after many attempts (usually more than 20-40 call attempts) it became apparent that further follow-up would be unproductive. Table 4 summarizes the households and reasons why the equipment is presumed to be lost.

TABLE 4:
SUMMARY OF UNRETURNED EQUIPMENT

HOUSEHOLD ID	NUMBER OF DEVICES	TRAVEL DATE	DESCRIPTION
2121859	1	6/8/01	Unable to be retrieved from respondent.
2119219	1	6/18/01	Unable to be retrieved from respondent.
2118930	3	6/29/01	Unable to be retrieved from respondent.
2198552	2	11/14/01	Devices thrown away by respondent
2024653	1	9/21/01	Unable to be retrieved from respondent
2174841	1	10/30/01	Device destroyed by respondent
2087862	1	10/19/01	Unable to be retrieved from respondent. Respondent moved and is untraceable.
2202844	1	10/26/01	Respondent indicates returned device, but never received.
2166406	2	11/13/01	Respondent indicates returned device, but never received.
2187464	1	11/28/01	Respondent indicates returned device, but never received.
2258101	2	12/12/02	Devices likely stolen from respondent.
2164905	1	11/8/01	Respondent indicates that they never received device though signature is on file with FedEx.
2204408	2	9/1/02	Unable to be retrieved from respondent.
2342649	1	2/27/02	Unable to be retrieved from respondent.
2345724	1	1/18/02	Unable to be retrieved from respondent.
2354951	1	1/17/02	Unable to be retrieved from respondent.
2356193	1	1/22/02	Unable to be retrieved from respondent.
2366654	1	1/25/02	Unable to be retrieved from respondent.
2388458	1	1/31/02	Unable to be retrieved from respondent.
2397953	1	1/16/02	Unable to be retrieved from respondent.
2421434	1	1/16/02	Unable to be retrieved from respondent.
2447998	1	2/7/02	Unable to be retrieved from respondent.
2469813	1	2/11/02	Unable to be retrieved from respondent.
2482494	1	2/21/02	Unable to be retrieved from respondent.
2535831	1	2/22/02	Unable to be retrieved from respondent.
2537628	1	2/11/02	Unable to be retrieved from respondent.
26	**32**	**Total**	

Table 5 summarizes the distribution of households by the number of eligible vehicles and vehicles providing GPS information. Approximately three-fourths (77.6 percent) of the households that used the GPS equipment had data collected from all eligible vehicles. However, having no GPS data for all household vehicles does not necessarily construe a lack of participation by the household; a very likely reason could be that the other household vehicles were not used on the Travel Day.

TABLE 5:
DISTRIBUTION OF HOUSEHOLDS BY NUMBER OF ELIGIBLE VEHICLES & VEHICLES PROVIDING GPS INFORMATION

NUMBER OF VEHICLES WITH GPS DATA	NUMBER OF ELIGIBLE VEHICLES			
	1	2	3	TOTAL
1	254	69	11	334
2	--	92	25	117
3	--	--	18	18
Total	**254**	**161**	**54**	**469**

Among the 631 GPS Leader™ devices that were installed in household vehicles, slightly less than one-half (48.3 percent) were installed and used on more than one day. On average, each installed device was used for 1.95 days (a 95% confidence interval, 1.8 to 2.1 days). Table 6 summarizes the distribution number of days that GPS travel was collected by the installed devices. The additional days of data provide an opportunity to investigate the temporal changes in travel for the same households. However, this analysis is beyond the scope of this project. Table 7 summarizes installed devices and corresponding households by day of the week.

TABLE 6:
NUMBER OF DAYS OF TRAVEL INFORMATION COLLECTED BY INSTALLED GPS LEADER™ DEVICES

NUMBER OF DAYS OF TRAVEL COLLECTED BY GPS DEVICE	NUMBER OF DEVICES	PERCENT OF DEVICES
1	326	51.7%
2	161	25.5%
3	78	12.4%
4	37	5.9%
5	9	1.4%
6	7	1.1%
7	4	0.63%
8	3	0.47%
9	4	0.63%
10+	2	0.31%
Total	**631**	**100%**

TABLE 7:
NUMBER OF HOUSEHOLDS AND DEVICES WITH GPS INFORMATION BY DAY OF THE WEEK

DAY OF THE WEEK	USING DATA FROM ALL DAYS OF TRAVEL		USING DATA ONLY FROM HOUSEHOLDS TRAVELING ON ASSIGNED TRAVEL DAY	
	NUMBER OF DEVICES[A]	NUMBER OF HOUSEHOLDS[A]	NUMBER OF DEVICES W/ DATA	NUMBER OF HOUSEHOLDS
Monday	165	134	78	59
Tuesday	199	164	69	52
Wednesday	261	210	160	121
Thursday	227	192	104	86
Friday	202	166	136	103
Saturday	82	74	--	--
Sunday	75	66	--	--
Total	**1,211**	**1,006**	**547**	**421**

a: Includes multiple days with the same device/household. Thus, devices/households that provided information on one day of the week more than once would be counted more than once for that day of the week (e.g., a household/device providing data on two separate Sundays would be counted as two separate devices on Sunday

