I T L S

WORKING PAPER

ITLS-WP-06-16

Pilot Testing a GPS Panel for Evaluating TravelSmart®

By

Peter Stopher, Camden FitzGerald & Tony Biddle

July 2006

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-06-16

TITLE: Pilot Testing a GPS Panel for Evaluating TravelSmart®

ABSTRACT: TravelSmart initiatives are an important element of the national program to reduce greenhouse gas emissions from cars. In this paper, the authors report on the design of a long-term panel for evaluating the effects of TravelSmart initiatives, and report on a pilot test of the methodology, using a combination of GPS surveys and collection of odometer readings from participating panel members. Two tests are documented, in one of which respondents provided data for one week, using portable GPS devices, while respondents in the second survey carried devices for four weeks. The results of the tests are compared, and conclusions drawn on the ideal period of measurement. We also speculate on the sample size requirements.

KEY WORDS: ***TravelSmart, Evaluation, GPS (Global Positioning Systems), VKT, Sample Size, Panel, Variability***.

AUTHORS: Peter Stopher, Camden FitzGerald & Tony Biddle

CONTACT: Institute of Transport and Logistics Studies **(C37)**
An Australian Key Centre
The University of Sydney NSW 2006 Australia

Telephone: +61 9351 0071
Facsimile: +61 9351 0088
E-mail: itlsinfo@itls.usyd.edu.au
Internet: http://www.itls.usyd.edu.au

DATE: July 2006

1. Introduction

Over the past decade, several governments in Australia have tested and implemented initiatives to reduce dependence on the car and increase the use of environmentally friendly travel alternatives, through a policy that is known as TravelSmart®[1] (DPI, 2006; TravelSmart Australia, 2006a). As described by TravelSmart Australia (2006a), this policy is one of the key elements in a national campaign to reduce greenhouse gas emissions in Australia, by using it to reduce the emissions from cars. The states of Queensland, South Australia and Victoria, and the Australian Capital Territory have joined with the Australian Greenhouse Office to form the National Travel Behaviour Change Program (NTBCP), which aims to promote the adoption of TravelSmart strategies in the three states and territory involved, as one of the contributions to reducing greenhouse gases in Australia under the Greenhouse Gas Abatement Program (GGAP) of the Commonwealth Government (TravelSmart Australia, 2006b).

An important element of this program of greenhouse gas reduction is to determine its effectiveness over the period of the Kyoto agreement from December 2007 to December 2012 (AGO, 2006). Whilst various claims have been made for the reduction in car kilometres of travel that can be attained with TravelSmart, it is currently unknown to what extent these reductions would be sustained, become greater, or erode over the long term and, therefore, whether or not the targets for greenhouse gas abatement that have been set for this initiative would be achieved. As a result, the NTBCP partners sought assistance in designing a long term monitoring procedure that might be capable of measuring the effectiveness of TravelSmart in attaining the GGAP goals for transport among the partner states and territory.

In the past, evaluations of voluntary travel behaviour change (VTBC) projects have concentrated on their short-term impacts, usually within a year or less of implementation, and sometimes followed by further evaluations for up to three years. We reviewed not only the strategies used in short-term evaluations, but also those used in other fields (particularly health and epidemiology) for both short and long term evaluations. One of the features of short-term evaluation is that the effect that is to be measured is that of a significant reduction in vehicle kilometres travelled (VKT) as a result of the implementation of the VTBC. Thus, the goal of short-term evaluations is to measure how big a change takes place in the behaviour of those who take up the VTBC program, compared to either prior behaviour or the behaviour of similar households that did not take up the program or to whom it was not offered. In this context, the task is to devise a measurement method that can reliably measure a change in behaviour that may range from an overall reduction of 1 or 2 percent to as much as 10 or 15 percent of car driver trips or VKT.

In long-term monitoring, a different situation arises. In this case, households will have taken up the VTBC at some earlier date, and will, presumably have reduced their car driver travel at that time. By the time the long-term monitoring takes place, it will not be an issue of looking for a change in behaviour, but rather of looking to see to what extent a previous change is being maintained. This makes the monitoring more difficult, mainly because we do not know what size of change we wish to measure, because it could be a zero change, a

[1] TravelSmart® is a name that was originally registered by the Government of Western Australia to apply to policies of this type. It is now used broadly throughout Australia for policies of this type.

further decrease in VKT and car driver trips, or an increase. Some explanation of this is in order.

For households that participated in VTBC, the expected result was a decrease in VKT (and car driver trips) within the months immediately following implementation of the VTBC. This is shown by the black line **in Figure 1**. After this change occurred, perhaps in the first six to twelve months, several things could happen. In the worst case, the change would be temporary, and the household would, after a year or two, revert to their pre-VTBC behaviour, as shown by the solid red line in Figure 1. A case that is not much better is represented by the dashed red line, and indicates that the household's VKT subsequently grows faster than the population at large, until it eventually equals the levels if no VTBC would have occurred. In the best case, the household would continue to reduce VKT and car driver trips, as shown by the solid green line. A good outcome, would however, be represented by the blue lines, where VKT and car driver trips either remain static (solid line), or rise more slowly than current growth rates in VKT (dashed blue line). In the meantime, in the population that has not been exposed to VTBC, two possibilities occur. The first is that VKT and car driver trips continue to grow, more or less as they have been for the past two decades and more. Second, there could be some diffusion of VTBC that would lead to either a lower rate of growth, or even some decline in total VKT. These situations are shown by the pink dotted line and the pink dash-dot line. Other factors will, of course, interfere with all of these projections, as noted below.

Any of the outcomes represented by the green and blue lines would be indicative of long-term success of travel behaviour modification. A result such as the red lines would be indicative of failure of the program in the long term. To know if the solid green or solid blue lines are descriptive of the result of the long-term effects of VTBC is relatively simple, because, over time, the trend will become apparent. To judge if the long term situation is either the dashed blue line or the dashed red line requires knowledge of the pink dotted line. However, if diffusion has taken place, it will be difficult to know whether what is appearing in the population that did not receive tools is a diffusion effect or no effect, because other economic and social forces could cause the pink dotted line to be flat, or even to decline. One would also have to be measuring the population that was not exposed to intervention, and is probably too far removed from the target populations to be affected by diffusion, in order to have any certainty about whether one is measuring the red dashed line, or the blue dashed line. That the pink dotted line is potentially realistic is illustrated by the following data. From 1979 to 2003, the population of Australia increased by about 37 percent (ABS, 2003a). In the same time period, passenger VKT increased by over 70 percent, and passenger VKT per person increased by over 26 percent (ABS 1998 and 2003b). In most years in that period, average population growth was around 1.4 percent per annum, but VKT grew at over 2 percent per annum.

1.1 Key Challenges for the Long-Term Monitoring Program

Clearly, the context for long-term monitoring of VKT is complex and brings with it a number of fundamental challenges, which could be summarised as how to:

- Measure VKT accurately (as an indicator of greenhouse gases)
- Repeat this measurement at regular intervals over an extended period of time (5 years)

- Do this for a representative sample so as to make robust statistical inferences about changes (or lack of changes) in VKT for particular populations (e.g., different socio-demographic groups, different types of urban location, etc.)
- Distinguish correctly changes in VKT that are due to the VTBC interventions from those that are due to other underlying social and economic factors
- Corroborate this detailed information with more macro measures/indicators of declining VKT from other sources
- Factor diffusion effects from TravelSmart interventions into this assessment
- Conduct the monitoring with minimal respondent burden
- Do all of this as cost-effectively as possible.

