Australasian Transport Research Forum 2010 Proceedings
29 September – 1 October 2010, Canberra, Australia
Publication website: http://www.patrec.org/atrf.aspx

CAN GPS REPLACE CONVENTIONAL TRAVEL SURVEYS? SOME FINDINGS

Peter Stopher, Christine Prasad and Jun Zhang

Institute of Transport and Logistics Studies, The University of Sydney

Email for correspondence: Peter.Stopher@sydney.edu.au

Abstract

Global Positioning System (GPS) devices have been considered as a substitute for conventional travel diaries for some time. However, only in the past year has a serious effort been made to trial replacement of travel diaries with GPS. GPS devices are very accurate at recording the time and positional characteristics of travel, but cannot record travel modes, trip purposes, or number of occupants in private vehicles, all of which are important attributes normally acquired in a household travel survey. The authors of this paper have developed software that is able to deduce the missing data from a combination of the GPS records, other data collected from respondents, and data available in GIS records. However, the level of accuracy has not been checked. As part of an ongoing GPS-only survey, a prompted recall survey is being conducted on a sample of households. In this paper, we report on comparisons between the results from the processing software and the prompted recall web survey, with respect to identifying trips, modes and purposes of travel, and number of household members travelling together. These results will be used subsequently to develop refinements to the processing software.

1. Introduction

In the mid-1990s, Global Positioning System (GPS) devices were first trialled (Wagner, 1997) as a means to measure people’s travel, as a direct outcome of a Conference held in Irvine, California on Household Travel Surveys (TRB, 1996). At that time, selective availability (intentional degradation of the GPS signal by the US government – NOAA, 2010) was still in place, GPS technology was in its infancy, and devices were cumbersome and required an external power source. Over a little more than a decade, selective availability has been turned off and the technology has improved enormously, as summarised by Wolf (2009). Since the outset of GPS use, the idea in the minds of the profession has been that one day GPS might replace the conventional interview or self-administered household travel survey (Wolf et al., 2001). However, in the early years of the development of GPS surveys, it was clear that neither the technology nor the processing software was yet ready for such a replacement to take place. Rather, most of the use of GPS was to validate travel surveys (Wolf et al., 2003; Stopher, 2009) and for evaluation of travel behaviour changes aimed at reducing daily vehicle kilometres of travel (VKT) (Stopher, 2009). However, these uses of GPS served both to provide the opportunity to improve and change the design of the devices and also to develop increasingly sophisticated processing software (Stopher et al., 2008).

With these developments in mind, the Ohio Department of Transportation (ODOT) commissioned the first GPS-only Household Travel Study, which is taking place at the time of writing this paper in the Greater Cincinnati Area of the Ohio-Kentucky-Indiana (OKI) metropolitan planning organisation region. This study was commissioned in early 2009, with a pilot survey to be conducted in April of that year and with the expectation of the main survey being conducted over a 12-month period from about August 2009 to August 2010. It is expected that at least 3,600 households will eventually be included in the sample, with all households using GPS devices for a period of two to three days. GPS devices are given to all members of a household aged 12 years and over, and brief diary surveys are used for those members of the household under the age of 12. In addition, a subsample of about 1,000 households is hoped to undertake a Prompted Recall (PR) survey (details of which are provided in the next section of this paper). The GPS survey has two purposes. The first purpose is to provide an opportunity to upgrade and improve the processing software, so that more complete and more accurate data can be obtained from the GPS records. The second purpose is to provide a database that can support the continued updating and improvement of travel demand models for the OKI region, and to do this with processed GPS data for the first time.

In the pilot survey, it was intended to recruit 250 households that would use the GPS devices, and to recruit 100 of these households to undertake the PR survey. As a result of lower–than-expected response rates to the Prompted Recall Survey, the pilot survey resulted in a sample of 120 households and 228 persons who provided GPS data. More details of the pilot survey are provided in Stopher and Wargelin (2010). It was decided to recruit all pilot survey households to undertake the PR survey. This resulted in 35 households providing usable PR data from a total of 46 individuals. While this appears to be a low response rate, it must be kept in mind that the PR survey is a web-based survey. General response rates to web-based surveys are reported to be in the 10-25 percent range in the US (Jones and Pitt, 1999; McDonald and Adam, 2003). The almost 30% response rate achieved in this survey is therefore considered high for a web-based survey. The PR survey provided data for one day of travel for each person who completed it. As a result of the pilot survey, changes were made to the PR survey in particular, and some selective incentives were introduced to try to increase the response to the PR survey in the main survey. Also, due to the rather low response rate to the PR survey in the pilot, it was decided in the main survey to recruit all households that provided GPS data to undertake the PR survey. It should be noted, however, that the overall response to the GPS survey was similar to or higher than that usually experienced with conventional surveys (Stopher and Wargelin, 2010).

This paper reports on analysis conducted on the first few months of the main survey, from August 2009 until March 2010, representing about one-third of the eventual aimed for PR data. At this point, PR data had been received from 214 persons, representing 142 different households.

2. The Prompted Recall (PR) Survey

The PR survey first appeared very early in the development of GPS applications in transport (Bachu et al., 2001) and was then developed further in a number of subsequent studies. Stopher et al. (2002) used a small pilot study to investigate the concept further, subsequently transitioning the survey from a paper PR to a web version (Stopher and Collins, 2005). This was followed by a number of further developments in Internet-based PR surveys in the next few years (Lee-Gosselin et al., 2006; Li and Shalaby, 2008, Auld et al., 2009). Auld et al. (2009) provide a more detailed history of the development of PR surveys over the past eight years or so.

The concept of the PR survey is that respondents who have earlier carried a GPS device with them for a day or more are subsequently sent information that allows display of the travel recorded on the GPS device. They are then asked to provide additional information about the travel, such as the mode of travel, the purpose of travel and the size of their travel party, as well as to indicate if there are any errors in the processed GPS data. The maps that display the travel recorded by the GPS device therefore act as a memory prompt to the individual, then allowing the individual to respond to questions about the travel. In the earliest form of the PR survey, maps of each day of travel undertaken by the respondent were printed and incorporated within a paper survey that then asked for further information about the travel and also offered the respondent the opportunity to indicate if there were errors in the processing or if there were gaps in the GPS record. In general, however, the paper survey was rather clumsy, in that the respondent could generally indicate only limited information about the displayed trips and correction of the processing. Indicating that a mapped stop was not a stop, or that a stop had been omitted at a certain location, or that entire travel had been omitted was generally difficult to accommodate in a paper format.