6.0 – DATA ANALYSIS

One of the key focuses of the SCAG GPS Study was to detect under-reporting of trips in respondent self-reports using a diary/telephone retrieval method of capturing travel information. To accomplish this, GPS-derived trip data were compared to respondent self-reported trip data for those households that participated both in the GPS study and in the travel diary study. Battelle defined the GPS trips through computerized and manual processes, as described in Section 4 of this report. For the CATI (diary/telephone retrieval) data, respondent self-reports were used to define the trips. For this analysis, only driver and passenger vehicle trips were used for the comparative analysis. There were 293 households for which both GPS and respondent self-reported vehicle trip data were available for analysis.

6.1 COMPARISON OF TRIP RATES FOR GPS STUDY AND NON-GPS STUDY HOUSEHOLDS

First, a direct comparison of the trip rates for these 293 households that participated in the SCAG GPS Study was made to the corresponding trip rates among households (with vehicles) that did not participate in the SCAG GPS Study. This analysis sheds light on whether the additional procedures utilized in the SCAG GPS Study, such as additional telephone contact, incentives and receipt of equipment, may have caused the participants in the GPS study to be more meticulous in reporting their travel than those participating in the main survey. Table 8 presents vehicle trip rates by household size for households that participated in the SCAG GPS Study and those that did not. These estimates of trip rates indicated that participants in the GPS Study were better reporters of travel than those persons not participating in the SCAG GPS study. That is, the additional recruitment efforts, receipt of equipment, and incentives increased the salience of recording trips in the travel diaries. It also appeared that the larger the household, the greater the impact on trip reporting.

TABLE 8:
COMPARISON OF VEHICLE TRIP RATES BY HOUSEHOLD SIZE

SAMPLE	TOTAL	1-PERSON HH	2-PERSON HH	3-PERSON HH	4-PERSON HH	5+-PERSON HH
GPS Study	8.26	4.72	8.61	9.72	12.03	15.47
Non-GPS Study	6.32	3.42	5.93	7.57	9.37	10.84

Note: Data are unweighted, linked vehicle trips.

6.2 QUANTIFICATION OF MISSED TRIPS IN RESPONDENT SELF-REPORTED DATA

However, even with more meticulous participants in the GPS study, there is still the possibility that they did not self-report trips that were captured by the GPS because the respondent did not understand the research definition of a "trip," forgot to record a trip, etc. Quantifying the precise number of "missed" CATI trips is challenging due to the presence of "missed" trips in both CATI and GPS data. See Table 9. A "missed vehicle trip" is defined as a driver or passenger vehicle trip that was not self-reported by a CATI respondent (i.e., missed in CATI) or not captured by the GPS equipment (i.e., missed in GPS).

A total of 4,085 vehicle trips were recorded either by GPS equipment "or" self-reported by respondents. Of these, 2,209 vehicle trips were found in both GPS and CATI data, 1,178 vehicle trips were found in GPS but not CATI, and 698 were found in the CATI data but not the GPS data. Battelle flagged the 698 missing vehicle trips in the GPS data with one of seven values: no GPS fix, bad latitude, bad longitude, bad speed, bad time, bad distance, or data truncated due to edit of time.