A troublesome issue here is that of the level of analysis. It may be considered most desirable to be able to undertake a fully disaggregate analysis, for example, tracking a household from before the intervention to whatever point it is possible to measure that household in the long-term monitoring period. However, doing this will raise a number of serious issues about how to deal with changes that take place within the household, as well as behaviour changes caused by such exogenous influences as prices of petrol, levels of unemployment, value of the dollar, capital works in transport in the vicinity, etc. To date, most of the reporting of the extent of short-run change from VTBC has been at an aggregate level. Generally, evaluation results are reported either for the general population, or for those households that actually participated in the VTBC. In the former case, results have usually been provided for the entire suburb where the intervention has taken place. In the latter case, evaluation results are reported for all households that participated in a specific project.

2. Choice of Measurement Method

2.1 Travel Diaries

Travel diaries have been the method of choice in short-term evaluations to date. However, they present a number of problems. First, while Richardson *et al.* (2003) have shown that it would be preferable to collect data for more than two days, it is too burdensome to have respondents complete a diary for more than about two days. Second, response rates to diary surveys range from 20 to about 55 percent, from one-time cross-sectional surveys. It can be expected that responses to repeated waves of a panel survey will decline well below these figures. Third, it has been shown (Wolf *et al.*, 2003) that diary surveys, where the data are retrieved by telephone, under-report trip making by 20 to as much as 60 percent. It is not known by how much a postal diary under-reports trip making, but it is expected to be greater than with telephone retrieval, because of the lack of interviewer intervention. There is no information to indicate whether such under-reporting would be consistent on a household-by-household basis in a repeated survey, such as a panel.

However, the measure of greatest importance in this evaluation is VKT. There are three ways to derive VKT from a diary. One is to ask people to report the distances that they travel on each trip reported in the diary. However, analysis against Global Positioning System (GPS) data suggests that distances are usually over-estimated by an average of 10 percent, but with substantial variability in measurement accuracy (Stopher *et al.*, 2005).

Second is to calculate minimum time paths between each origin and destination. However, people are notoriously poor at providing accurate address information for the places they visit, and it is impossible to get actual routes chosen, and the minimum time path is probably not a good measure of actual distances driven. Neither of these methods provide reliable information on VKT, and would make it very difficult to ascertain the extent of changes in VKT, especially when those changes are relatively small. Third, odometer readings could be requested from respondents. Two odometer reports would be required for each vehicle to determine a distance travelled over some period of time. In a diary survey, one could request an odometer reading at the start of the first diary day, and one at the end of the diary period. However, exact compliance with recording odometer readings at those specific times is impossible to monitor. Experience also shows that many respondents will remember to provide the first odometer reading, but forget the second one (Stopher *et al.*, 2006a).

Finally, the diary collects much more information than is required for long-term monitoring. To obtain VKT, the first two methods require people to provide detailed trip-by-trip reporting for one, two, or more days for each person in the household. This is far more information than is required for this monitoring program. Indeed, it can be argued that the only part of the above information that is really needed is a periodic odometer reading. However, this will also not allow us to assess changes in overall trip making, nor in mode use.

2.2 Interviews

An alternative to postal travel diaries is an interview. Face-to-face interviews involve interviewers travelling to the homes of respondents and interviewing them about their travel behaviour. This has a similar level of burden to the travel diary, but is much more expensive, although the response rates are generally higher, at 75 to 85 percent. Also, GPS validation shows that underreporting in face-to-face interviews is much lower at 7-12 percent (Stopher *et al.*, 2005). Hence, face-to-face interviews provide more accurate information about travel than self-administered diaries. Also, a face-to-face interview can achieve accurate collection of odometer readings at the time of the interviewer visit. However, if two odometer readings are required, it will be necessary to have two interviewer visits. Already, face-to-face surveys are very expensive, perhaps costing in excess of $350 per completed household. As with postal diaries, this method also collects much more information than is required to measure changes in VKT. Hence, it is not a cost-effective solution.

An alternative method is to use a postal diary with a telephone interview to retrieve the data. In North America, this has been found to be a more effective alternative to the postal survey, but still suffers from considerable underreporting of travel. It is less expensive than the face-to-face interview, but otherwise has the same disadvantages of all diary surveys of collecting more information than required, and of being perceived as burdensome to the household.

2.3 The ABS Survey of Motor Vehicle Use

According to the Statistical Clearing House register "…the main purpose of the SMVU is to satisfy the information needs of Commonwealth and State government agencies responsible for the allocation of funds for road development, the design and construction of highways, the regulation of road transport operators, accident exposure and energy use analysis…" (ABS, undated). The introductory notes to the *Survey of Motor Vehicle Use: Year Ending 31 October, 2003* warn users:

"This survey has been designed to provide a measure of total distance travelled and tonne-kilometres for each state/territory of registration by type of vehicle. While comparisons are made between 2003 survey results and earlier iterations of the SMVU, *the survey has not been designed to provide accurate estimates of change*." (ABS, 2003b, italics added).

From an analysis of the probable errors in state level figures from the SMVU, it was determined that the SMVU does not enable us to even say (statistically speaking) whether VKT has actually increased or decreased. Fundamental to the evaluation of the TravelSmart program is an ability to measure sustained change over the first Kyoto protocol period. From an error standpoint, the SMVU cannot accurately estimate change and is unlikely to be useful for long-term monitoring of VTBC. It is unlikely that the ABS can separate the participating from the non-participating population, it does not measure VKT at the household level, and it cannot accurately estimate change. Hence, this would not be a potential source for long-term monitoring of VTBC.

2.4 Passive Measurement - GPS Surveys

A novel method of collecting the data is to use GPS devices. The use of GPS devices to track travel is a recent development, made possible with recent technology changes that have permitted devices to be developed that can be carried by individuals. We have pioneered the use of these devices in Australia, both in evaluating TravelSmart initiatives and in validating household travel survey data. The advantages of the GPS are that it:

- Is a passive method of data collection that requires very little from the respondent other than to carry the device with her or him for the period requested
- Records data very accurately about routes used, distance travelled, time taken, and when and where the trip takes place
- Provides a means to obtain travel data over a number of days, with very little additional burden for respondents
- Records distances for all modes of travel, and permits the analyst to infer the mode of travel. Hence, VKT and PKT can be estimated much more accurately than from diaries, and also walking and bicycling travel can be captured.

The GPS also has disadvantages that it:

- Can easily be left at home and not carried by the respondent

- May not record at the beginning of a trip, when a person is in a tunnel and also when a person is in an area in which there are major interferences with signal reception
- Is expensive to purchase and to deploy by courier to and from each sampled household
- Requires substantial work to process the data and convert the information to useful statistics.

In the newest form factors, the devices are available in the shape and weight of a mobile telephone (Stopher *et al.*, 2006b). This offers considerable potential to overcome past problems with wearable devices. At the same time, in-vehicle devices are already quite well accepted, and have been used successfully in a study in Sydney. In applications in Australia to date, the devices have been used for up to one month.

2.5 Respondent Burden

The other clear intuitive appeal of moving away from a diary-based approach is reduced respondent burden. While some may argue the use of GPS is more burdensome, recent design initiatives (miniaturisation and passivity of devices) suggest this is fast becoming a redundant argument. On the downside, we recognise that the desire to form a panel for (potentially) five years or more, and to increase the length and frequency of monitoring will place significant demands on individuals. The request to carry a GPS device for a week or a month and then return it to the survey firm is a low burden activity, compared to any type of diary survey.

3. Sampling Mechanism

3.1 Repeated Cross-Sectional Samples

Repeated cross-sectional samples are samples that are drawn independently from a target population at each period. Such samples are relatively cheap and easy to obtain, compared to other alternatives. However, a repeated cross-sectional sample has to be a large sample to measure small changes, because of the independence (Stopher and Greaves, 2006). We have to account for both variability between the two separate samples and the variability in VKT over time. This requires very large samples for robust inferences. In addition, such samples cannot permit any form of disaggregate analysis, in which one would compare the behaviour of a given household or small group of households over a period of time.