Thus, the transition of the PR survey from paper to the Internet, providing an interactive environment in which respondents could indicate corrections to the GPS processed record, was extremely important to the continuing use of the PR survey. There remain two problems associated with the PR survey, however. The first is that the survey requires that respondents are familiar with maps and map reading, and have the ability to understand the implications of a series of trips shown on a map. Second, the ability to read a map may require an alternative dimension of literacy than is often required for standard paper and pencil surveys. Second, the survey requires access to and familiarity with the Internet. This necessarily reduces the proportion of households and household members who could respond to a PR survey administered over the Internet.

At the outset of the Greater Cincinnati Area Household Travel Survey (GCAHTS), it was proposed to conduct PR surveys by both paper and pencil and the Internet. However, as the specification of the survey was developed in an Internet environment, it rapidly became clear that a comparable survey by paper and pencil could not be developed within reasonable resource constraints. The decision was made, therefore, that the PR survey would be conducted only by Internet. While this could be considered to generate some bias in the responses, there is, in fact, no need for the PR survey to be undertaken by a representative sample of the population, because the purposes of the survey are not to expand the PR results to the entire population of the region.

In the case of the GCAHTS, the purposes of the PR survey are to provide “ground truth” about the travel undertaken by a subsample of people, against which to check the results of the processing of GPS data, and also to provide a data source for potentially improving the processing software. Neither of these uses demands a representative sample. There is no question that a representative sample would be nice to have, but it is not a requirement for the use of the data. At the same time, the fact that respondents to the PR survey are both Internet and map savvy is unlikely to have any biasing effect on either of the uses of the PR data.

Whilst the decision was made to limit the PR survey to an Internet version, it was also decided to develop the PR survey by using Google® Maps, so that respondents would be likely to find some similarity between the PR survey and maps that they may possibly be familiar with in their own use of the Internet. It is necessary for processing to be undertaken on the GPS data, prior to creating the maps for the PR survey. The procedures for data processing and analysis and creation of the PR survey are described in the next section of this paper.

3. Data Processing and Analysis and Generation of the PR Survey

In this survey, the GPS devices used are the GPS-PPAL device of the Institute of Transport and Logistics Studies (ITLS). The device is shown in When the survey period is completed by a household, the devices (which are logging devices) are returned to the survey team who download the data. What is obtained is a modified stream of data from the GPS device, giving the second-by-second position of the device for the two or three days during which the sampled respondent had the device. These data are then processed by a series of software programs developed at ITLS. The first of these programs uses a number of rules to delete spurious data (generally the data collected whilst the device is at rest at the end of a trip, or possibly in the middle of a trip when there is a lengthy delay in movement, such as may occur at a traffic signal) and to split the data stream up into what are assumed to be individual trips. At the completion of this process, maps are generated by the software, along with a summary file showing the assumed start and end locations of each trip, the time (to the nearest second) when the trip started and ended, and some of the other characteristics of the trip (distance, elapsed time, average speed, etc.).

Figure 1. Each person 12 years of age or older in each sampled household is asked to carry one of these devices for about three days. Each device is identified by a unique number and each participant is required to carry the same device for the duration of the GPS survey. The device is set to record position every second and is equipped with a vibration sensor. If no vibration is sensed for 3 minutes, the device turns itself off. As soon as the device is vibrated again, it turns on and seeks a position. If the time it has been off is less than about an hour, then the position is usually acquired within a matter of 10 to 15 seconds. However, if the device has been off for more than an hour, position acquisition may take from 10 or 15 seconds up to about a minute or so, depending on the speed of movement and location of the device.

When the survey period is completed by a household, the devices (which are logging devices) are returned to the survey team who download the data. What is obtained is a modified stream of data from the GPS device, giving the second-by-second position of the device for the two or three days during which the sampled respondent had the device. These data are then processed by a series of software programs developed at ITLS. The first of these programs uses a number of rules to delete spurious data (generally the data collected whilst the device is at rest at the end of a trip, or possibly in the middle of a trip when there is a lengthy delay in movement, such as may occur at a traffic signal) and to split the data stream up into what are assumed to be individual trips. At the completion of this process, maps are generated by the software, along with a summary file showing the assumed start and end locations of each trip, the time (to the nearest second) when the trip started and ended, and some of the other characteristics of the trip (distance, elapsed time, average speed, etc.).

Figure 1: The GPS-PPAL Device

Because it is not possible to craft rules that will work 100 percent of the time, the next step in the process is what ITLS calls ‘map editing’. In this process, trained staff at ITLS review the maps for each day using TransCAD® software and look for possible spurious data that may not have been deleted in the initial processing, for possible stops in a trip that the software identified as a single trip, and trips that might be split into two or that may be missing due to loss of GPS signal. Trips may not be split correctly by the software due to a rule that dictates that an identifiable stop (after removing spurious data) must last for at least 120 seconds to define the end of a trip. Because there are a number of activities that will take less than 120 seconds to accomplish that should also define the end of a trip (such as picking up or dropping off a person), and also because the deletion of spurious data is also done conservatively by the software, it is necessary to inspect the map and make some edits to the list of trips provided by the software processing. The map editing process, in this case, would have removed the agglomeration of data points at each of the three locations on the left and bottom of the map, would have inserted stops at those locations, and would also have created a stop at the end of the trip near the top of the map. Edits of this type are needed so that the respondent is not expected to understand how such agglomerations of data points occur and to provide a map that is more clearly representative of the travel undertaken. All editing of trips is done to a trip list in text format so the original visual map that was generated will remain unchanged.