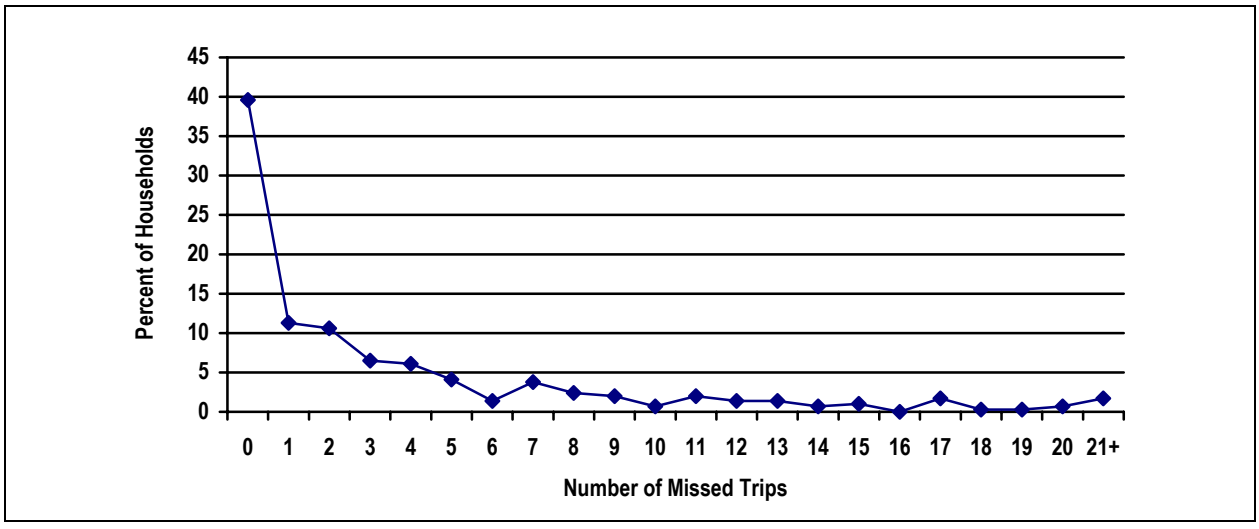
TABLE 9:
SUMMARY OF MISSED VEHICLE TRIPS - GPS AND CATI

		MISSED IN GPS		TOTAL
		NO	YES	
Missed in CATI	No	2209	698	2907
	Yes	1178	0	1178
Total		3387	698	4085

The rest of this analysis focuses on the 1,178 vehicle trips that were captured by the GPS but missing from the CATI data. Of the 3,387 vehicle trips captured by the GPS, 1,178 vehicle trips were missing in the self-reported CATI data, indicating a "missed" trip percentage of 34.8%. It is important to point out that this percentage represents both missed driver *and* missed passenger vehicle trips. The identification of both driver and passenger trips was enabled through the use of the PDA data. Prior studies of missed trips using GPS methods, focused solely on missed driver trips. Thus, direct comparisons of the percent of missed trips to other studies may not be appropriate. It would be expected that the percent of missed driver trips only for the SCAG data would be lower than the 34.8% noted above.

Figure 3 summarizes the incidence of missed trips among the 293 households for which both GPS and diary data were available for analysis. The number of missed trips ranged from zero (for 116 households) to 46 (for 1 household). These data indicate that most households are fairly accurate reporters of their travel. The vast majority of "missed" vehicle trips came from a relatively small percentage of households.

FIGURE 3:
INCIDENCE OF MISSED VEHICLE TRIPS IN RESPONDENT SELF-REPORTS
(N = 293 HOUSEHOLDS)



The households in the bottom right corner of Figure 2 could be considered "outliers" in terms of subsequent analysis. For example, the household that missed the 46 vehicle trips turned out to be a special case. It was a two-person household of retirees, who owned four vehicles that they used for delivering flowers. They did not record the flower deliveries in their diary but these were captured by the GPS. The next highest number of missed trips was 34 vehicle trips. This household was comprised of four members – two adults and two children ages 9 and 11. The missed vehicle trips were trips on which the children accompanied the adults. The children's' diaries indicated only trips from home to school and back home, whereas the GPS units indicated more trips on which the children were passengers, in

accompaniment of an adult. In subsequent studies of "missed" trips in respondent self-reports using GPS data for validation, it is important to ensure that the definition of the "trip" that is missing is comparably defined in for both GPS and CATI analysis. In this particular analysis, a missed trip was defined as either a driver or a passenger trip.

The results of this study indicate that there are some households (roughly 10 percent) for which an extreme number of vehicle trips were not self-reported in telephone retrieval interviews. However, the bulk of the households had a much lower percentage of trips that were not recorded during the telephone interviews. These data indicate that not all households are equal in the accuracy (or lack thereof) with which they reported trips in their diaries.

6.3 Evaluation of the Accuracy of Trip Duration and Purpose in Self-Reported Data

An additional project purpose was to assess the accuracy in reporting of two trip characteristics: trip duration and trip purpose. Trip durations for the matched trips were compared using a paired t-test. On average, the GPS trips (person-based) were 6.4 minutes shorter than the self-reported trips. However, the self-reported start and end times were within five minutes of the time reported by the GPS equipment for roughly 70 percent of all matched person-trips. Approximately 20 percent of all person trips reported by respondents during the telephone interview had start or end times that were more than 10 minutes from the times recorded by the GPS equipment. The matched information also provided an opportunity to compare trip purposes. The vast majority of trip purposes were in agreement and this association was statistically significant.

7.0 – DEVELOPMENT OF ADJUSTMENT FACTORS FOR "MISSED" TRIPS

An important objective of the SCAG GPS Study was the continued testing an experimental methodology being developed at NuStats to use GPS data from a subset of households in a household travel diary survey to correct the larger overall diary sample. This methodology relies on a multivariate logit model. The database of vehicle trip records was used to test the model. For purposes of this analysis, a missed trip was defined as a driver or passenger vehicle trip that was captured by the GPS equipment but missed in the self-reported CATI data.

