Multi-day Household Travel Surveys: Sampling Issues

Peter Stopher[1], Stephen Greaves[1], Eoin Clifford[1]
[1] University of Sydney, Sydney, NSW, Australia

1 Introduction

Traditional household travel surveys ask respondents to report their travel behaviour for a 24-hour period, although it is well known that travel patterns vary from day to day. While this provides an indication of average household behaviour, or the behaviour on an average weekday, previous evidence suggests this may not be the most cost-effective way to collect the data because the day-to-day variability in travel is in fact substantial, requiring larger sample sizes. In addition, collecting multi-day data provides a richness of information that simply cannot be captured with a one-day survey providing insights into, for example, weekday and weekend travel, the impact of flexible working hours across the week on travel etc. Despite the intuitive appeal of multiday surveys, there are few examples and little information for designers on sampling issues, particularly sample size requirements. With this in mind, the issues explored in this paper are: 1) the reasons given for not doing multiday surveys, which primarily centre around respondent fatigue, 2) why these issues are fast becoming redundant through the use of new passive data recording technologies such as GPS, which have facilitated the possibility of collecting several days if not weeks of data with little additional respondent burden, 3) the sample size implications of extending the survey and whether in fact we reach a point of optimality in terms of the ‘ideal’ survey duration, and 4) the potential cost savings of conducting multiday surveys over 1-day surveys, even accounting for the use of new technologies.

2 Reasons for Not Conducting Multi-Day Surveys

For about the past thirty years, most household travel surveys have used a diary format that is administered prospectively either through a telephone recruitment and postal diary survey, a face-to-face interview at the household’s door, or an entirely postal survey. The earliest diaries were trip diaries, in which respondents were asked to report each trip that they undertook in a day. In most diaries, a trip was defined as being the travel from an origin to a destination, without intermediate stops except for changing travel mode or for traffic-related stops, such as traffic signals (Parvataneni et al., 1982). In other cases, the diaries were designed to collect details about each trip segment, where a segment was defined as that part of a trip that was carried out on a single mode of travel (Axhausen, 1995). In the former case, the diary required the reporting of what generally averaged about four trips per person, together with all of the details about modes used, time started, time ended, purpose, persons accompanying, etc. In the latter case, this might increase to an average of perhaps twelve or more trip segments, especially if it included the walk segments at the beginning and ending of each car trip.

Subsequently, starting nearly two decades ago, the trip diary was replaced by an activity diary in at least some surveys (Stopher, 1992). Concentrating on out-of-home activities, the reporting task increased to a requirement for an average of about 8 to 10 activities per day per person. Subsequently, a time-use diary was introduced (Goldenberg et al., 1995), which increased the number of events to be reported to perhaps 15 or even 20 on average, depending again on what limits were placed on the activities of interest. As can be readily appreciated, the increasing desire for activities or time use, and increasing detail about what a person is doing has led to increases in respondent burden for completing these surveys.

The vast majority of household travel surveys, especially those conducted for mainstream modelling activities, are still one-day diary surveys. Despite their intrinsic appeal and (potential) sample size reduction implications, examples of large-scale multi-day household travel surveys are in actuality relatively rare. Table 1 provides a summary of recent multi-day travel surveys:

Table 1: Examples of Multi-Day Household Travel Surveys

Region	Year	No. of Households	Survey Method	Survey Type	Days
Uppsala, Sweden	1971	296	Self-completed diary	Travel	35
Reading, UK	1973	136	Personal interview, self-completed diary	Activity	7
Dutch Panel	1984-1989	1687-1928	Weekly trip diary	Travel	7
Puget Sound Panel	Since 1989	1,700	Phone/CATI	Activity	2
San Francisco	1990	10,838	Phone/Phone	Travel	1, 3, and 5 day
Raleigh-Durham	1994	2,000	Phone/CATI	Time Use	2
German Mobility Panel	Since 1994	750-800	Weekly travel diary	Travel	7
Portland	1994/5	4,451	Phone/CATI	Time Use	2
Lexington	1995	100 persons	Pre-notification letter & phone/ PDA & GPS	Travel	7
San Francisco	1996	2,000	Phone/CATI	Time Use	2
'MOBIDRIVE' conducted in (Karlsruhe & Halle)	1999	139 (317 persons)	Face-face interviews, weekly travel diaries	Travel	6 weeks
San Francisco	2000	Not published	Phone/CATI	Time Use	2
Michigan	2005-6	15,000	Phone/CATI	Travel	2