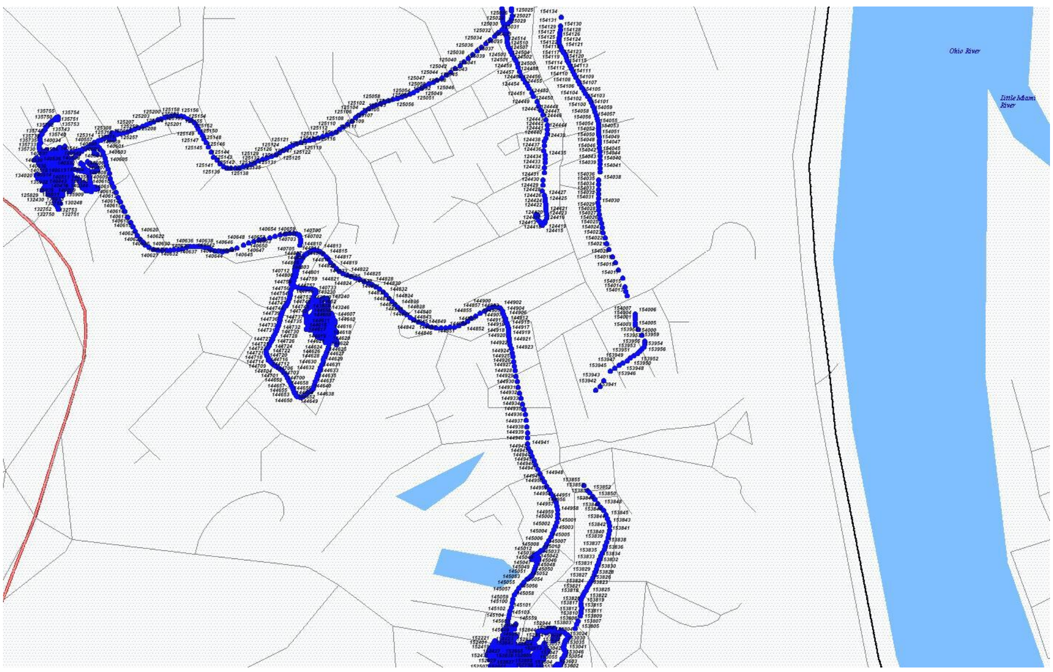
After map editing is complete, the data are then run through several processes prior to producing the data for the PR survey. One of the processes applies the changes from the amended trip list file to the original trip database to remove data points, and split or join trips. Another process compares the address information collected from respondents to the locations of trip ends in the modified trip list and records any matches for input into the PR survey, so that home, work, educational establishment, or grocery shopping locations can be shown on the map and the possible purpose of the trip can be shown.

Figure 2 shows part of what the processing software identified as a single trip. However, it can be seen that there are some clusters of points at three locations and another location where the respondent appears to have travelled to a point and then returned, without a stop of any noticeable duration. These all suggest the possibility that the trip should be broken into a number of different trips.

The map editing process, in this case, would have removed the agglomeration of data points at each of the three locations on the left and bottom of the map, would have inserted stops at those locations, and would also have created a stop at the end of the trip near the top of the map. Edits of this type are needed so that the respondent is not expected to understand how such agglomerations of data points occur and to provide a map that is more clearly representative of the travel undertaken. All editing of trips is done to a trip list in text format so the original visual map that was generated will remain unchanged.

After map editing is complete, the data are then run through several processes prior to producing the data for the PR survey. One of the processes applies the changes from the amended trip list file to the original trip database to remove data points, and split or join trips. Another process compares the address information collected from respondents to the locations of trip ends in the modified trip list and records any matches for input into the PR survey, so that home, work, educational establishment, or grocery shopping locations can be shown on the map and the possible purpose of the trip can be shown.

Figure 2: Part of a 'Single' Trip Identified by the Software



Not all of the steps to create the web link or Uniform Resource Locator (URL) for the web survey are explained here, because some require a detailed knowledge of other aspects of the survey and would be excessive for this paper. It is sufficient to understand the broad process used to generate the web survey, as described here. Figure 3 shows an example of the PR survey as it is used in the main survey, displaying at the outset to the respondent the full day's worth of travel that was recorded by the GPS. This is intended to orientate them and to show the overall task that the respondent is to undertake as part of the PR survey. The survey then proceeds by displaying to the respondent one trip at a time as shown in Figure 4.

With each trip displayed, the respondent is given a number of options. In the pilot survey, respondents were allowed to amend the time that a trip started, the time it ended, the trip's distance and speed, as well as to fill in the purpose of the trip and the mode of travel used. Respondents could also indicate if a stop had not occurred where it was shown (i.e., combining two or more trips into one), or if a stop had been made that was not shown on the map (i.e., splitting a trip into two or more separate trips). These options were reduced for the main survey.

In the questions shown on Figure 4, when the respondent clicks on a red question mark, a window pops up with a list of available responses to the questions about the activity, means of travel, and accompanying household members. The numbers across the top of the screen in Figure 4 show the total number of trips that are in the day's travel. These turn to red as the respondent opens each one, indicating that there is information required to be entered. If all information has been entered, but not everything confirmed, then the number turns orange. Finally, when all is entered, the circle changes to green. Once all the circles are green, the respondent can return to review and edit any one of the trips. There are also zoom capabilities that allow the respondent to zoom in on a trip end or some location along the trip, if he or she thinks that a stop occurred somewhere along the trip.