7.1 VARIABLES ASSOCIATED WITH MISSED TRIPS

Based on analyses that were conducted previously for the California Department of Transportation (Caltrans), it was determined that four demographic variables and two trip characteristics might be significant predictors of "missed" trip reports in the SCAG household travel survey.[1] These variables were: number of vehicles owned, household size, household income, respondent age, trip duration (as recorded by GPS), and trip purpose (as recorded in the PDA). Table 10 presents summary statistics for these variables associated with numbers of "missing" trips. For purposes of the analysis shown in Table 10, trip purpose was recorded as a dichotomous variable that is 1 if a trip was discretionary and 0 if it was non-discretionary (defined as work or school).

TABLE 10:
DISTRIBUTION OF MISSING VEHICLE TRIPS FOR HOUSEHOLD SIZE, VEHICLE OWNED, HOUSEHOLD INCOME, TRIP DURATION, AGE, & TRIP PURPOSES (WEIGHTED DATA)

VARIABLE	MISSING VEHICLE TRIPS	TOTAL VEHICLE TRIPS	% MISSING OF TOTAL
Household Size			
1	121	446	27.1
2	240	787	30.5
3	180	446	40.4
4	306	788	38.3
5+	269	912	29.5
Total	**1116**	**3379**	**30.0**
Number of Vehicles			
1	389	1220	31.9
2	518	1566	33.1
3	159	465	34.2
4+	51	129	39.5
Total	**1117**	**3380**	**33.0**

[1] 2000-2001 California Statewide Household Survey, *Final Report*, prepared by NuStats for the California Department of Transportation.

TABLE 10: CONT.

VARIABLE	MISSING VEHICLE TRIPS	TOTAL VEHICLE TRIPS	% MISSING OF TOTAL
Household Income			
Less than $24,999	107	314	34.0
$25,000 -$34,999	104	234	44.4
$35,000 - $49,999	226	608	37.2
$50,000 - $74,999	265	774	34.2
$75,000 - $99,999	233	680	34.2
$100,000 or more	147	519	28.3
Total	**1082**	**3129**	**34.6**
Trip Duration (in min)			
0 - 5.4	392	874	27.1
5.5 - 9.4	257	825	31.2
9.5 - 17.4	234	851	27.5
17.5+	234	830	28.2
Total	**1117**	**3380**	**33.0**
Age			
Less than 25 years	241	847	28.5
26 – 40 years	187	946	19.7
41 - 55 years	102	593	17.2
56+ years	49	456	10.7
Total	**579**	**2842**	**20.4**
Trip Purpose			
Non-Discretionary	150	508	29.5
Discretionary	704	2609	26.9
Total	**854**	**3117**	**27.4**

Note: Based on vehicle trips records for 293 households subsequent to final edit checks. Not all vehicle trips had an age or trip purpose recorded in the GPS data.

A multivariate regression technique (Logistic Regression) was used to determine which variables, of the ones available for this analysis, had the most impact on "missed" vehicle trip reports in the SCAG GPS Study. 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 equipment logged trip records and 0 otherwise and x_i is a vector of associated characteristics that influence whether a trip will be "self-reported" in the telephone retrieval interview or not. This analysis estimated the conditional distribution of y_i given x_i, $\Pr[y_i | x_i]$. The variables that turn out to be statistically significant are household income, respondent age, and trip duration. Table 11 displays the output.

TABLE 11:
RESULTS OF LOGISTIC REGRESSION

VARIABLE	B	S.E.	SIG
Respondent Age	-.022	.003	.000
Trip Duration	-.023	.004	.000
Household Income	-.122	.026	.000
Recoded Trip Purpose	-.175	.131	.183
Number of Vehicles	.019	.060	.749
Household Size	-.006	.029	.831
Constant	.416	.250	.096

As indicated in Table 11, the logit analysis identified three variables as being significantly associated with "missed" vehicle trips in the respondent self-reports – respondent age, household income, and trip duration. The remaining three variables were found to be insignificant.

7.2 CALCULATION OF GPS ADJUSTMENT WEIGHT

Based on the logit results, a variable "class" was created representing the 3-way cross table of the three significant variables. Each of the significant variables was recoded as noted in Table 12 to create the 3-way cross tab.

TABLE 12:
CODING PROTOCOL TO CREATE "CLASS" VARIABLE

VARIABLE	VALUE	LABEL
Trip Duration	1	0 – 5.5
	2	5.5 – 9.5
	3	9.5 – 17.5
	4	17.5+
Household Income	1	<$34,999
	2	$35,000 - $49,000
	3	$50,000 - $74,999
	4	$75,000 - $99,000
	5	$100,000+
Age	1	<25
	2	26 - 40
	3	41 - 55
	4	56+

Note: Missing data for household income was imputed using hot deck imputation using zip code, number of workers and number of household vehicles. Missing data for age was imputed by identifying the ages of other household members and making assumptions about related spousal or off-spring ages.