3.2 Panel Surveys

Panel surveys are those in which the same people are surveyed on each occasion. There are four major reasons to propose a panel. First, a panel will enable the tracking of change over time for specific cohorts of households. Such a *dynamic* assessment of change is essential for true understanding of how sustained any changes in behaviour are. Second, the sample size required for measuring a change in behaviour is very much smaller (Stopher and Greaves, 2006). Third, a panel is more conducive to the formation of target and control groups, for separating out exogenous and endogenous change. Finally, while initial costs of recruitment are higher, in the long-term a panel *may* prove a cheaper option than a cross-sectional survey with an equivalent number of respondents (Armoogum *et al.*, 2006). In fact, even if the unit cost per respondent is higher in a panel survey it is probable this will be outweighed by the significantly smaller sample sizes required to achieve similar levels of statistical reliability.

There are also challenges associated with using panels. First, panels are always subject to attrition, i.e., premature drop out of people or households. This arises for a number of reasons, principally from people changing their minds about participating, as well as moving away from the survey area, death, or dissolution of a household. Panel attrition in the USA has been estimated to run at about 20 to 30 percent per year for diary-type surveys. Second, initial recruitment is harder and more expensive than a simple one-off cross-sectional survey. An incentive may be needed to have a household join the panel, and it is usually necessary to maintain contact with households to keep their interest in continuing to participate. Further, when updating characteristics from a previous wave of the panel, there is a greater analysis task in retrieving and reproducing for the panel members the information provided on the last occasion. To reduce respondent burden, each household should be provided with a copy of their household characteristics from the prior wave and asked if anything in the household has *changed*. Third, conditioning may occur, in which, as the person or household continues in the panel, their participation in the panel may cause them to change the behaviour being measured. The problems caused by attrition, recruitment, and conditioning raise particular issues when it comes to the use of the results to infer changes at the population level.

3.3 Panel Design

The simplest panel design is one in which participants are recruited and then retained through the duration of the study. There are two possible ways to deal with the inevitable attrition that such a panel incurs. The first is to start with a panel of sufficient size that anticipated attrition will reduce the panel to the desired size by the end of the time period for which the panel will function. This is called a **subsample panel**. Thus, supposing that a panel of 500 households was desired for a period of 5 years, with anticipated attrition of 20 percent per annum, one would start out with a sample of 1225 households, which would be expected to decline to 980 in the second wave, 785 in the third wave, 625 in the fourth wave, and 500 in the fifth wave. This would involve a total number of 4,115 surveys.

The second method is called partial replacement. In this method, the number of households that leave due to attrition each year are replaced by new panel members. This is called a **refreshed panel**. Thus, one might recruit a panel of 625 households, and expect to replace 125 households at the second and each subsequent wave. This would mean that every pair of waves would have 500 households, whose data could be compared across the two waves, although of the original 625 households, only 256 would, in this case, be expected to remain at the end of 5 waves, and to have provided data throughout the entire study. However, this would involve 3,125 surveys, or about 1,000 less than the first method.

A major problem with the subsample panel is that, by the end of the measurement period, the panel may be quite a bit different from the population it is supposed to represent. This can also happen in the refreshed panel, if the replacement members are selected to be as similar as possible to those lost by attrition. One method of benchmarking this is to use a design called a **split panel**. At each wave of measurement of a split panel, which may be either a subsample panel or a refreshed panel, a separate cross-sectional sample is also drawn and surveyed. This provides greater accuracy about the changes occurring in the population, but is also a very expensive design, in that the cross-sectional sample must still be fairly large, and the panel size cannot be reduced significantly. A variant on the split panel is where the cross-sectional survey is conducted less frequently than the panel waves. However, this loses much of the benefit of the split panel, and is useful only as an occasional check on the make up of the panel.

The fourth panel design is known as a **rotational panel**. A rotational panel deals with attrition by recruiting panel members for a pre-defined amount of time that is less than the measurement period. For example, in the US, the Bureau of the Census uses a rotational panel for a quarterly income and expenditure survey of households, which is a continuing panel. Households are recruited for this survey and asked to remain in the panel for a period of about three years. At the end of three years, the household is replaced. The rotation is designed in such a way that only a fraction of the panel is replaced at each wave. The three major advantages of this type of panel are that it puts a limit on the total burden to participants of being in the panel, it can remain more representative of the underlying population from which it is drawn, and the design minimises conditioning, by keeping respondents in the panel for a short enough time that conditioning is relatively minor. An additional benefit of a rotational panel in this context is that it is possible to ask questions about TravelSmart initiatives as households leave the panel.

For the purposes of the pilot test reported in this paper, we have used a refreshed panel. However, the intent is that the long-term monitoring would actually use a rotational panel. Because of the shortness of the pilot survey, a rotational panel was not feasible. It is expected that there will be a maximum of three waves of the pilot panel, spanning one year in total duration.

3.4 Attrition Reduction Strategies

Non-response, fatigue, and loss of interest are the major forms of attrition that can be mitigated. In the past, it has been found that continued contact and sharing of results with the household help to maintain interest and reduce non-response. Therefore, a series of re-contacts should be planned for each household, including sending Christmas or Holiday greetings around the Christmas/New Year season, and sending birthday greetings to household members at the appropriate time. Sending households details of responses may be inadvisable because of the potential of such information to condition behaviour.

Second, participation can occur every three or six months. This frequency of contact should help to keep people more interested and involved in the panel. Also, for those households that are contacted by telephone, if the same interviewer contacts the household each time, a rapport will be built up between the interviewer and the household – empirical evidence has shown this issue of interviewer maintenance (which is easily over-looked) to be of extreme importance in maintaining panel participation (Hensher, 1987). Third, incentives can be offered to panelists.

3.5 Frequency of Monitoring

The frequency of the survey activity depends on the measurement method employed – the higher the burden, the less frequent the survey. Thus, if a diary survey were used, frequency would probably have to be restricted to a maximum of once per year. On the other hand, for this monitoring task, the more frequent the survey the better. Using a GPS survey, an annual or semi-annual panel survey would seem to be best, although it is necessary to determine the ideal frequency through pilot surveys. In the case of a GPS survey, there is a potential trade-off between the frequency of the GPS survey – this could be once or twice per year, with all panel members being surveyed at approximately the same time, or spreading the survey throughout the entire year – and the length of time that GPS recording is performed – this could be as little as a few days to one week, or could be as long as one month.

The longer the period for which data are recorded by the GPS device, the greater the accuracy of measurement of such variables as VKT, and therefore either the sample size required will be smaller, or the survey can be undertaken less frequently. Measuring all panel participants at the same time each year provides for accurate information on VKT change from year to year. However, measurement throughout the year allows a more accurate estimate to be made of annual VKT (because there is information on seasonal variation), and therefore total greenhouse gas emissions. Measuring panel participants twice per year, with measurements spread through the entire year will provide increased accuracy on annual VKT and emissions reductions, without requiring the large number of devices that would be required if all panel members were to be surveyed within a period of a month or two.

4. The Pilot Survey

Because there are a number of unanswered questions about the potential of a GPS survey to provide the needed measurements for long-term monitoring, the NTBCP partners contracted the ITLS to undertake a pilot survey. The pilot survey is still underway. This paper reports on early findings from the survey. It was decided first to take advantage of the fact that a major TravelSmart intervention was being conducted in South Australia, under the auspices of the GGAP, beginning in mid 2005. In that intervention, it had already been decided that the short-term evaluation would be done by ITLS using a combination of two panels – one undertaking GPS surveys and the other undertaking an odometer survey. The odometer survey aspects of this pilot survey are not described in this paper.