Figure 3: Overview of Travel from the PR Survey

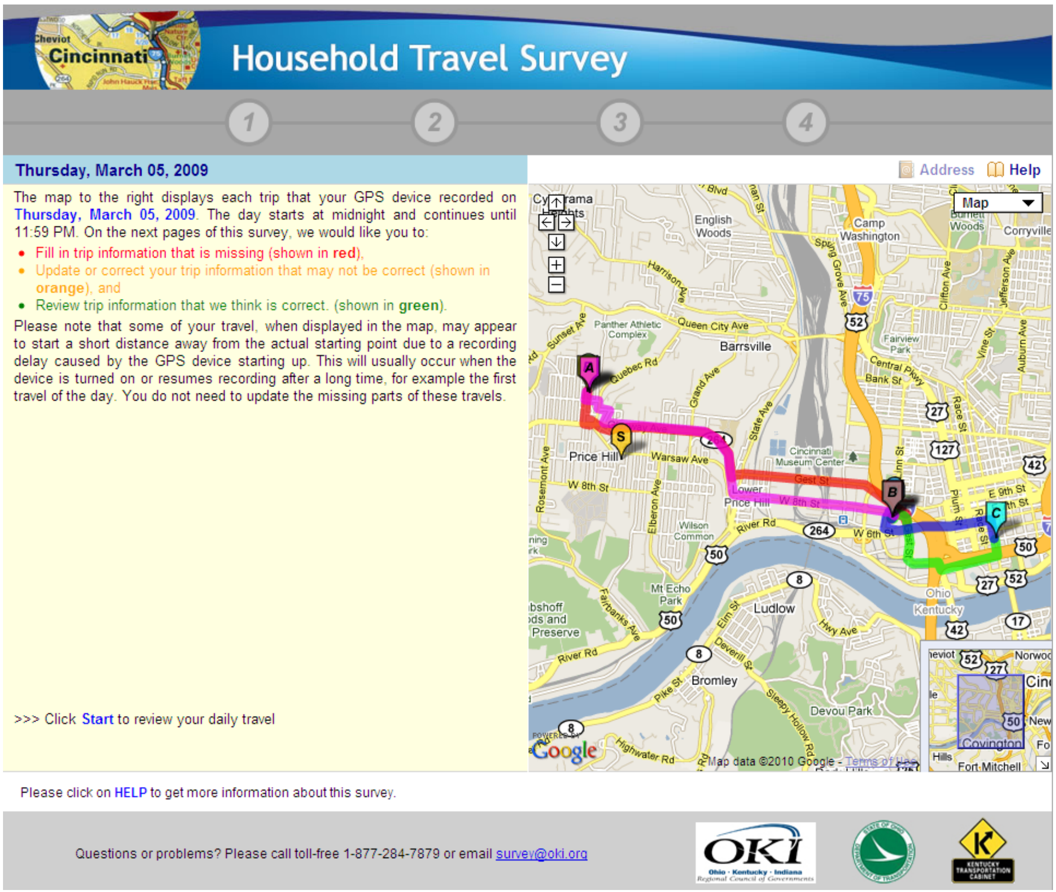

Household Travel Survey
Cincinnati
1
2
3
4
Thursday, March 05, 2009
Address
Help
The map to the right displays each trip that your GPS device recorded on Thursday, March 05, 2009. The day starts at midnight and continues until 11:59 PM. On the next pages of this survey, we would like you to:
Fill in trip information that is missing (shown in red),
Update or correct your trip information that may not be correct (shown in orange), and
Review trip information that we think is correct. (shown in green).
Please note that some of your travel, when displayed in the map, may appear to start a short distance away from the actual starting point due to a recording delay caused by the GPS device starting up. This will usually occur when the device is turned on or resumes recording after a long time, for example the first travel of the day. You do not need to update the missing parts of these travels.
>>> Click Start to review your daily travel
Map
Price Hill
Barrsville
Ludlow
Bromley
Devou Park
Covington
Map data ©2010 Google
Please click on HELP to get more information about this survey.
Questions or problems? Please call toll-free 1-877-284-7879 or email survey@oki.org
OKI

Figure 5 shows a zoomed-in version of the trip that ended at B. Finally, Figure 6 shows how the display appears when all editing has been completed.