Based on this coding protocol, a matrix representing the cross-tab of the three significant variables was created, which was then used to derive the adjustment weight for specific household "classes". See Table 13 (variable labels for the values in columns under trip duration, household income, and age are identified in Table 12). Within each of the resulting household classes, the total number of driver or passenger vehicle trips (as defined by the universe of GPS equipment logged trip records) was divided by the number of self-reported driver or passenger vehicle trips to derive a GPS adjustment weight.

TABLE 13:
MATRIX TO DERIVE GPS ADJUSTMENT WEIGHTS BASED ON HOUSEHOLD CLASS

CLASS	TRIP DURATION	HH INCOME	AGE	SELF-REPORTED VEHICLE TRIPS	TOTAL VEHICLE TRIPS (W/ GPS)	GPS ADJUSTMENT WEIGHT
111	1	1	1	13	22	1.6923
112	1	1	2	27	45	1.6667
113	1	1	3	19	27	1.4211
114	1	1	4	10	11	1.1000
121	1	2	1	15	45	3.0000
122	1	2	2	43	50	1.1628
123	1	2	3	35	49	1.4000
124	1	2	4	18	19	1.0556
131	1	3	1	7	9	1.2857
132	1	3	2	32	47	1.4688
133	1	3	3	14	19	1.3571
134	1	3	4	16	27	1.6875
141	1	4	1	41	56	1.3659
142	1	4	2	39	61	1.5641
143	1	4	3	21	32	1.5238
144	1	4	4	10	13	1.3000
151	1	5	1	30	34	1.1333
152	1	5	2	31	32	1.0323
153	1	5	3	23	31	1.3478
154	1	5	4	40	49	1.2250
211	2	1	1	23	28	1.2174
212	2	1	2	25	38	1.5200
213	2	1	3	21	22	1.0476
214	2	1	4	15	16	1.0667
221	2	2	1	19	47	2.4737
222	2	2	2	46	47	1.0217
223	2	2	3	27	30	1.1111
224	2	2	4	16	16	1.0000
231	2	3	1	50	70	1.4000

CLASS	TRIP DURATION	HH INCOME	AGE	SELF-REPORTED VEHICLE TRIPS	TOTAL VEHICLE TRIPS (W/ GPS)	GPS ADJUSTMENT WEIGHT
232	2	3	2	45	65	1.4444
233	2	3	3	17	22	1.2941
234	2	3	4	21	24	1.1429
241	2	4	1	64	72	1.1250
242	2	4	2	31	49	1.5806
243	2	4	3	27	34	1.2593
244	2	4	4	12	13	1.0833
251	2	5	1	20	26	1.3000
252	2	5	2	30	33	1.1000
253	2	5	3	20	24	1.2000
254	2	5	4	41	43	1.0488
311	3	1	1	28	32	1.1429
312	3	1	2	24	27	1.1250
313	3	1	3	19	20	1.0526
314	3	1	4	21	23	1.0952
321	3	2	1	4	21	5.2500
322	3	2	2	29	30	1.0345
323	3	2	3	19	24	1.2632
324	3	2	4	19	22	1.1579
331	3	3	1	46	59	1.2826
332	3	3	2	53	60	1.1321
333	3	3	3	39	41	1.0513
334	3	3	4	26	27	1.0385
341	3	4	1	48	59	1.2292
342	3	4	2	28	46	1.6429
343	3	4	3	27	33	1.2222
344	3	4	4	9	11	1.2222
351	3	5	1	58	65	1.1207
352	3	5	2	43	47	1.0930
353	3	5	3	39	45	1.1538
354	3	5	4	39	42	1.0769
411	4	1	1	23	23	1.0000
412	4	1	2	28	40	1.4286
413	4	1	3	24	25	1.0417
414	4	1	4	17	18	1.0588
421	4	2	1	14	56	4.0000
422	4	2	2	49	50	1.0204

Class	Trip Duration	HH Income	Age	Self-Reported Vehicle Trips	Total Vehicle Trips (w/ GPS)	GPS Adjustment Weight
423	4	2	3	18	23	1.2778
424	4	2	4	12	14	1.1667
431	4	3	1	38	53	1.3947
432	4	3	2	61	72	1.1803
433	4	3	3	28	30	1.0714
434	4	3	4	16	16	1.0000
441	4	4	1	10	13	1.3000
442	4	4	2	44	54	1.2273
443	4	4	3	28	30	1.0714
444	4	4	4	7	9	1.2857
451	4	5	1	55	56	1.0182
452	4	5	2	52	55	1.0577
453	4	5	3	26	31	1.1923
454	4	5	4	43	45	1.0465
Total	**---**	**---**	**---**	**2265**	**2844**	**1.2556**

7.3 APPLICATION OF GPS ADJUSTMENT WEIGHTS TO ENTIRE REGIONAL TRAVEL SURVEY SAMPLE

All computational work up to this point was relegated to the SCAG GPS Study sample only. After the calculation of the GPS adjustment weights, the computational work was expanded to incorporate all households in the Post-Census Regional Travel Survey. All vehicle trip records by "class" (class is defined by variables presented in Tables 12 and 13) were adjusted by the GPS adjustment weight. Table 14 presents the impact of using the GPS adjustment weight on survey estimates.