A short-term panel of 200 households was established that would undertake a one-week GPS survey annually for three years, with the first wave of this panel taking place in late 2005, and subsequent waves targeted for October 2006 and 2007. As a supplement to this, and to address questions raised by the proposed long-term evaluation methodology, a second panel of 50 households was established in October-November of 2005, with a view to having these households undertake a one month GPS survey, which would be repeated at six-monthly intervals. At the time of writing, the first wave was completed in December 2005, and the second wave will be completed by May 2006. In this paper, we report on results obtained from the first wave of each of the one week and one month GPS surveys. However, for the one week survey, only compliance and response rates are discussed.

4.1 Effect of Length of Time of the GPS Task on Compliance

The South Australia panel was established first and recruitment for that was done by telephone. The aim for the first wave was to recruit sufficient households to obtain complete GPS records from 150 households, because we had already recruited and obtained complete data from 53 households in a pilot survey conducted back in July-August. These households were recruited to become a part of the main panel. We estimated that we would need to recruit about 175 households, because our previous experience showed that about 90 percent of the households that are recruited actually complete the GPS task (much higher than our experience with diary surveys, where about 60 percent completion appears to be more usual). To achieve a recruitment of 175 households, we estimated we would need an initial pool of 1,000 households with telephone numbers. The reason for needing such a large number is that our procedure was to draw a sample by address from a parcel-based GIS, and then to phone match to the addresses. Because the available lists of reverse-listed telephone numbers are now substantially outdated (July 2004), we expected about 25 to 35 percent of the matches to be ineligible (i.e., the telephone number is no longer in operation, or the number does not match the sampled address). Of the eligible numbers, we expected about a 30 percent success rate for recruitment. We used a similar recruitment procedure for the one-month sample. We also assumed a similar productivity of the address sampling, so for a goal of 50 households, we started with 300 phone-matched households. The actual results of our recruitment are shown in Table 1.

From Table 1, we can see that the actual response rates for the two surveys were very similar, and the non-compliance was also almost the same in the two surveys, indicating that the length of the GPS task did not have much effect on either of these two aspects of the response. On the other hand, Table 2, shows the number of days for which people actually provided data. In interpreting these results, it must be borne in mind that most people do not travel every day. In a GPS multi-day survey, we have no way of knowing when a person stays home all day and when he or she simply forgets the GPS device, or intentionally leaves it at home. Based on various anecdotal reports (Stopher et al. 2006c), we believe that average non-mobility is between 1 and 2 days per week. Therefore, it is probably close to correct to assume that those who provide data on at least 5 of 7 days and at least 20 of 28 days have complied fully with the request to take the GPS with them at all times. Of course, it is possible to do a further disaggregated analysis in which we might distinguish between those who are working or being educated and those who are retired or have no paid work outside the home. The former might be expected to travel on about 6 days per week, while the latter may be expected to travel less than 5 days per week. However, such analysis is yet to be undertaken and is somewhat questionable, because we do not have immobility rates by these categories.

Table 1: Sample Disposition for the Main Panel and the One-Month Pilot

Disposition	**Main Panel**	**One-Month Pilot**	**Total**
Initial Phone-Matched Sample	1000	300	1,300
Unused Sample	63	12	75
Used Sample	937	288	1225
Known Ineligible Households	265	81	346
Refusals	420	150	570
Unknown Disposition	85	0	85
Completion of Wave 1	151	50*	201
Households failing to comply†	16 (10%)	7 (14%)	23 (11%)

* Two of these households returned the devices early
† Households that returned no data on their GPS devices

Table 2: Number of Days for Which Data Were Recorded (Persons)

Number of Days	**Main Panel**	**One-Month Pilot**	**Total**
All days	65 (23%)	1 (1%)	66 (17%)
Up to 6 days per week	56 (20%)	13 (12%)	69 (18%)
5 to less than 6 days per week	46 (16%)	10 (10%)	56 (14%)
3 to less than 5 days per week	76 (27%)	25 (24%)	101 (26%)
1 to less than 3 days per week	40 (14%)	32 (30%)	72 (19%)
More than 0 but less than 1 day per week	0 (0%)	24 (23%)	24 (6%)
Total	283 (100%)	105 (100%)	388 (100%)

As Table 2 shows, the compliance was lower for the 28-day survey than for the 7-day survey. Allowing that 5 or more days per week on average should represent complete reporting, Table 2 shows that 59 percent of the one-week sample provided that much data, while only 23 percent of the 28-day sample did so. Clearly, this suggests that, while the recruitment rates may not vary much by the length of time that people are asked to undertake the GPS task, the actual compliance, once they have the GPS devices, is significantly poorer for the longer time period. We suspect that this means that people are more likely to forget to take the GPS with them in the longer survey, and are also more likely to get tired of doing the task and intentionally leave the GPS at home. In fact, we note from the data that in the one-week survey, 5.7 percent of people provided one day of data, while 9.5 percent of the 28-day survey also provided only one day of data. Also, in the one-week survey, 8.5 percent provided 2 days of data, and 9.5 percent of the 28-day sample provided 2 days of data. This suggests that there is a similar initial drop-out rate, regardless of the number of days for which data are requested. (Given the difference in sample sizes, 5.7 percent and 9.5 percent are not significantly different at 95 percent confidence.) If we were to assume that all those who provided more than 10 days of data in the 28-day survey would have provided at least 10 days if this had been the number of days requested, we find that 59 (56 percent) of the 105 persons would apparently have provided 10 days of data.

4.2 Effect of Length of Time of the GPS Task on Variability of Travel

The primary reason for having people carry a GPS device for a longer period of time is to determine the true variability of travel more accurately. To do this, we analysed the data from the one-month GPS, because it provides the greatest ability to ascertain the effects of longer periods of time of the GPS task. Our prior expectation was that we would see the variance in trip making tending to increase at first, as more days of recording are added, and then to level off or start to decline. As we increase the number of days of travel data from one, we would expect to see an increasing variance, because additional days will begin to account more correctly for the inherent variability in people's daily travel. However, as the number of days continues to increase, we would expect that day-to-day variability would level off or even begin to decline, because of repetitions in patterns between one week and the next. If this is the case, then there will come a point where additional days of data are not changing the mean or the variance of the number of trips per day, or other related data.

To undertake this analysis, we used the person level data, because in a household, we will get variability in how many people remember to take their GPS devices with them, and also there will be potential problems about how to treat a household in which one person has provided data for 21 days, while another person in the same household only provided data for 6 days. It should also be noted that different people started the GPS survey on different days of the week, and the analysis concerns the first, second, third, etc. days of each person's survey task. In addition, The figures that are used in the following analysis are obtained by successively averaging the results. For example, if we are looking at the average number of trips per day per person, then the day 1 data are all the observations of what people did on the first day of travel that was recorded by the GPS device. The data reported for Day 2 represent the average of the first and second days of travel recorded by the GPS. If there was no travel recorded on a day, it is assumed that the person did not leave home and that the zero travel is a valid and correct representation of their travel on that day. However, we comment further on this later.