The vast majority of household travel surveys, especially those conducted for mainstream modelling activities, are still one-day diary surveys. One of the major reasons for the concentration on one-day surveys has been the perception that the respondent burden of a one-day survey is already at about as high a level as most survey professionals would wish to go, and that extensions to two or more days will result in substantially lower response rates, and substantial increases in respondent burden. Indeed, in a recent two-day postal diary survey conducted in New South Wales, Australia, it was found that respondents used the device of stating that they did not travel anywhere on the second day as a means to decrease the level of burden in a two-day survey. It was noticeable that the apparent rate of non-mobility jumped about ten times on day two, compared to day one (Stopher et al., 2006). In the most recent multi-day survey in Michigan, the Metropolitan area of Detroit opted for a one-day diary in their add-on to the statewide survey, mainly as a result of their concerns with the drop off in reporting on the second day. This appears to confirm the concern that multi-day diaries are perceived as burdensome by respondents. Anecdotally, respondents have observed that it becomes very tedious to report on very similar travel for two or more days, where many people who go to work and school will more or less repeat their travel on each day of a multi-day diary.

3 Technological Developments and Multi-day Surveys

The arguments against multi-day surveys are largely based on self-report procedures for undertaking such surveys. Given that the two major arguments against multi-day diaries are that they are burdensome to respondents and that, as a result of the level of burden, there is a tendency for the reporting to become less accurate as the survey duration increases, any method of undertaking a multi-day survey that did not rely on self-report procedures may offer a potential mechanism for conducting such surveys. Before moving into that, however, it may be worthwhile considering why there is potential merit in multi-day surveys. To do this, we examine some results that have been obtained from self-report based multi-day surveys, such as those listed in the previous section.

3.1 Day-to-day Variability in Travel

The key driver of interest in multi-day surveys is the potential sample size reductions which result from the intra-person day-to-day variability in travel. Evidence suggests that while much of what we do is based on routine, there is considerable variability in travel behaviour. For instance, Pas and Sundar (1995) analysed the variability in trip rates, chaining, and daily travel times from the three-day Seattle data set. They found the intra-personal variability in trip rates is 38 percent, compared to 50 percent from the Reading data set, something attributed to the longer reporting period (five days for the Reading data set). In a more recent study, Pendyala compared intra-personal variability in trip rates and travel times with these earlier studies using GPS data collected for the Lexington pilot study (http://www.fhwa.dot.gov/ohim/gps/travel_char.htm). He reports intra-personal variability in trip rates for the three-day weekday sample at 49 percent, which is higher than the 38 percent reported by Pas and Sundar (1995). For the 3-5 day sample, this variability increased to 62 percent, which is higher than the (directly comparable) five day Reading survey. The higher intra-person variability captured by GPS is attributed to the fact it is better able to measure infrequent and irregular behaviours that tend to be missed in self-reported diary surveys.

To our knowledge, the longest duration survey completed is the six week MobiDrive survey (Axhausen et al., 2002), which was completed in 1999 in the German cities of Halle and Karlsrhue. The extended nature of the survey enables unique analyses of variability in behaviour. For instance, Richardson (2003) analyses the variability of a number of measures of travel behaviour, a summary of which are provided in Table 2. While this does not specifically show the impacts of extending the period to two, three, four, or more days, it is nevertheless interesting to note the reduction in variability that occurs from (in effect) sampling people on the same day of the week for six weeks and sampling them for one week.