Figure 4: Details of One Trip on the PR Survey

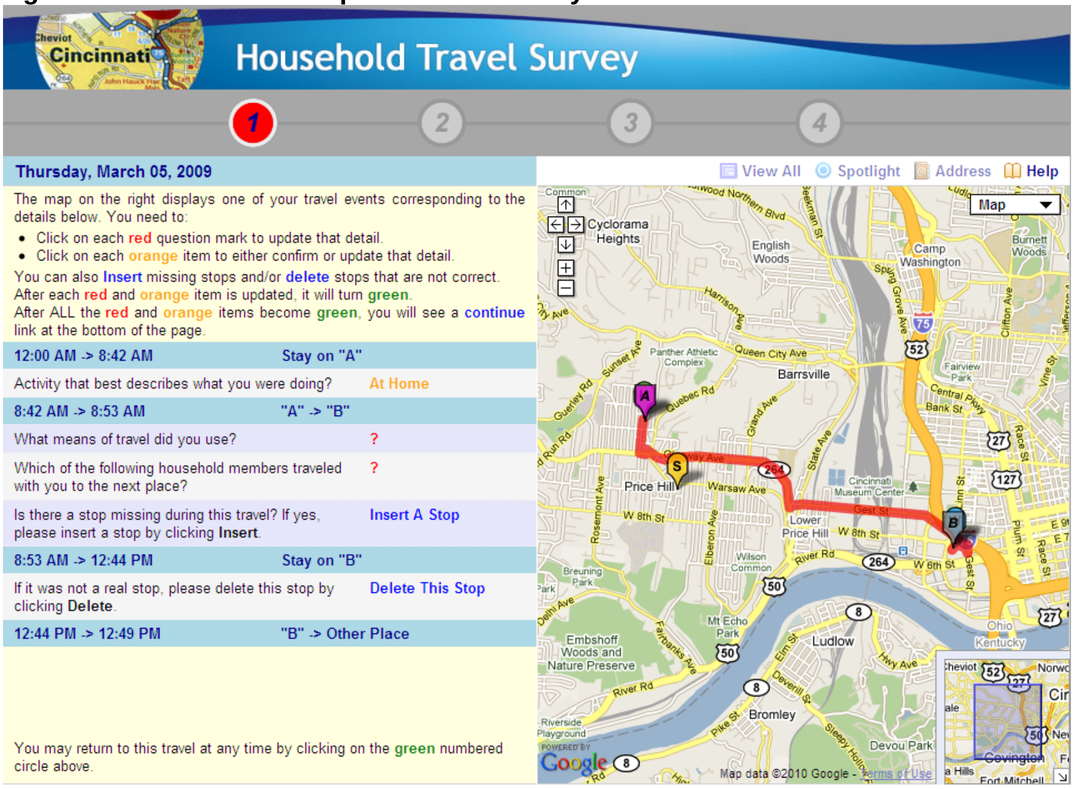


Figure 5: Zoomed-In Map for Trip End at B

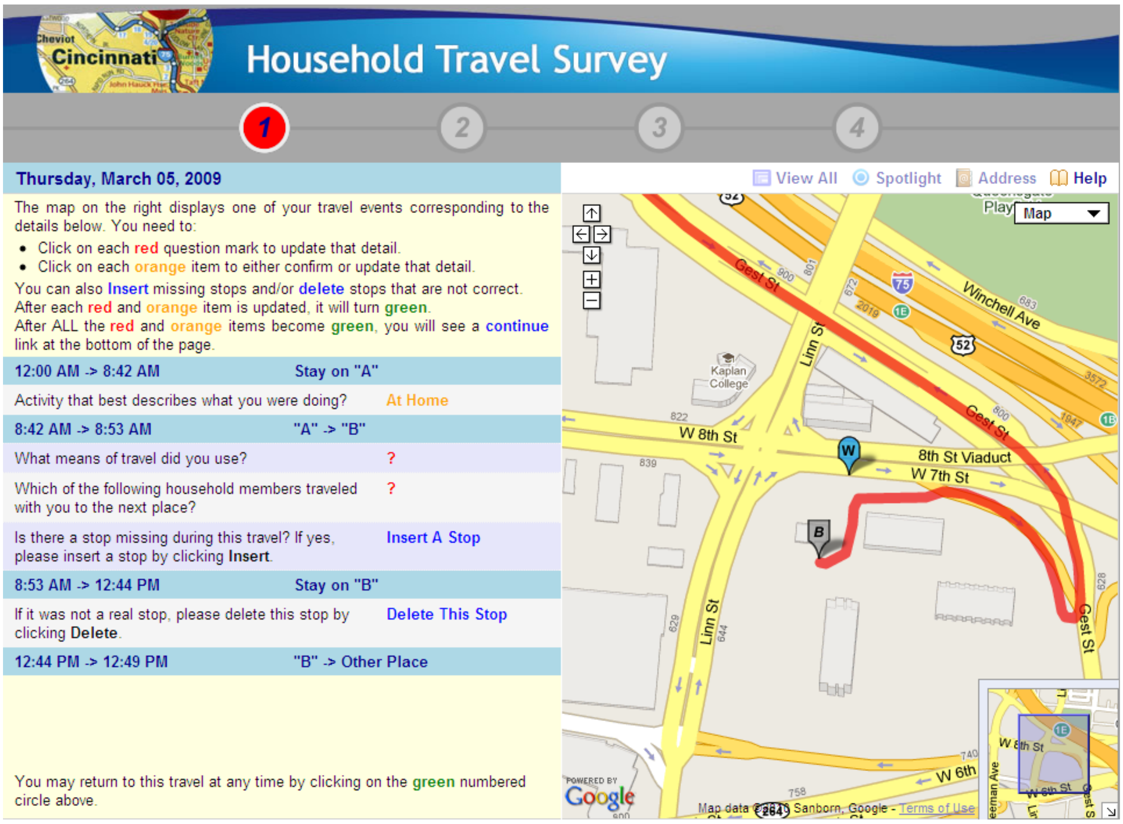
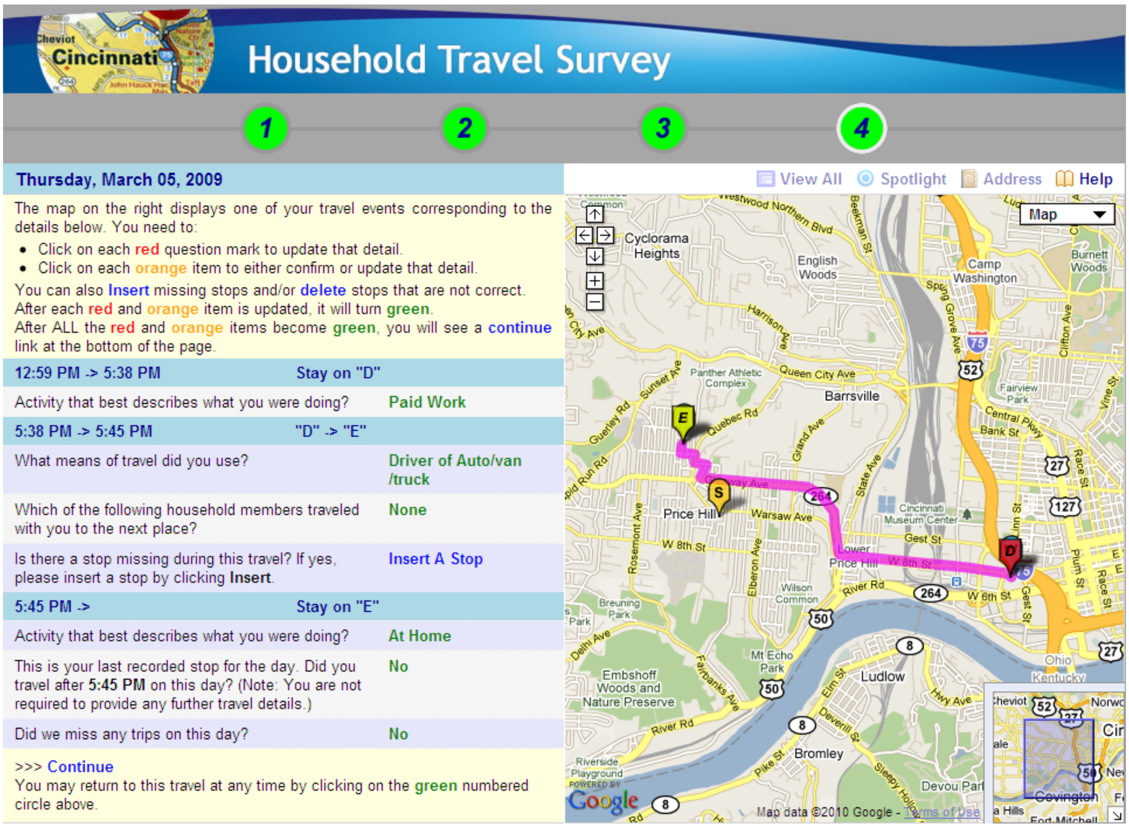


Figure 6: Completed PR Survey



Once the respondent completes the editing of the data, a final screen appears in which the respondent may make any comments or ask any questions. The survey then concludes and the data are saved to the server. It should be noted that the survey does not allow the respondent to change the times at which trips start and end, unless the respondent is deleting a stop or inserting a new stop. We permitted this in the pilot survey and found that respondents often disagreed with the GPS times, sometimes as little as rounding off the time to the nearest 5 minutes, and sometimes disagreeing by hours. However, the GPS does not lie, so we decided to remove this option, which had seriously complicated the subsequent analysis, by removing data on which a match can be made.

4. Analysis of the Prompted Recall and GPS Records

This paper provides preliminary findings from those who completed the PR survey prior to the beginning of April. For the most part, these respondents were reporting on travel that took place in August through November. The analysis focuses first on the identification of trips. Respondents were not able to change the times that trips started and ended, but could split a trip into more than one trip or delete a stop so as to join two trips into one. It should be noted that the software splits multi-modal trips into separate trips for each mode, and that trips made during work are included, if the respondent carried the GPS device with them during the working day. Second, the analysis focuses on mode of travel. The software has limited ability to identify modes of travel, being restricted to private motorised vehicle, bus, walk, bicycle, and other. Respondents were asked, however, to choose from a list of fifteen travel modes that included distinction between driver and passenger for various motorised modes, including (separately) privately owned cars, vans, or trucks, carpools, vanpools, motorcycle or moped. Third, the analysis focuses on purpose. Here again, the software capabilities are quite restricted in the present software, being limited to home, school, work, shop, and other. Respondents to the PR survey were asked to indicate their trip purposes from a list of 15 activities.