Figures 2 and 3 show the variation in the mean and variance of per person trip rates and PKT per day for those persons who provided data for at least 21 days (although there will be days of zero travel within that period). There were 41 such persons in the one-month data set out of 105 persons who accepted GPS devices and returned some data. As can be seen in Figure 2, the data exhibit very much the expected shape. The variance rises over the first four days, then declines slowly and levels off to a value of about 10. In fact, this curve suggests that the optimum may be to obtain data for about 10 days, because, by this point, the variance has settled fairly clearly to a variance of around 10. Figure 3, on the other hand, shows that the variance in PKT per person declines for the first four days and then more or less levels out, although there is some continuing slow decline in value, due largely to increasing numbers of zero travel days (people forgetting to take the GPS with them, or omitting to recharge it over night) as duration lengthens. We found almost identical distributions for the variance in travel time per trip and distance per trip as in PKT, as shown in **Figures 4 and 5**. In all cases, a decision to have people record data for about 10 days appears to be the optimum.

We also examined the trend in the estimates of the mean. The trend in the mean trips per day and PKT per day are shown in **Figures 6 and 7**. As can be seen, the average number of trips per day starts at about 3.7, drops slightly and then increases to a maximum of 4.1 by day 6, after which it declines by day 10 to 3.9, and then slowly declines further through the remainder of the period to just over 3.6 at day 21. The average PKT starts high at 29, drops in the next three days, then increases to about 29.5 on day 6, and then declines to about 25.6 on day 14, with a continuing slow decline to about 24.9 on day 21. The latter declines in both of these figures are not expected. Rather, the figures should stabilise. Therefore, we examined the number of days of no travel reported. This started out at 9 days for the first day of recording, dropped to a cumulative average of 7.7 by day 3, and then rose steadily through the period to a maximum of 10.2 for the cumulative average by day 21. Looking at the numbers of days of no travel per day, this shows a noticeable increase, especially after about day 11. From this, we conclude that the reason for the continuing decline in average number of trips per day and PKT per day through the latter part of the period is a result of people not taking the GPS device with them or not travelling in the latter part of the period more than at the outset. Again, this seems to confirm that around 10 days is optimal, because it is after the tenth day that the number of no travel days increases and remains high for the rest of the period.

One other aspect of the zero travel days is the issue of true non-mobility. As noted earlier, anecdotal evidence suggests that average daily non-mobility is about 20 percent of persons. If we apply the figure of 20 percent to the sample used in this study, we would expect about 8.2 days of no travel per day. The mean is never quite as low as 8.2, over the 21 day period, except on the third day, where it dropped to 7.7. However, over the period from the fifth day through the tenth day, it averages between 8.4 and 8.6. After the tenth day, the average trends upwards fairly steadily, suggesting that there is an increasing tendency for people to leave their GPS devices at home. This would again suggest that about 10 days is optimal.

4.3 Sample Size

The issue of estimating a sample size is more complex, particularly as to how one treats multi-day data for the purposes of estimating sampling error. However, for the purposes of this paper, we make a simplifying assumption that we can treat each person day as an independent observation. This is not strictly correct, because what a person does on one day is likely to have some effect on another day. Because we have already noted that the mean and variance of trip making become quite stable by around 7 or 8 days, and assuming that people are asked to record data for ten days, it appears that this assumption may not introduce much error. The second issue in determining sample size is to specify the level of error that one is willing to accept in the results. If we are undertaking a short-term evaluation, as discussed earlier in this paper, it is rather simple to estimate the acceptable level of error. For a long-term evaluation, this is not so simple. Table 3 shows the number of households that would be required for a ten-day GPS survey, based on various levels of error, all specified at 95 percent confidence. This table is based on several important assumptions. First, we assume that the variance of PKT will be the same at each wave of the panel and is the value of 1,500 determined from about ten days of reporting (Figure 3). We also need to know the covariance or correlation between panel measurements. We have not yet established this, because we are still in the field for the second wave. However, we have used three assumed levels of correlation in Table 3, one where the correlation is very low (0.3), one where it is of a medium level (0.6), and the third where it is high (0.9). As noted earlier, we have assumed that we can treat each person day of travel as an independent observation, which is not strictly correct, and means that the sample sizes here are probably a bit too low. Finally, we have assumed that the average number of GPS devices per household (to convert the sample size from persons to households) is 1.7, which is the figure we found in our pilot GPS panel for the first wave for those persons who gave us at least 21 days of valid data.

Table 3: Estimated Sample Sizes for a GPS 10-Day Panel of Households

Maximum Sampling Error (kms)	**Variance**	**Sample Size**		
		Low Correlation (0.3)	**Medium Correlation (0.6)**	**High Correlation (0.9)**
±0.05	1500	60000	34286	8571
±0.1	1500	15000	8571	2143
±0.25	1500	2400	1371	343
±0.5	1500	600	343	86
±0.75	1500	267	152	38
±1.0	1500	150	86	21
±1.25	1500	96	55	14
±1.5	1500	67	38	10
±1.75	1500	49	28	7
±2.0	1500	38	21	5
±2.25	1500	30	17	4
±2.5	1500	24	14	3

A sampling error of ±0.1 kilometres represents a 95 percent confidence limit of ±0.75 percent of average daily PKT from our sample data. Similarly, a sampling error of ±0.25 kilometres represents a maximum allowable error of ±1.85 percent at 95 percent confidence. This helps to set these sampling errors in context. Given the earlier discussion relating to the size of the difference that we might desire to measure, probably an error on the order of ±2 percent at 95 percent confidence is the maximum we would wish to incur. With the figures in this table, a low correlation would require a sample size of almost 2,500 households to achieve this level of error, while a high correlation would reduce this to around 350 households.

5. Conclusions

From the first wave of the pilot survey, with people undertaking a 28-day GPS survey, several conclusions can be drawn. First, there appears to be no significant difference in the recruitment rate between a 7 and a 28 day survey. Second, there appears to be little difference in the drop out rate for the first two days of the GPS survey, among those who were recruited, although the 28-day survey shows some signs that the drop out rate may be higher in the first two days, but lower, thereafter, although the sample size is sufficiently small that this is not a reliable finding. Third, the optimal period for which to ask people to undertake a GPS survey appears to be around 10 days, based on the evidence to date from this pilot survey. It may be best to ask people to undertake the task for about 14 days, because it appears that there is always drop out, and requesting 14 days may ensure a larger proportion of the sample providing 10 days of data. Fourth, it appears that 10 days will give an acceptable level of zero travel days, averaging only slightly above the expected 20 percent non-mobility level. Fifth, the average values of trips per day, PKT per day, and time and distance per trip all appear to stabilise by between 8 and 10 days, with only the slight continuing decline that appears to be a result primarily of more people forgetting to take their GPS devices with them as the measurement period becomes longer.

Also, using a variety of assumptions, which the second and subsequent waves of the pilot panel will need to confirm or correct, it appears that a maximum error in measuring change in PKT from one time to another of about ±2 percent with 95 percent confidence would require a sample of between 350 and 2400 households, depending on the correlation between the panel waves. If the stipulated error was relaxed to ±3.5 percent with 95 percent confidence, the sample sizes decrease to a range from 86 to 600, whereas tightening the stipulated maximum error to ±1 percent would require much larger samples, ranging from 8,600 to 60,000 households. Clearly, significant thought must be given to what constitutes an acceptable level of error for measuring the long-term effects of VTBC, and with what error one is prepared to state that the reduction in greenhouse gas emissions has been achieved.