Table 2: Day-to-Day Variability from the MobiDrive Data

Measure	Daily		Stratified by Day of Week		Weekly	
	Mean	CV	Mean	CV	Mean	CV
Person Car Driver Trips	1.19	75%	1.19	56%	8.37	29%
Household Car Driver Trips	2.72	77%	2.72	67%	19.1	30%
Person car driver distance (km)	12.4	118%	12.4	63%	86.8	45%
Household car driver distance (km)	28.4	99%	28.4	79%	199	45%
Person car driver travel time	22.3	107%	22.3	59%	156	42%
Household car driver time	50.9	93%	50.9	72%	357	38%

The only other analysis of this type we have found is from the five-week 1971 Uppsala survey, where Hanson and Huff (1988) found the average number of home-to-home journeys and stops for one selected week were identical to those from the five week period.

3.2 Sample Size Implications

While there are many studies of inter-day variability, few have dealt with the issues specific to this paper of how this (potentially) impacts sample size requirements. The suggestion is that the savings in sample size and costs are significant, even if one allows for the additional expense and attrition attributed to additional days of data collection. For instance, using the data collected in Reading, Pas (1986) showed that three-day data from a 75-person sample and two-day data from a 91-person sample gave the same level of precision in parameter estimates as a 1-day sample of 136 persons.

From these analyses, it appears that there is considerable potential value in collecting multi-day data in that it will permit analysis of day-to-day variability, and will also have some interesting sand useful implications for sample sizes in household travel surveys. These two reasons would be sufficient to justify interest in any procedure that might allow multi-day surveys to be undertaken, where the issues of respondent burden and fatigue in reporting would not play a role.

3.3 Technological Developments

Within the past decade, the use of Global Positioning System (GPS) devices to track people's travel has emerged as a serious contender for a new method to collect household travel survey data. A brief history of the use of GPS devices in travel monitoring can be found elsewhere (see, for example, Wolf, 2006; Stopher et al., 2007). Suffice it to say that GPS devices for monitoring travel have progressed rapidly in the past decade to the point where a device is now available that can be carried as a personal GPS device, that can store weeks of data for an individual, that weigh less than 50 grams, and are smaller than a mobile telephone. In work undertaken by the authors, using a device that predate this latest device, we have found that it is quite feasible to ask people to carry the devices with them for periods as long as one month. While we have not tried periods longer than this, there would not appear to be any obvious reason why people would not be willing to carry the devices for even longer periods of time.

In terms of respondent burden, there is little incurred with these devices. The respondent still is asked to complete a short self-report survey about characteristics of themselves and their household, and to provide addresses for certain places visited on a regular basis (e.g., work place, schools, shops used frequently). We also ask respondents to complete a brief card that indicates on which days during the study period, the respondent did not leave home, and to also indicate any days on which the device may have been left at home inadvertently, or had run out of charge and was not recharged in time. Other than that, all the respondent is asked to do is to carry the device with them throughout the requested study period, and to remember to recharge it when the opportunity presents (e.g., over night, or when driving in a car).

These devices make it almost painless to collect many days of data about personal travel. In our use of these devices, we have, so far, found very few objections on the grounds of invasion of privacy, perhaps because there is nothing displayed on the devices that indicates the data being recorded and also because the data are not collected from the individual in real time. Thus, although the respondent carries the GPS device with them at all times, the researcher does not know where the respondent has travelled until the stored data are subsequently downloaded and analysed. This may occur days to weeks after the data were collected.

Response rates have been quite insensitive in our work to the length of time for which we ask people to carry the devices. In work undertaken in South Australia, we have found little evidence of a difference in either recruitment rates or the failure to comply rates among recruited households between requests for people to carry GPS devices for periods from one week to four weeks (Stopher, Swann and FitzGerald, 2007). We have not experimented with periods longer than four weeks. In an initial panel asked to undertake four weeks of data collection, we were disappointed to find only a handful of respondents who appeared to have been diligent in taking the device with them for the full period (Stopher, Swann, and FitzGerald, 2007). However, in the second wave of this panel, the number who appeared to have taken the device with them most or all of the time jumped enormously. This leads us to some confidence in asserting that the number of days for which data can be collected by this means may be even higher than the 28 that we have already attempted successfully.