There are two issues to be examined here. First, it is important to determine how well the current processing software and manual editing of the GPS data are doing in relation to providing accurate representations of the trip-making behaviour of respondents. Second, however, it is also important to try to gauge how well respondents to the PR survey respond by providing accurate information. It must also be remembered in assessing this that, by having the PR information, respondents are in a far better position to provide accurate information about their travel than in a conventional travel survey. Therefore, to whatever extent there may be evidence of errors in respondent information, it must be assumed that these errors would have been much greater in a conventional household travel survey.

4.1. Trip Analysis

By the end of March 2010, 214 individuals had provided PR data for one day, for which there was also a GPS record, covering at least 1,200 trips[1]. Of those 214 respondents, 133 (62%) agreed with the number of trips identified by the GPS processing software. A further 11 individuals reported that there was one or more trips missed at the end of the day, while everything else agreed between their recollection and the GPS record. Thus, in total, there were 144 individuals (67%) who agreed with the GPS software, apart from a possible omission at the end of the day. Trip omissions are likely due to the device's battery running out at the end of the day or respondents forgetting to carry their GPS devices for trips made after they return home, for example, when making a pickup trip after a day at work. A further 34 individuals disagreed by only one trip, with most of these being cases where the GPS identified one more trip or stop than the respondent. Therefore, 178 respondents (83%) agreed within plus or minus one trip with the software results. The remaining 36 disagreed by more than one trip. In no case did the GPS software fail to identify more than 8 trips

[1] This number is explained and discussed further later.

different from the individual, but there were 13 instances where the difference between the respondent and the GPS was either 3, 4, 5, or 8 trips, in all cases the majority of these being additional trips identified by the GPS than the respondent. The one respondent who disagreed on 8 trips combined 7 trips into one trip lasting 4 hours and 48 minutes. In total, the GPS software failed to identify 8 trips reported by individuals in the PR. However, only one of those 8 trips lasted for more than 1 minute, so that the edit by the respondent appears to be incorrect. On the other hand, individuals reported that 81 trips identified by the GPS software were not trips, or did not involve stops for other than traffic purposes. In 18 of these cases, the respondent combined together GPS trips to produce a total trip of several hours duration, which appears to have defined a round trip or a tour rather than a one-way trip. In another 38 cases, the GPS identified a stop of longer than 2 minutes duration which was generally not consistent with a traffic stop, so that these were most likely cases where the respondent forgot a stop or considered it was not relevant. In 11 cases, the respondent deleted what appeared to be a significant trip both in time and distance, and provided no means for the respondent to have moved to the next trip starting point. In 14 cases, it appeared that map editing had resulted in splitting a trip that should not have been split. These last are the only cases where it appears that the GPS data were potentially in error and represents an error rate of just 1.2 percent of the total trips.

There are a total of at least 1,200 trips in the records from these respondents. The reason that it is stated as 'at least' is that there are 20 cases where the respondent indicated that one or more trips were missing at the end of the day. Because we do not know if this was one trip or several, we cannot be sure of the exact number of trips in the data set. In looking at the agreement between the GPS record and the PR responses, we find that the 133 respondents who agreed completely with the GPS record made 649 trips, or a trip rate of 4.88 trips per day. For the 11 respondents who agreed with all but one or more missing trips at the end of the day[2], there were 71 trips on which they agreed, plus 11 or more trips that they believed should have been included in the record. This brings the total of trips in agreement to 720 for 144 respondents. Because the trips missed at the end of the day were most probably because the respondent forgot to take the GPS device with him or her on the last trip(s) of the day, this represents a very modest respondent error in the survey. Also, adding in these additional respondents, we reach a daily trip rate of 5.0, not including the missing trips at the end of the day.

For the respondents who disagreed on one or more of the identified trips on the PR survey, there were 480 trips reported either by the GPS or the respondent. (Just 8 of these 480 trips were ones that the respondent identified as missing.) These 480 trips were made by 70 respondents, representing a trip rate of 6.86 trips per person per day. However, of those 480 trips, it appears that 14 trips were incorrectly identified by the software and map editing procedure, reducing the trip total to 466 trips, or 6.66 trips per person per day.

4.1. Mode Analysis

There are 1,015 trips in the data file where a comparison can be made between the software result and the answers provided by respondents to the PR Survey. Table 1 provides a comparison between the results of GPS processing for mode and responses obtained from respondents on trips that matched. As can be seen, on an overall basis, the processing software is performing well. Summing drivers and passengers in private vehicles shows 885 private vehicle trips in the PR data, compared to 873 in the GPS processing. There are 12 bus trips in the PR data, and 8 were identified in the GPS processing, while there are 15 bicycle trips identified by GPS processing and none in the PR survey responses.

[2] Respondents may have returned home, say from work, and then gone out again in the evening, but forgot to take their GPS device with them. This would lead to omitted trips at the end of the day.

Table 1: Comparison of Mode of Travel between PR and GPS Software

Mode of Travel	Prompted Recall	GPS Processing
1. Motor Vehicle (GPS)	N/A	873
2. Bus (GPS)	N/A	8
3. Walk	85	105
4. Bicycle	0	15
5. Driver of Auto/Van/Truck (PR)	820	N/A
6. Passenger of Auto/Van/Truck/Motorcycle (PR)	63	N/A
7. Driver of Carpool (PR)	0	N/A
8. Passenger of Carpool (PR)	1	N/A
9. Driver of Vanpool (PR)	0	N/A
10. Passenger of Vanpool (PR)	1	N/A
11. Bus (Public Transport) (PR)	6	N/A
12. Demand Response Bus (PR)	0	N/A
13. School Bus (PR)	6	N/A
14. Taxi (PR)	0	N/A
15. Motorcycle/Moped (PR)	6	N/A
96. Other	15	8
98. Unknown	13	6
TOTAL	1015	1015

It is useful to look at the extent to which mode was classified correctly by the software, compared to what respondents reported in the PR survey. Table 2 shows this information for those modes that are reported in each of the two sources.