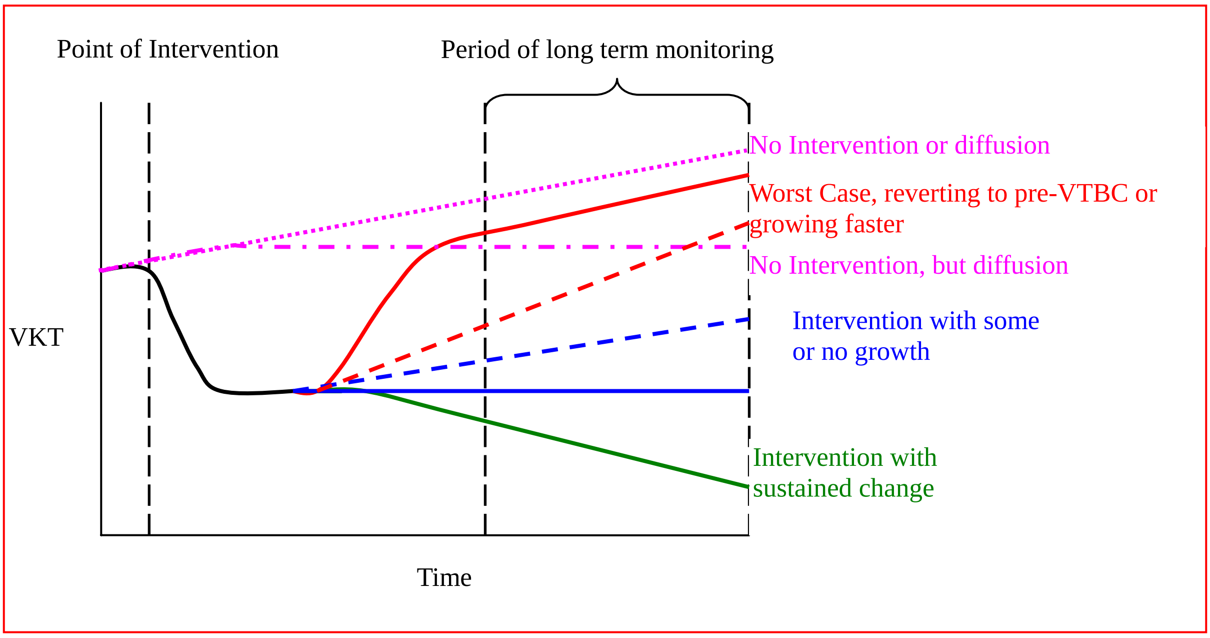

Figure 1: Potential Changes in VKT over the Long Term

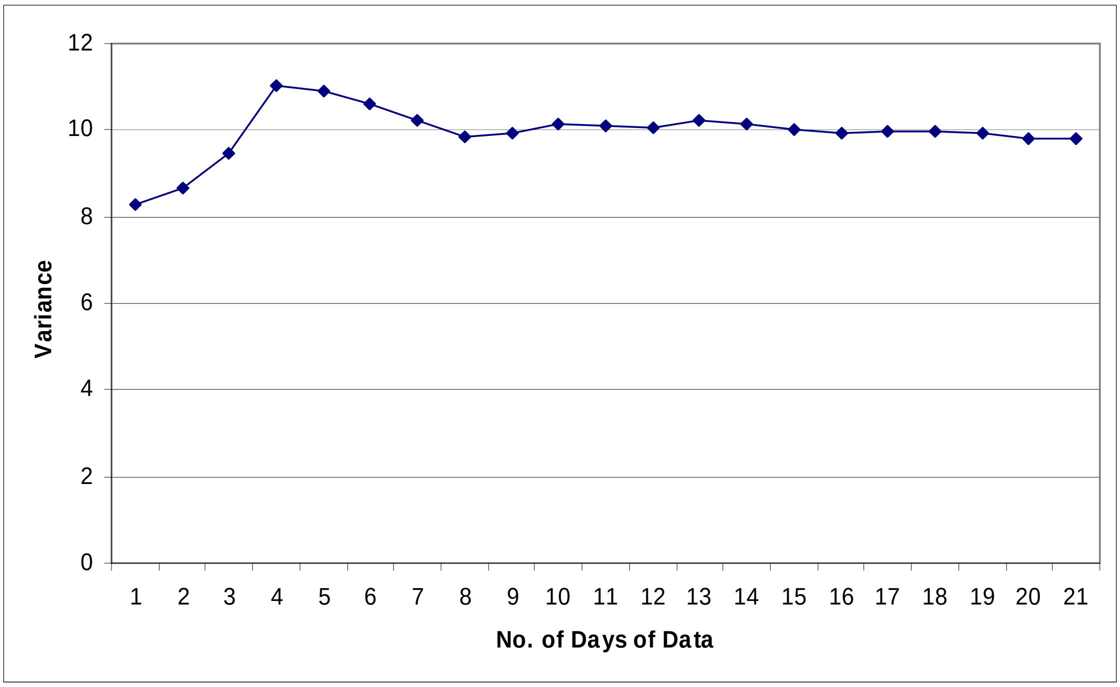

Figure 2: Variance of Person Trips per Day

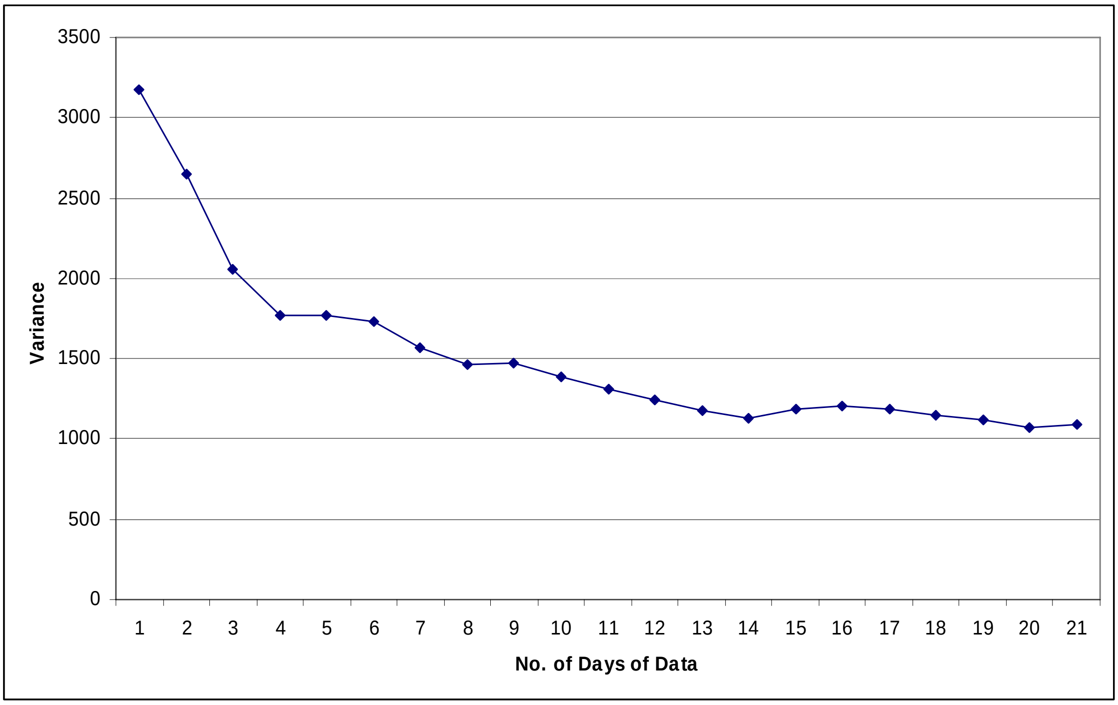

Figure 3: Variance of Person Kilometres of Travel per Day

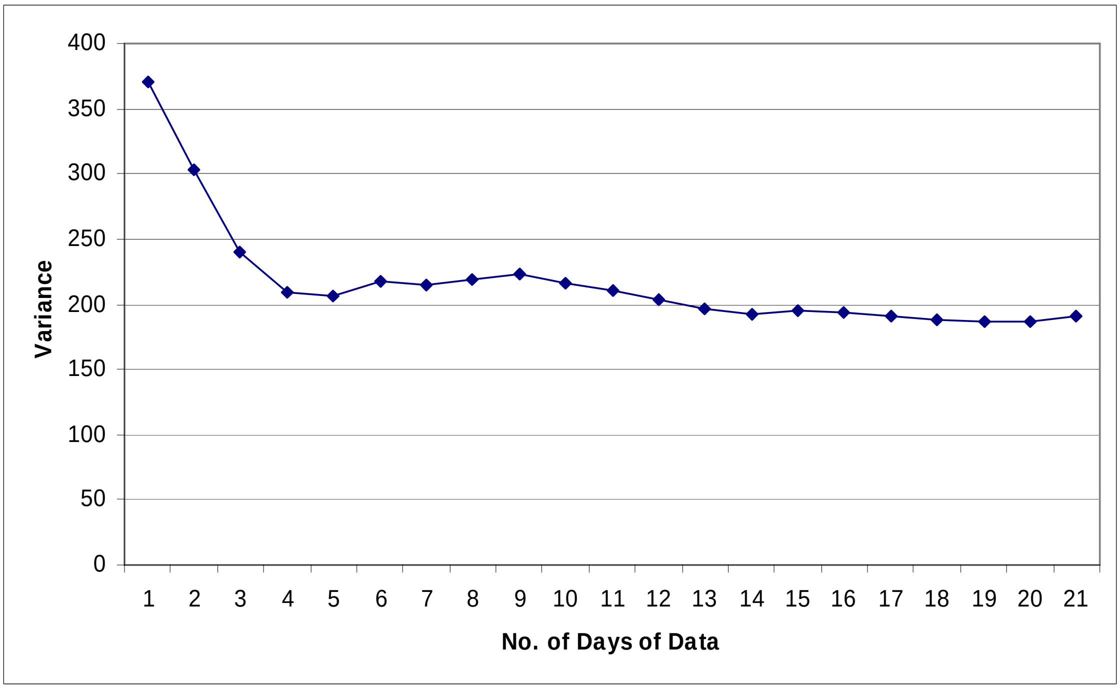

Figure 4: Variance in Travel Time per Trip

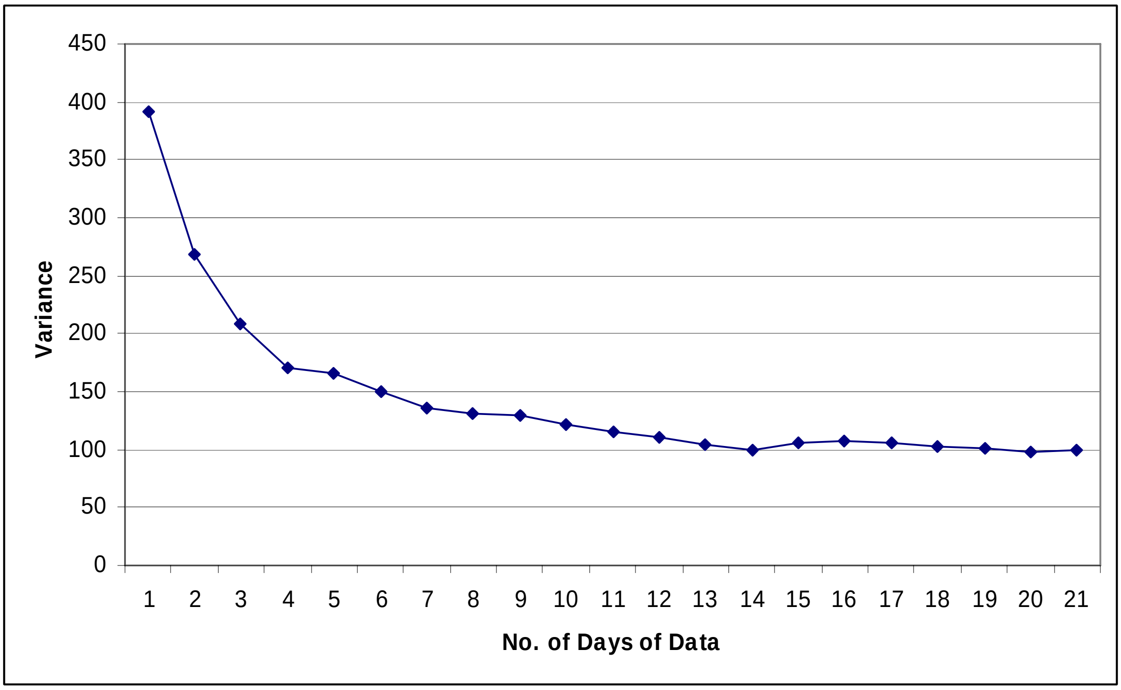

Figure 5: Variance in Travel Distance per Trip

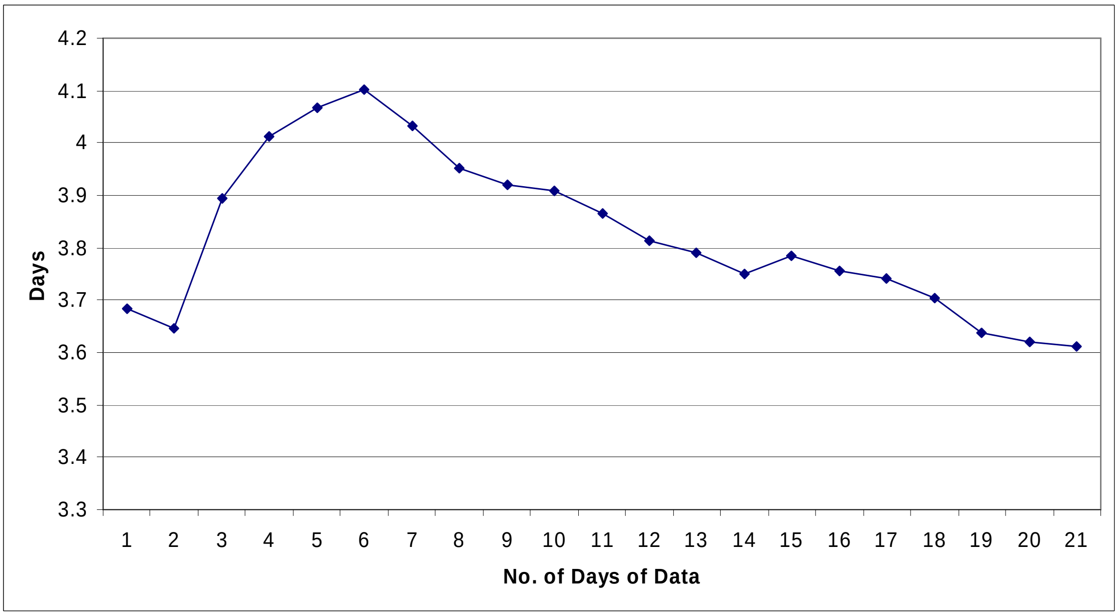

Figure 6: Average Number of Trips per Day

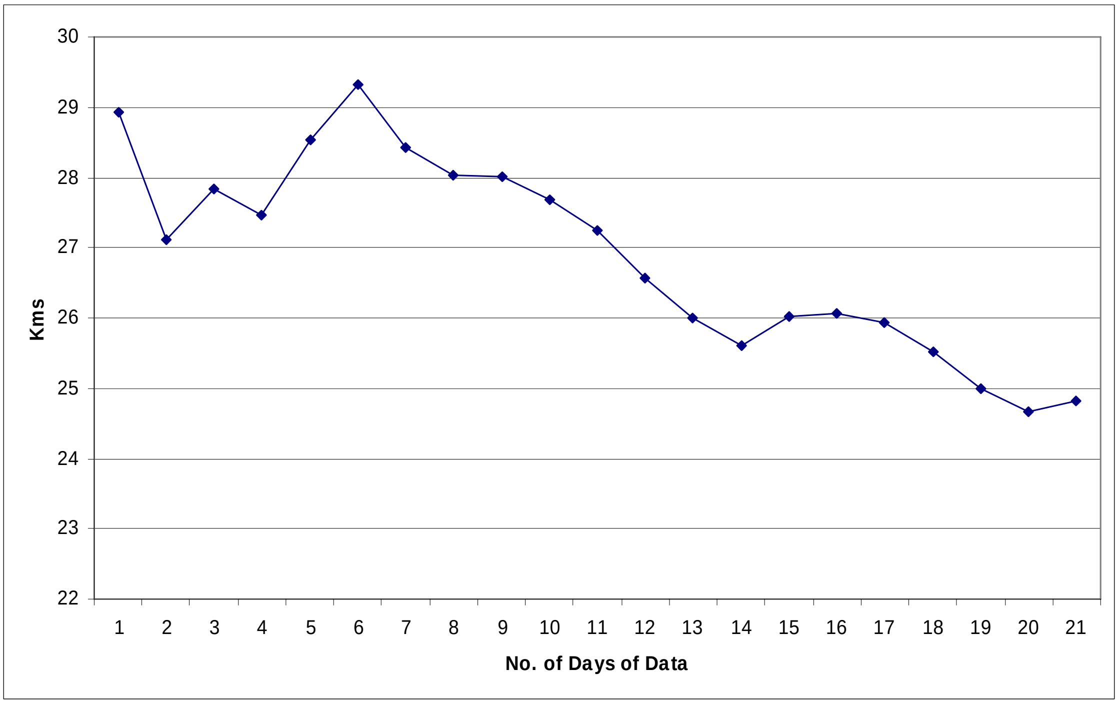

Figure 7: Average PKT per Day

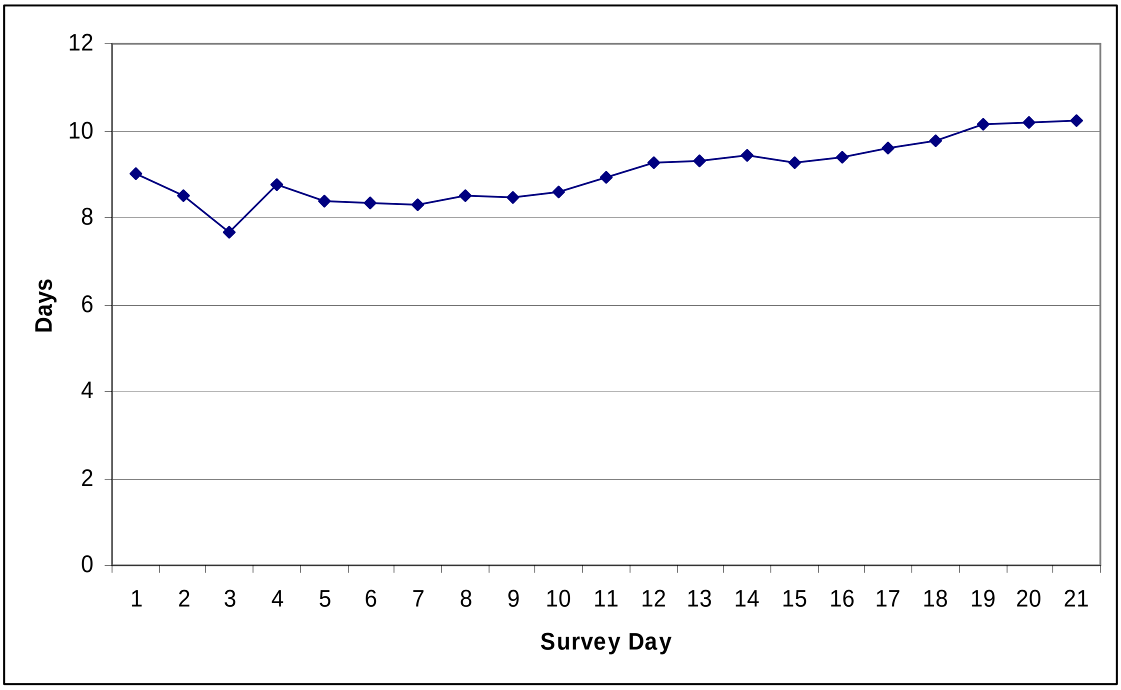

Figure 8: Cumulative Average of Days of No Travel

References

ABS (1998) *Information Paper – Survey of Motor Vehicle Use: Historical Comparison*. Retrieved: 9th September, 2004, from

http://www.abs.gov.au/Ausstats/abs@.nsf/Lookup/407EBF957FC03B29CA2568A9001394D

ABS (2003a) *Australian Historical Population Statistics - 1.Population size and growth: Table 1. Population by sex, State and Territories, 31 December, 1788 onwards*. Retrieved: 9th September, 2004, from http://www.abs.gov.au/Ausstats/abs@.nsf/lookupresponses/c5f4bf9d61d632c8ca256db00001f41?opendocument