With the exception of the 1991 MobiDrive project and the much earlier Uppsala Project (1971), there are no instances of data being collected successfully by diary for longer than a few days. Therefore, obtaining as many as 28 days of data from each individual without any significant concerns about respondent burden is clearly a major breakthrough in the collection of multi-day household travel data. Elsewhere, we have documented much about the variability of personal travel that we are able to derive from 28 days of data (Stopher, Clifford, and Montes, 2007). In this paper, our main concerns relate to the issue of sample size gains that could be achieved through a multi-day survey.

4 The Effect of Multi-Day Data on Sample Size

With the use of GPS for measuring personal travel, we are confronted with an issue that has largely been ignored in the past by the transport planning profession, although it is actually present in most of the data collection that has been done for the past 50 years. Sampling theory and sampling statistics always make the assumption that each observation of a sampling unit in a data set is independent of any other observation. The only situation in which this lack of independence is openly acknowledged is in the collection of panel data, where a correction is made for the lack of independence by estimating the variance of the difference of a statistic between two waves as being the sum of the variances of the statistic in each of the two waves, minus twice the covariance between the waves. In actuality, most of the transport data collected in household travel surveys should be regarded as not meeting the strict definition of independence, in that the sampling unit is generally an individual person, whilst the data are collected by sampling households. Within a household, travel decisions are not usually made independently, so the assumption of independence breaks down at the person level. This means that sampling errors are actually mis-estimated when calculated at the level of a person or a trip.

When collecting multiple days of data from the same individual, the issue of independence of the observations is much more obviously apparent. This means that, if we were to treat all observations as though they were independent, we would underestimate the actual population variance. Underestimation appears intuitively obvious because one person's behaviour over multiple days is likely to be more similar than would be the behaviour of a number of people on any given day. The question to be addressed, then, is whether or not we can estimate the increase in variance that must be undertaken to compensate for the lack of independence in the observations. It would be this increased variance that would then need to be used in estimating sampling error (increasing it from an assumption of independence) and in estimating the sample size required for a specified sampling error (also increasing it from what one would obtain by assuming independence of observations).

4.1 Estimating Variance

Suppose that an individual is sampled for *D* days, one of which is designated *d*. Suppose further that we sample *n* individuals for *D* days each and an individual is designated by *i*. From this, we measure a behaviour of interest. For the purposes of this example, we will assume that the key variable is person kilometres of travel (PKT) per day. We have, therefore, measured PKT per day with two components, as shown in equation (1):

$$\overline{Y_D} = \overline{y_n} + \overline{y_{iD}} \qquad (1)$$

where:

$$\overline{Y_D} = \text{observed sample mean PKT per person per day (in this case)}$$
$$\overline{\mathrm{y}_\mathrm{n}} = \text{actual mean PKT per person per day}$$
$$\overline{\mathrm{y}_\mathrm{iD}} = \text{mean random day-to-day variation in PKT for individual i}$$

In other words, the sample mean we would estimate if we took each day of observation of each person as an individual observation, and computed the mean PKT from all days and all individuals would be the observed sample mean in the above equation. The mean of the observations for all individuals, aggregated over the days of observation would allow us to estimate the actual mean PKT per person per day. The mean random day-to-day variation in PKT for an individual is not observable and could only be estimated by subtracting the actual mean PKT per day from the observed sample mean.

We assume that each of these components of the overall measurement of PKT has a variance. These variance components are σ_n^2 for y_n and σ_{iD}^2 for y_{iD}. Then, we can state that the variance of the observed mean PKT per person per day is given by equation (2):

$$V(\overline{Y_D}) = \frac{\sigma_n^2}{n} + \frac{\sigma_{iD}^2}{nD} + \text{covariance terms} \qquad (2)$$

The covariances would arise if there is a correlation between the two elements of PKT, such as might be shown if households with high average PKT were also to show higher than average day-to-day variability. For the time being, we will ignore the covariance terms. Again, as we had before, we can estimate the first variance term in equation 2 by estimating the variance from aggregating each person across the days of travel they have reported. We can also estimate the variance of the observed sample mean from the entire data set. We can only estimate the second of the two variances on the right-hand side of equation (2) by subtracting the first term from the left hand side.