TABLE 14:
COMPARISON OF TRIP RATES WITH AND WITHOUT GPS ADJUSTMENT

TRIP DATA (LINKED TRIPS)	WITHOUT GPS ADJUSTMENT	WITH GPS ADJUSTMENT
Total Households	5,386,491	5,386,491
Total Persons	15,904,849	15,904,849
Mean Total Trips Per HH	7.66	9.76
Motorized Trips per HH	6.70	8.80
Total Trips per Person	2.59	3.31
Driver Trips per Household	4.67	5.86
Total Trips	41,235,382	52,575,580
Vehicle Driver Trips	25,150,652	31,554,856

The unadjusted estimates were taken from Table 9: Year 2000 Post-Census Regional Travel Survey Key Trip Statistics, page 9 of the Final Report of Survey Results. Both statistics with and without GPS adjustment are for persons age 5+, and both have been weighted and expanded to represent total trips in the region, excluding external-to-external trips.

8.0 – CONCLUSION

The SCAG GPS Study provided useful information on the quantity and quality of missed driver and passenger vehicle trips in the self-reported data provided by respondents in the Post-Census Regional Travel Survey. The results of this study indicate that there are some households (roughly 10 percent) for which survey participants did not report an extreme number of trips. But the bulk of the households had a much lower percentage of trips that were not reported during the telephone interviews. However, this by itself does not imply that there is no under-reporting of trips by participants in the main part of the sample. Households participating in the GPS subsample were subjected to increased involvement in the survey due to additional telephone contacts, incentives, and the receipt of equipment through an overnight delivery service. This additional attention could have generated a feeling of greater participation and commitment among households participating in the GPS subsample, which may have caused these households to be more meticulous in terms of recording and reporting their trips during the telephone interview than households in the main portion of the survey that did not receive this additional attention. However, the results of this study do suggest that study participants are generally aware and can successfully interpret the research definition of a "trip" and can, when properly motivated, accurately report trips on a given travel day.

The SCAG GPS Study also introduces a methodology for correcting or adjusting vehicle trip records based on information derived from the GPS data. This approach is currently still experimental and more research in this area needs to be done. However, if the proposed approach (or an improved model) could become a standard method for adjusting and correcting survey estimates more accurate information on household travel behavior would be available for transportation planning and policy making.

There are a number of additional research questions that were beyond the scope of this project that could be examined through this data. The SCAG GPS Study provides an opportunity to investigate many different aspects of the trips taken by participants of the study. For example, the GPS information collected could be matched to routes, and this matched data could then be used to evaluate whether algorithms based upon shortest path accurately depict route choices by Southern California residents. In addition, since roughly one-half of all of the deployed devices provided travel information on multiple days, it may be possible to use the information collected to gain a better understanding of the temporal changes in travel within a household. In this regard, the collected data truly provide a wealth of information to be explored. For example, because the GPS equipment collected latitude and longitude information at one-second intervals, analyses could be conducted to examine items such as route choices, driving patterns/decisions, average time to stop, etc. Additionally a comparison of the actual driving distance could be compared to distance imputed using algorithms such as "shortest path."

APPENDIX A – DATA DICTIONARY

There are a number of data files associated with the GPS subsample. These files include the following:

- ***Raw Output:*** These files represent the raw output from the GPS equipment as it was directly uploaded from the equipment. There is one file for each device.
- ***GIS Files:*** These files represent the final files containing the latitude and longitude information, at one second or 30 second intervals, for all data collected by the GPS equipment. These data have been processed to remove extraneous information. There is one file for each device.
- ***Trip File:*** This file contains one record for each recorded person-trip. The beginning and ending latitude and longitude coordinates are included on each record, but the intermediate position coordinates are not. It contains trip information from every household providing GPS data.
- ***Matched File:*** This file contains one record for each person trip that was recorded by the GPS equipment or during the telephone interview. Only data from households with both GPS and interview data are included. Trips for each source have been matched and appear as a single record. Unmatched trips are also included.

The first three files are comma delimitated ASCII files. The Matched File is formatted as an Excel spreadsheet and SAS dataset. The following tables summarize the data elements in each set of files.