From Table 2, it is apparent that public transport (bus) trips were misidentified with most being classed by the processing software as car, while the trips identified as bus by the software were actually trips where the respondent was driving a car/van/truck. A significant number of walk trips were incorrectly identified as car and vice-versa. All of the trips identified as bicycle were either walk or car, apart from one school bus trip. In total, the number of trips that were correctly identified were 916 out of 1068, or about 86 percent. This is lower than our previous experience, and may be a result of some issues with incomplete GIS layers for bus routes and bus stops, and to developments in the mode identification software program to incorporate identification of school bus mode. However, further work is needed to distinguish among car, walk, and bicycle, as well as improving the identification of bus trips.

Table 2: Detailed Comparison of Mode Identification between GPS and PR

PR identified modes	GPS identified modes						
Mode	Car	Bus	Walk	Bicycle	Other	Unknown	**TOTAL**
Walk	21	0	59	3	0	2	**85**
Driver of Auto/Van/Truck	753	7	37	11	8	4	**820**
Passenger of Auto/Van/Truck/Motorcycle	62	0	1	0	0	0	**63**
Passenger of Carpool	0	0	1	0	0	0	**1**
Passenger of Vanpool	0	0	0	0	0	0	**0**
Bus (Public Transport)	5	0	1	0	0	0	**6**
School Bus	4	0	1	1	0	0	**6**
Motorcycle/Moped	5	1	0	0	0	0	**6**
Other	10	0	5	0	0	0	**15**
Unknown	13	0	0	0	0	0	**13**
TOTAL	**873**	**8**	**105**	**15**	**8**	**6**	**1015**

4.2. Activity Analysis

Similar to the situation with mode where the software has a subset of possible options, at present the software is able to identify home, work, school, social-recreational, shop, and other as the activities at the origin and the destination end of the trip. Again, there were 1,015 trips for which a comparison can be made between what the respondent reported in the PR and what the GPS software identified as the activity. In this case, for the origin activity, the software classified only 435 of the 1015 origins correctly and, for the destination activity, it classified 431 out of 1014 destinations correctly. This represents correct identification of only 43 percent of activities. Both of these we consider to be unacceptably low, and they clearly indicate a need for major improvements to the software. Overall, Table 3 shows the outcome of the activity identification.

Table 3: Comparison of Activities between PR and GPS Software

Source	Home	Work	School	Social Rec	Shop	Other	Total
PR-Origin	277	163	36	51	142	346	1015
GPS-Origin	244	90	12	174	58	437	1015
PR-Destination	297	155	29	53	136	344	1014
GPS-Destination	243	87	13	176	57	438	1014

Within the comparisons shown in Table 3, there are a number of mismatches, where the Table might seem to imply a match. For example 79 percent of home activities (191/244 and 195/243) are correctly identified, while 74 percent of work activities (66/90 and 63/87), only 15 percent of school activities (2/12 and 1/13), and 36 percent of shopping activities (21/58 and 20/57) are correctly identified. School is most frequently confused with work. The mismatch on shopping is less hard to understand, because the information provided is only of the two most frequently used grocery shopping locations, whereas there may be many trips to other shopping locations. Social-recreational purposes are identified only as those involving multiple occupants, which is only a partial definition. Hence a good match on this purpose was not expected. Because the software does not yet use land use information, nor duration of stay at the location, substantial improvements are anticipated for refinements to the software that takes into account such information.

Improvements to mode and purpose are the next steps that will be taken, now that we are accumulating a significant number of responses to the PR survey. This will be undertaken by analysing the misidentification of each of mode and purpose, and also using improved information on bus routes and bus stops, school bus related information, and land use information. A detailed representation of the matching and mismatching of the PR Survey and the GPS software is shown in Table 4 and Table 5. These two tables will be the basis of further work on how to refine the software. We will examine the cases of mismatch that occur, which are highlighted in the two tables. For origins, there is a total of 464 trip origins that should have been possible to identify correctly, but which were not so identified. Similarly, there are 456 trip destinations that should have been possible to identify correctly, but which were incorrectly identified. Correction of these alone would raise the correct identification to around 88 percent, which would be considered to be much more acceptable for purpose identification. Further improvements may be possible beyond that by using geographic files of land use, which have not yet been used.

Table 4: Comparison of PR and GPS Software Origin Activities

PR Origin Activity	GPS Origin Activity						Total
	1 At Home	2 Paid Work	3 School	6 Social, Recreational, Church	9 Shop	15 Other	
0 Unidentified	0	1	0	0	0	12	13
1 At Home	191	2	1	31	6	46	277
2 Paid Work	5	66	1	13	3	75	163
3 School	4	1	2	12	1	16	36
4 Volunteer Work	2	0	0	1	0	6	9
5 Pick up/Drop Off person	4	3	6	8	4	17	42
6 Social, Recreational, Church	5	1	0	10	6	29	51
7 Catch a Bus, Train, or Airplane	1	0	0	1	0	2	4
8 Transfer from One Bus, Train or Airplane to Another	0	2	0	0	1	7	10
9 Shop	8	3	2	32	21	76	142
10 Personal Business	5	1	0	19	4	40	69
11 Eat Meal	5	2	0	18	4	25	54
12 Go for a Drive	2	0	0	0	1	6	9
13 Work Related	2	8	0	2	3	35	50
14 School Related	1	0	0	4	1	7	13
99 Don't Know/Refused	9	0	0	23	3	38	73
TOTAL	244	90	12	174	58	437	1015

Table 5: Comparison of PR and GPS Software Destination Activities

PR Destination Activity	GPS Destination Activity						Total
	1 At Home	2 Paid Work	3 School	6 Social, Recreational, Church	9 Shop	15 Other	
0 Unidentified	1	1	0	0	0	12	14
1 At Home	195	2	1	40	7	52	297
2 Paid Work	1	63	1	14	2	74	155
3 School	4	0	1	11	1	12	29
4 Volunteer Work	1	0	1	2	0	5	9
5 Pick up/Drop Off person	3	4	7	7	4	16	41
6 Social, Recreational, Church	6	1	0	10	5	31	53
7 Catch a Bus, Train, or Airplane	0	0	0	1	0	3	4
8 Transfer from One Bus, Train or Airplane to Another	0	1	0	0	1	6	8
9 Shop	5	3	2	30	20	76	136
10 Personal Business	5	2	0	17	5	40	69
11 Eat Meal	6	2	0	17	4	24	53
12 Go for a Drive	3	0	0	0	1	7	11
13 Work Related	1	8	0	3	3	34	49
14 School Related	1	0	0	2	1	7	11
99 Don't Know/Refused	11	0	0	22	3	39	75
TOTAL	243	87	13	176	57	438	1014

5. Conclusion

The PR data show that a number of respondents insist on combining trips in a tour into a single trip. Also, there are a few cases where respondents have attempted to split a trip, but created a second trip of less than 1 minute duration, which may represent an error in trip splitting. Only one trip was added that was missed by the GPS, and the overall combination of map editing and software processing for trip identification appears to be achieving a high level of accuracy.