Clearly, at one extreme, if there is no day-to-day variation, then the second term on the right hand side of equation (2) becomes zero, and there is no gain to obtaining multi-day data. At the other extreme, if there were no interpersonal variation, but only day to day variability within a person, the first term on the right hand side of the equation would be zero, and there would be no point in measuring more than one individual, although we want as many observations as possible on that one individual. The reality is presumably somewhere between these two extremes.

To see further how these relationships would work, suppose that the day-to-day variability is about one half of the interpersonal variability, i.e., σ_{iD}^2 equals $0.5\sigma_n^2$. In this case, we can write the overall variance as shown in equation (3):

$$V(\overline{Y_D}) = K = \sigma_n^2 + \sigma_{iD}^2 = 1.5\sigma_i^2 \qquad (3)$$

The variance of the observed average PKT per day from the multi-day data is then given by equation (4):

$$V(\overline{Y_D}) = \frac{K}{1.5*n} + \frac{K}{3*n*D} \qquad (4)$$

Suppose, for example, that *D* were to be 15 days. In this case, the variance and standard deviation would be as given in equation (5):

$$V(\overline{Y_D}) = \frac{K}{1.5*n} + \frac{K}{45*n} = \frac{31K}{45*n} = 0.689\frac{K}{n}$$
$$sd(\overline{Y_D}) = 0.83\sqrt{\frac{K}{n}} \qquad (5)$$

In this case, 15 days of data would only reduce the sample size requirements by 17 percent over the situation that would arise with one day of data. On the other hand, suppose that the day-to-day variation has a variance that is equal to the population variance. In this case, equation (3) becomes equation (6):

$$V(\overline{Y_D}) = K = \sigma_n^2 + \sigma_{iD}^2 = 2\sigma_i^2 \qquad (6)$$

and equation (4) becomes equation (7):

$$V(\overline{Y_D}) = \frac{K}{2*n} + \frac{K}{2*n*D} = \frac{K(1+D)}{2nD} \qquad (7)$$

Now, if *D* is set at 15 days, then *(1+D)/2D* is equal to 0.533 and the square root of this is 0.73, leading to a 27 percent reduction in the sample size compared to one-day data.

To put this theory to practical use, we need to know the relationships among the variance components, and also we need to test to see if there is correlation between the random day-to-day variation and the population average. This can be done in the following way.

First, we can determine the variance in PKT by person day from one wave of a panel survey. Using Wave 1 data from a survey conducted in South Australia by the authors (Stopher, Swann, and FitzGerald, 2007), we have 1554 person days of data for which we have PKT per day. The mean PKT per day is 25.57 and the variance of this mean is 1088.17. This represents the observed sample mean PKT per person per day and its variance. Second, if we aggregate to persons, we find that the mean PKT per day is now 24.74 and the variance is 249.55. These are the estimates of y_n and $\&_n^2$ respectively. By subtraction of the latter values from the former values, we can estimate the random day-to-day variation. The average random day-to-day variation in PKT is, therefore, 0.83 kms, but the variance is 838.62. In other words, the day-to-day variation is 3.36 times as great as the variation among persons.

If we assume that there is no covariance between the random day-to-day variations and the person to person variations, then *K* in equation (3) is equal to $4.36\&_n^2$. Hence, equation (4) will now become equation (8):

$$V(\overline{Y_D}) = \frac{K}{4.36*n} + \frac{3.36*K}{4.36*n*D} = \frac{(D+3.36)*K}{4.36*n*D} \qquad (8)$$

The factor by which the standard deviation is now reduced is $\sqrt{[(D+3.36)/4.36D]}$. If D is set to 7 days, then the reduction in the standard deviation is 58 percent and the sample size required is 34 percent of that required for a one-day sample, while if D is set to 15, then the reduction is in the standard deviation is 53 percent and the reduction of the sample size is to 28 percent of that required for one-day data.