TABLE A-1: DATA ELEMENTS OF THE RAW OUTPUT FILES[1]

LINE IDENTIFIER (FIELD NO. 1)	DATA TYPES	FIELD NO.	DESCRIPTION	VALUE/FORMAT
$L Occupant Information	Character	2	Lookup Type	O = Occupant
	String	3	First name of driver/occupant	Maximum 30 characters.
	Number	4	Unique identifier for driver/occupant	
$H Header Information	Number	2	GPS device number	
	Number	3	Household ID number and vehicle ID number	Separated by hyphen
	String	4	PDA software version number	
	Number	5	Application type	
	Number	6	Micro version number	
$A Administration Information	Number	2	Date device was programmed	MMDDYY
	Number	3	Time device was programmed	HHMMSS
	Number	4	GPS power	0 = power-up, 1 = start trip
	Number	5	PDA power	0 = power-up, 1 = manual power
	Number	6	Differential GPS	0 = off, 1 = on
	Number	7	Trip source	0 = power, 1 = ignition, 2 = event port.
	Number	8	GPS sample rate	Seconds between samples
	Number	9	Analog sample rate	Seconds between samples
	Number	10	Idle speed	Kilometers per hour
	Number	11	Idle Sample rate	Seconds between samples
	Number	12	Idle time	Seconds
	Number	13	Timer	Minutes
$S Chain Start	Number	2	Date of chain start	MMDDYY
	Number	3	Time of chain start	Military time, HHMMSS
$E Chain End	Number	2	Date of chain end	MMDDYY
	Number	3	Time of chain end	Military time, HHMMSS.
$s Trip Start	Number	2	Date of trip start	MMDDYY
	Number	3	Time of trip start	Military time, HHMMSS
$e Trip End	Number	2	Date of trip end	MMDDYY
	Number	3	Time of trip end	Military time, HHMMSS
$+p Vehicle Power On	Number	2	Date of vehicle power on	MMDDYY
	Number	3	Time of vehicle power on	Military time, HHMMSS
$-p Vehicle Power Off	Number	2	Date of vehicle power off	MMDDYY
	Number	3	Time of vehicle power off	Military time, HHMMSS
$+i Vehicle Ignition On	Number	2	Date of vehicle ignition on	MMDDYY
	Number	3	Time of vehicle ignition on	Military time, HHMMSS
$-I Vehicle Ignition Off	Number	2	Date of vehicle ignition off	MMDDYY
	Number	3	Time of vehicle ignition off	Military time, HHMMSS
$O Occupant Trip Purpose	Character	2	Occupant type	D = driver, P = passenger
	String	3	Occupant code	
	String	4	Major trip purpose	
	String	5	Specific trip purpose	
$G Standard GPS Record	String	2	Date of GPS record	MMDDYY
	Number	3	Time of GPS record	Military time, HHMMSS
	Number	4	Latitude from GPS device	Decimal degrees
	Number	5	Longitude from GPS device	Decimal degrees
	Number	6	Speed	Kilometers per hour

1. Filename identifies GPS unit, household, and vehicle number in XXX_XXXXXXX_X format.

TABLE A-2:
DATA ELEMENTS OF THE GIS FILES[1]

FIELD NAME	DATA TYPE	FIELD NO.	DESCRIPTION	VALUE/FORMAT
GPS_ID	Number	1	Unique identifier for individual GPS record	
Latitude	Number	2	Latitude from GPS device	Decimal degrees
Longitude	Number	3	Longitude from GPS device	Decimal degrees
Speed	Number	4	Speed from GPS device	Miles per hour
Time	Number	5	Date and time stamp recorded by GPS device	MM/DD/YY HH:MM:SS
Flag	Number	6	Flag for GPS record	0 = No problem
				1 = No GPS fix
				2 = Bad latitude. Outside US
				3 = Bad Longitude. Outside US
				4 = Bad Speed
				5 = Bad Time, current time is before previous time
				6 = Bad distance
				8 = Data truncated due to edit of time
Chain_Seqno	Number	7	Trip chain sequence	
Trip_Seqno	Number	8	Trip number within trip chain	

1. Filename identifies household and vehicle number in XXXXXXX-X format.

TABLE A-3:
DATA ELEMENTS OF THE TRIP FILE

FIELD NAME	DATA TYPE	FIELD NO.	DESCRIPTION	VALUE/FORMAT
DeviceOwner	Number	1	Household number and vehicle number separated by hyphen	XXXXXXX-X
Chain_Seqno	Number	2	Trip number within trip chain	
Trip_Seqno	Number	3	Trip number within a trip chain	
Trip_StartDate	String	4	Start date of trip	MM/DD/YYYY
Trip_StartTime	String	5	Start time of the trip	HH:MM:SS AM/PM
Trip_EndDate	String	6	End date of trip	MM/DD/YYYY
Trip_EndTime	String	7	End time of the trip	HH:MM:SS AM/PM
Trip_PDATrip	String	8		True = Trip information was recorded using PDA False = Non-PDA trip
Trip_Distance	Number	9	Distance of trip	Miles
Trip_Duration	Number	10	Time duration of trip	Minutes
Trip_ImportTripSet	Number	11	Battelle data upload flag	
Occupant_Type	String	12	Occupant type	D = Driver P = Passenger
OccupantLU_SuppliedID	String	13	User-supplied occupant ID (Same as person number from recruitment file)	
MajorPurposeLU_SuppliedID	Number	14	User-supplied trip purpose ID (top level)	See Table A-3b
SpecificPurposeLU_SuppliedID	Number	15	User-supplied trip purpose ID (sub level)	See Table A-3b
GPS_Latitude	Number	16	Ending latitude from GPS device	Decimal degrees
GPS_Longitude	Number	17	Ending longitude from GPS device	Decimal degrees