ABS (2003b) *Survey of Motor Vehicle Use: 12 Months ended 31st October 2004*. Retrieved: 9th September, 2004, from

http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/Lookup/A97A3B6D99640F48CA256DAC000224CB/$File/92080_01%20nov%202001%20to%2031%20oct%202002.pdf

ABS (undated) *Motor Vehicle Use (Survey of), Year ending October 2002* Retrieved: 1st December, 2004, from

http://www.sch.abs.gov.au/SCH/A1610103.NSF/SurveysNameM?OpenView&M

AGO (2006) National Travel Behaviour Change (NTBC)

http://www.greenhouse.gov.au/ggap/ntbc.html , Australian Greenhouse Office,

Department of the Environment and Heritage, Australian Government, Accessed on 20/04/06.

Armoogum, J., B. Chlond, J. Madre and D. Zumkeller (2006). Panel Surveys, in P. Stopher and C. Stecher (eds), *Travel Survey Methods – Quality and Future Directions*, Elsevier, pp. 363-398.

DPI (2006) TravelSmart, http://www.dpi.wa.gov.au/729.asp Department for Planning and Infrastructure, Western Australia, Accessed on 20/04/06.

Richardson, A.J., Seethaler, R. and Harbutt, P. (2003) "Design Issues for Before and After Surveys of Travel Behaviour Change", *ATRF(26)*, CD-ROM.

Stopher, P., M. Xu, and C. FitzGerald (2005). "Assessing the Accuracy of the Sydney Household Travel Survey with GPS", paper presented to the 28th Australasian Transport Research Forum, Sydney, September.

Stopher, P. and S. Greaves (2006). Measuring a Change in Behaviour from Before and After Surveys – Sample Size Issues, *Transportation* (to be published).

Stopher, P., N. Swann, and T. Bertoia (2006a) Establishing and Using a Before-and-After Panel: Case Study of New South Wales, paper submitted for presentation to the 22nd ARRB Conference, Canberra, October-November.

Stopher, P., C. FitzGerald and J. Zhang (2006b) Advances in GPS Technology for Measuring Travel, paper submitted for presentation to the 22nd ARRB Conference, Canberra, October-November.

TravelSmart Australia (2006a). About TravelSmart,

http://www.travelsmart.gov.au/about.html , Commonwealth of Australia, Accessed on 20/04/06.

TravelSmart Australia (2006b). Government,

http://www.travelsmart.gov.au/government/index.html , Commonwealth of Australia, Accessed on 20/04/06.

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