It would be useful to see if these relationships hold from a second wave of GPS data. Examining wave 2 of the same GPS panel, we find the following statistics for PKT still: mean PKT per day based on 1,986 person days of data is 30.41 kms, with a variance of 3999.04. Aggregated to the 73 persons, the mean is 24.95 kms and the variance is 217.96. In this case there is a major increase in the day-to-day variation because of a holiday period that was included for some participants. We can deduce from this that the day-to-day variation in the mean PKT has a mean of 5.46 kms and a variance of 3,781.08. In this case, the day-to-day variation is 17.348 times the person-to-person variance. The factor of interest is now $(D+17.348)/18.348D$. This would equal 0.19 for 7 days, indicating that the sample size for this case would be just 19 percent of that required for one-day data and that the standard deviation of the person-to-person variability is 43.5 percent of the total measured standard deviation. For 15 days of data collection, the sample size is only 11.8 percent of the sample size required for one-day data, and the standard deviation is 34 percent of the measured multiday standard deviation.

In this latter case, it would probably be safer to exclude the extreme distance days from the data before computing the statistics, since these generate an unusual situation of increasing the day-to-day variability in the person level data. This results from about 5 individuals who had anything from around 600 kms of travel to well over 1100 kms of travel on two or three days of the collection period. While such outliers are genuine data points that should be included in most analysis, they should be removed to give a more conservative estimate of the sample sizes required. With the outliers removed, the mean and variance based on 1968 person days of data are 25.78 kms and 1322.36. Using these figures instead of the earlier ones, the day-to-day variability has a mean of 0.83 kms, and a variance of 1104.4, which is 5.07 times the person-based variance. This is not dissimilar to the result from the first wave, although it is a little higher. One might conclude that a reasonable figure to use, based on these statistics would be that the day-to-day variability is about 4.75 times the person-to-person variability. This means that multi-day data would result in significant gains, with the sample size required being about 35 percent of the one-day sample size for a 7-day survey, and about 28 percent for a 15-day survey.

Before leaving this, we need to consider the issue of the covariance terms. These would be determined through examining the data first to see if there is a significantly high correlation between average PKT and day-to-day variability in PKT. We found that the correlations are significant although with an R value of around 0.73 or an R^2 of about 0.5. We have therefore not proceeded at this time to estimate any covariance terms, but this will be examined in further research.

4.2 Sample Sizes

With respect to sample sizes, this will largely depend on the use to which the data are to be put. However, we will consider some specific examples to illustrate, and will assume that the reduction in sample size estimated in the previous section applies, i.e., that a 7-day survey using GPS would reduce the sample size to 35 percent of a one-day survey, and that a 15-day GPS survey would reduce the needed sample size to 28 percent of the one-day survey sample size.

Suppose that the data are to be used for modelling purposes. In many instances, a minimum sample size of 3,000 households is often specified for modelling purposes, and there may be some specific subsample requirements, such as for a given number of public transport users in the sample. Using a conventional diary survey for one day, 3,000 households would be expected to comprise data from about 6,900 persons, making about 27,000 trips. On average, for example, this would result in having about 3,000 work trips. In an average Australian setting, this might also include about 300 public transport trips. Based on the analysis performed in the previous section, a 7-day GPS survey would require a sample of roughly 1,000 households, while a 15-day GPS would require a sample of 850 households. The former of these would be likely, at the placement rates we have experienced, to involve 2,100 people carrying GPS for 7 days, producing roughly 12,000 days of data, or about 48,000 trips. The 15-day GPS survey would produce about 1,800 people making approximately 72,000 trips. In both cases, significantly more trips would be available for analysis, but this is required because of the smaller sample size of households.