TABLE A-3B:
TRIP PURPOSE CODES FOR TRIP FILE

MAJOR TRIP PURPOSE ID	SPECIFIC TRIP PURPOSE ID		
1=Change mode of transportation	1=Change mode of transportation	13=Medical
2=Pick-Up/Drop-Off	2=Pick up someone or get picked up	14=Fitness activity (playing sports; gym)
3=Personal	3=Drop off someone or get dropped off	15=Recreational (vacation; camping)
4=Work	4=ATM; buy gas; quick stop for coffee; newspaper	16=Entertainment (watching sports; movies; dance)
5=Education/Childcare	5=Shopping	17=Visit friends/relatives
6=Eat Meal	6=Banking; post office; pay bills	18=Community meetings/political/civic event
7=Medical	7=Work (regularly scheduled)	19=Occasional volunteer work
8=Recreation/Entertainment	8=Work-related (sales call; meeting)	20=Church; temple; religious meeting
9=Social/Civic/Religious	9=School (attending classes)	21=Go along for the ride
10=Go along for the ride	10=Other school activities (sports; extra curricular)	22=Other personal
11=Other	11=Childcare; day care; after school care	23=Other
12=Return Home	12=Eat Meal	24=Return Home

TABLE A-4:
DATA ELEMENTS OF MATCHED DATA FILE

FIELD NAME	DATA TYPE	DESCRIPTION	VALUE/FORMAT
GPS_ID	Number	Unique identifier for GPS trip record	
Diary_ID	Number	Unique identifier for diary trip record	
HHID	Number	Household ID number	
MatchType	Number	Battelle flag indicating reason for match	1 = Trips have exact match (rounded to the minute) for both start and end time
			2 = Trips have nearly exact match (± 10 minutes) for start and end time
			3 = Start and end times within in 1 hour, ending longitude and latitude within ±1000 feet of each other
			4 = Trips have exact match (rounded to the minute) for start time and identical trip purposes
			5 = Trips have nearly exact match (± 5 minutes) for start time for start and end time and identical trip purpose
			6 = Manual match
			-1 = No match
GPSPerno	Number	Person ID number	
GPSMode	Character	Description of occupant	D = Driver
			P = Passenger
GPSTrip	Number	Chain and trip number	(Chain #)-(Trip #)
GPSVeh	Number	Vehicle ID number	
GPSStart	Date/Time	GPS-recorded start date and time	MMDDYY HH:MM:SS (Military time)
GPSEnd	Date/Time	GPS-recorded end date and time	MMDDYY HHMMSS (Military time)
GPSDur	Number	GPS-recorded duration of trip	Minutes
GPSPurpose	Number	User-supplied ID number of trip purpose	
GPSEndLat	Number	GPS-recorded latitude at end of trip	Decimal degrees
GPSEndLong	Number	GPS-recorded longitude at end of trip	Decimal degrees
GPSDist	Number	GPS-recorded distance	Miles
DiaryPerno	Number	Diary-recorded Person ID number	
DiaryMode	Number	Description of occupant	D = Driver
			P = Passenger
DiaryTrip	Number	Trip number	
DiaryPurp	Number	Diary-recorded trip purpose	
DiaryVeh	Number	Diary-recorded vehicle ID number	
DiaryStart	Date/Time	Diary-recorded start date and time	MMDDYY HH:MM:SS (Military time)
DiaryEnd	Date/Time	Diary-recorded end date and time	MMDDYY HHMMSS (Military time)
DiaryDur	Number	Diary-recorded duration of trip	Minutes
DiaryLat	Number	Diary geo-coded latitude at end of trip	Decimal degrees
DiaryLong	Number	Diary geo-coded recorded longitude at end of trip	Decimal degrees
StartTime	Date / Time	Start time associate with trip	Set to GPS start time for a GPS/Diary match. Set to diary start time for an unmatched diary entry.
Non_hh Member	Number	Occupant was non-household member	0 = no
			1 = yes
GPS_not_Diary	Number	GPS trip without matching diary trip	0 = no
			1 = yes
Diary_not_GPS	Number		1 = Trip not possible
			2 = Trip not in household vehicle
			3 = GPS equipment not installed
			4 = Person not identified by GPS equipment
			5 = GPS equipment not installed in this household vehicle
			6 = Unit malfunction
Exclude_HH	Number	Exclude household	0 = no
			1 = yes