Mode of travel is reasonably well identified, although there appear to be some problems with bicycle and bus trips in particular. Correction of these problems will produce a much better result from mode identification. Discrepancies between walk and car may not be possible to eliminate, if these are predominantly short and slow trips, such as driving from one house to a nearby house in the neighbourhood. For the most part, however, mode appears to be reasonably accurate and there are few major discrepancies.

Trip purpose is more troublesome, with a significant number of mismatches from the software that should be possible to eliminate. This will be the primary subject of software improvement in the next few months. Utilisation of land use GIS files and other improvements to the software are expected to assist on this work.

Overall, however, the accuracy with which it is possible already to identify trips, mode, and purpose, and currently in-progress work on occupancy, show considerable promise for the potential that GPS can replace conventional travel diaries. It appears that the GPS devices now in use, such as the GPS-PPAL being used in the GCAHTS study, are sufficiently accurate and sufficiently easy for respondents to carry, that the GPS devices themselves provide very accurate information on travel behaviour. The key now to further advancing the use of GPS as a replacement for conventional diary surveys lies in further refinements to the processing software. The extensive PR survey now underway as part of the GCAHTS study should provide a wealth of data to permit changes and improvements to be made. Future surveys may also benefit from having a small subsample of respondents undertake the PR survey, to validate the accuracy of the software, especially under circumstances where the mix of modes or the overall available transport network differs significantly from that used in this research.

References

Auld, J.; Williams, C.; Mohammadian, A.; Nelson, P. (2009). An Automated GPS-Based Prompted Recall Survey with Learning Algorithms. Transportation Letter, 1 (1), 59-79.

Bachu, P.K.; Dudala, T.; Kothuri, S.M. (2001). Prompted Recall in Global Positioning System Survey: Proof-of-Concept Study. Transportation Research Record 1768, pp. 106-113.

Jones, R. and N. Pitt (1999). Health Surveys in the Workplace: Comparison of Postal, email and World Wide Web Methods, *Occupational Medicine*, **49** (8), 556-558.

Lee-Gosselin, M.E.; Doherty, S.T.; Papinski, D. (2006). An Internet-Based Prompted Recall Diary with Automated GPS Activity-trip Detection: System Design. Proceedings of the 85th Annual Meeting of the Transportation Research Board, January 2006, Washington, DC.

Li, Z.J.; Shalaby, A.S. (2008). Web-Based GIS System for Prompted Recall of GPS-Assisted Personal Travel Surveys: System Development and Experimental Study. Proceedings of the 87th Annual Meeting of the Transportation Research Board, January 2008, Washington, DC.

McDonald, H. and S. Adam (2003). A Comparison of Online and Postal Data Collection Methods in Marketing Research, *Journal of Marketing Intelligence and Planning*, **21** (2), 85-95.

NOAA (2010). Selective Availability. Space-Based Positioning, Navigation and Timing, US Department of Commerce, http://pnt.gov/public/sa/ Accessed 2 August, 2010.

Stopher, P.R.; Bullock, P.; Horst, F. (2002). Exploring the Use of Passive GPS Devices to Measure Travel. Proceedings of the 7th International Conference on Applications of Advanced Technologies to Transportation, edited by K.C.P. Wang, S. Medanat, S. Nambisan, and G. Spring, ASCE, Reston VA, pp. 959-967.

Stopher, P.R.; Collins, A. (2005). Conducting a GPS Prompted Recall Survey over the Internet. Proceedings of the 84th Annual Meeting of the Transportation Research Board, January 2005, Washington, DC.

Stopher, P.; FitzGerald, C.; Zhang, J. (2008). In Search of a GPS Device for Measuring Personal Travel. Special Issue of *Transportation Research C*, on Emerging Commercial Technologies, 16, 350-369.

Stopher, P.R. (2009). Collecting and Processing Data from Mobile Technologies. In Bonnel, P.; Lee-Gosselin, M.; Zmud, J.; Madre, J.-L. (editors), Transport Survey Methods: Keeping Up With a Changing World, Emerald Group Publishing Ltd., England, pp. 361-391.

Stopher, P.; Wargelin, L. (2010). Conducting a Household Travel Survey with GPS. Paper to be presented to the World Conference on Transport Research, Lisbon, Portugal, July.

TRB (1996). Conference on Household Travel Surveys: New Concepts and Research Needs. Transportation Research Board Conference Proceedings No.10.

Wagner, D.P. (1997). Lexington Area Travel Data Collection Test: GPS for Personal Travel Surveys. Final Report for OHIM, OTA, and FHWA, USDOT.

Wolf, J.; Guensler, R.; Bachman, W. (2001). Elimination of the Travel Diary: An Experiment to Derive Trip Purpose from GPS Travel Data. Transportation Research Record 1768, pp. 124-134.

Wolf, J.; Loechl, M.; Thompson, M.; Arce, C. (2003). Trip Rate Analysis in GPS-Enhanced Personal Travel Surveys. In Stopher, P.R.; Jones, P.M. (editors), Transport Survey Quality and Innovation, Pergamon Press, Oxford, England, pp. 483-498.

Wolf, J. (2006). Applications of New Technologies in Travel Surveys. In Travel Survey Methods: Quality and Future Directions, Stopher, P. and Stecher, C. (editors), Elsevier, Oxford, pp. 531-544.

Wolf, J. (2009). Mobile Technologies: Synthesis of a Workshop. In Bonnel, P.; Lee-Gosselin, M.; Zmud, J.; Madre, J.-L. (editors), Transport Survey Methods: Keeping Up With a Changing World, Emerald Group Publishing Ltd., England, pp. 393-401.