A second example is provided by the situation in which one wishes to evaluate a travel behaviour change project. Stopher and Montes (2007) have previously estimated that data would be required from approximately 450 households in two waves of a panel to determine that a change of ±0.5 kms was significant at 95 percent confidence. This was based on using just the household variances and covariances, averaged from 28 days of data. Assuming that the variances and covariances were representative of the results from a single day in each of two waves, this would lead to a sample size of 158 households for 7-day data and 126 households for 15-day data. At this level of sample size, the issue of sample size becomes one that is more of a policy nature than statistical. In other words, those who have to make decisions based on the results from the surveys may desire a higher sample size than is actually required for statistical validity. This is probably true even for the sample sizes for a household travel survey.

5 Cost Implications

So far, there is little question that the execution of a GPS survey is more expensive per household than is the execution of a one-day diary survey, by comparable methods. The reason for the higher cost is partly the need for more expensive means of delivery and collection of GPS devices than would normally be used for diaries, and is partly because of the extensive data processing that is required from the large quantities of data produced by the GPS devices. As an example, we will consider the case of a household travel survey that is to be conducted by telephone recruitment, followed by posting out the surveys and having them returned by post. On the average, such a survey would cost on the order of $175 per completed household, assuming that there are about four reminders used to attempt to retrieve data from the households. Included in this cost is the average non-response rate for such a survey and also processing of the resulting data to the point of being able to present trip summaries that would be usable in travel demand models.

For a 15-day survey of households, with an average of 2.2 persons per household undertaking the survey, we estimate that the costs would be around $500 per household, based on the same methodology of a telephone recruitment followed by couriering the devices to the household and having them picked up again by courier. One of the major reasons for the higher cost is that, although we have automated the processing of the huge amount of GPS data that would be collected in a 15-day survey, we still find it necessary to undertake a visual check of the results of the processing for each day of data from each device. However, if we compare the costs of the situation for a modelling exercise, in which conventional wisdom would require a sample of at least 3,000 households, this conventional survey would cost approximately $525,000. As noted in the previous section of this paper, the GPS survey would require a sample of 850 households undertaking a 15-day GPS survey, which would cost $425,000. Therefore, the GPS survey is actually a less expensive survey, based on this sample size, and produces almost three times the number of person trips of the conventional diary survey for that lower cost.

For other survey methods, we would expect a smaller increment of cost for the GPS survey per household compared to the conventional survey method. The reason for this is that the cost of the visual editing will remain fixed and other costs that would be added will not increase in the same proportion. For example, a face-to-face survey of households currently in Australia costs in excess of $350 per completed household. This is more than double the cost of the telephone recruitment and postal retrieval survey. Assuming that approximately $200 per household of that cost is for interviewer time, over the cost of the telephone recruitment and postal return, then this same amount should be added to the cost of the GPS survey performed by face-to-face visits, while the cost of couriering devices should be deducted from the GPS cost. We believe that this would lead to a net increase in the costs of the GPS survey of about $180 over the telephone recruitment method, so that the comparative costs would now be, say, $350 for the conventional survey and $680 for the GPS survey. Comparing costs on a 3,000 sample size for modelling purposes for the conventional diary survey to those of an 850 household 15-day GPS survey, we have costs of $1,050,000 for the conventional survey against $578,000 for the GPS survey. Indeed, if for other than statistical reasons, it was decided to increase the sample for the GPS survey to, say, 1,500 households, rather than the 850 required for statistical purposes, the cost would still be less than the conventional survey, at a total of $1,020,000. However, this sample would produce about 198,000 trips, compared to the 27,000 from the conventional survey.

6 Conclusions

In this paper, we have outlined reasons why multi-day surveys of travel behaviour have been the exception rather than the rule in past transport studies. We have also shown that this is the case, despite the overwhelming attraction of multiday surveys to explain much more about person travel behaviour than can be obtained from examining one day of travel. We have described the potentials that are now offered by recent technological developments, especially in the area of personal GPS devices.

We have addressed the issue that has largely been unexplored in previous work in transport relating to the sample size implications of collecting multi-day data. We have shown that, given actual data results from multi-day data, the potential is for such data to reduce the sample size requirements by a very substantial amount. For a 7-day GPS survey, the sample size would be reduced by about 65 percent from that of a conventional one-day diary survey, while a 15-day GPS survey, which our experience has suggested would be optimal for a number of reasons, would reduce the conventional sample size by over 70 percent. Further, we have shown that, while GPS surveys are still quite a bit more expensive to conduct on a per household basis than a conventional one-day survey using a diary method, the reduction in sample size required results in significant cost savings if a GPS survey is used with the sample size that is necessary statistically, and that, even a substantial increase in sample size can be obtained without exceeding the costs of a conventional survey.

7 Acknowledgements

The authors would like to acknowledge the University of Sydney for providing the funds to undertake this research and Professor Kara Kockleman of the University of Texas at Austin, who provided guidance and comment on the variance estimation procedures described in this paper. All errors in this paper remain the responsibility of the authors.

8 References

Axhausen, K.W. (1995). *Travel Diaries: An Annotated Catalogue 2nd Edition*, Draft Working Paper, Institute für Straßenbau und Verkehrsplannung, Leopold-Franzens-Universität, Innsbruck, Austria.

Axhausen, K.W., A. Zimmermann, S. Schonfelder, G. Rindsfuser, T. Haupt (2002).Observing the rhythms of daily life: A six-week travel diary. *Transportation* **29**, 95-124.

Goldenberg, L., C. Stecher and K. Cervenka (1995). "Choosing a Household-Based Survey Method: Results of the Dallas-Fort Worth Pretest," paper presented to Fifth National Conference on Transportation Planning Methods Applications, Seattle, Washington, April 17-21.

Hanson, S. and J. Huff (1988) Systematic variability in repetitious travel. *Transportation* 15, 111-135.

Parvataneni, R., P.R. Stopher, and C. Brown (1982). Origin-Destination Travel Survey for Southeast Michigan, *Transportation Research Record No. 886*, pp. 1-8.

Pas, E.I. (1986) Multiday Samples, Parameter Estimation Precision and Data Collection Costs for Least Squares Regression Trip-Generation Models. *Environment and Planning A*, 18, 73-87.

Pas, E.I. and S. Sundar (1995) Intra-personal variability in daily urban travel behavior: Some additional evidence. *Transportation 22*, pp. 135-150

Pendyala (2003) Measuring Day-to-Day Variability in Travel Behavior Using GPS Data. Available at http://www.fhwa.dot.gov/ohim/gps/travel_char.htm. Accessed 13/06/07.

Richardson, A.J., (2003) Temporal Variability of Car Usage as an Input to the Design of Before & After Surveys, Paper presented at the *82nd Annual Meeting of the Transportation Research Board*, Washington, D.C.

Stopher, P.R. (1992). "Use of an Activity-Based Diary to Collect Household Travel Data," *Transportation*, Vol. 19, pp.159-176.

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

Stopher, P.R., E. Clifford, J. Zhang, and C. FitzGerald (2007). Deducing Mode and Purpose from GPS Data, paper presented to the *TRB Conference on Transportation Planning Applications,* Daytona Beach, Florida, May.

Stopher, P.R., N. Swann, and C. FitzGerald (2007). Using and Odometer and a GPS Panel to Evaluate Travel Behaviour Changes, paper presented to the *TRB Conference on Transportation Planning Applications,* Daytona Beach, Florida, May.

Stopher, P.R., E. Clifford, and M. Montes (2007). Further Analysis of Multi-Day Data, paper to be presented to the 30th ATRF Meeting, Melbourne, Victoria, September.

Stopher, P.R. and M. A. Montes (2007). Pilot Testing of a GPS Survey for the Long-Range Monitoring Procedure for Voluntary Travel Behaviour Change, Final Report to the NTBCP partners, Sydney, March.

Wolf, J. (2006). Application of New Technologies in Travel Surveys, in *Travel Survey Methods -- Quality and Future Directions*, P. Stopher and C. Stecher (editors), Elsevier, Oxford.