Reducing Burden and Sample Sizes in Multi-day Household Travel Surveys

Peter R. Stopher, Professor of Transport Planning (Corresponding Author)
The Institute of Transport and Logistics Studies
The University of Sydney, NSW 2006, Australia.
Telephone: +61 (0)2 9351 0010
Fax: +61 (0)2 9351 0088
peters@itls.usyd.edu.au

Kara Kockelman, Associate Professor of Transportation Engineering
Department of Civil, Architectural and Environmental Engineering,
University of Texas, Austin, Texas, USA

Stephen P. Greaves, Senior Lecturer
The Institute of Transport and Logistics Studies
The University of Sydney, NSW 2006, Australia.

and

Eoin Clifford, Research Associate
The Institute of Transport and Logistics Studies,
The University of Sydney, NSW 2006, Australia.

The following paper is a pre-print and the final publication can be found in
Transportation Research Record No. 2064:12-18, 2008.
Presentation at the 87th Annual Meeting of the Transportation Research Board,
January 2008

ABSTRACT

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 travel on an average weekday, evidence suggests this may not be the most cost-effective way to collect the data because day-to-day travel variability is substantial, requiring larger sample sizes. In addition, collecting multi-day data provides a richness of information that simply cannot be captured with a 1-day survey – offering insights into, for example, differences in weekday versus weekend travel, the longer-term travel impacts of flexible work hours, and trip substitution and cycling patterns that emerge over the course of a week or more. Despite the intuitive appeal of multiday surveys, there are few examples and little information on sampling issues and sample size requirements. With this in mind, this paper explores the reasons given for not doing multiday surveys (which centre around respondent fatigue), why these issues are fast becoming irrelevant (through the use of new passive data recording technologies), the sample size implications of extending a survey, and the potential for estimator efficiency and cost savings when conducting multiday surveys (even accounting for the cost of new technologies). Using GPS-based travel distance data from Adelaide, Australia, we find that the reductions in sample size are large, and that collecting multi-day data is feasible - offering a richness not available in one-day data, along with cost-effective gains in estimator efficiency. *(248 words)*

INTRODUCTION

Since their widespread adoption, several decades ago, most household travel surveys have used a diary format that is administered prospectively, either through a telephone recruitment and postal (mail out-mail back) diary survey, a face-to-face interview at the household's door, or an entirely postal survey. Most have been trip diaries, in which respondents are asked to report each trip that they undertake in a day. In most cases, a trip is defined as being the travel from an origin to a destination, without intermediate stops (except for changes in travel mode or for traffic-related stops, such as traffic signals) *[1]*. In other cases, the diaries are designed to collect details about each trip segment, where a segment is defined as that part of a trip that is carried out on a single mode of travel *[2]*. In the former case, the diary requires the reporting of what generally averages 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 results in average segment counts of perhaps twelve or more, especially when including walk segments at the start and end of all car trips.

Nearly two decades ago, the activity diary appeared in some travel surveys *[3]*. Concentrating on out-of-home activities, the reporting task required details on an average of 8 to 10 activities per day per person. Still later, a time-use diary was introduced *[4]*, increasing reportable events to 15 or more, on average. As can be readily appreciated, the increasing desire for information on activities or time use increased respondent burden.

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 remain 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 Focus*	*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
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 and 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	Unpublished	Phone/CATI	Time Use	2
Michigan	2005-6	15,000	Phone/CATI	Travel	2

Note: Data sets are listed in chronological order.

The vast majority of household travel surveys, especially those conducted for mainstream modelling and planning activities, are still one-day diary surveys. One of the major reasons for the concentration on one-day surveys has been the perception that respondent burden is already about as high as most survey professionals would wish to impose, and that extensions to two or more days will substantially lower response rates, and substantially increase respondent burden. Indeed, in a recent two-day postal diary survey *[5]* in New South Wales, Australia, it was found that respondents often stated that they did not travel anywhere on the second day as a means to decrease the response burden. This caused the reported rate of non-mobility to jump by a factor of roughly 10 between days one and two. Recently, the State of Michigan undertook a two-day survey, while the Detroit Metropolitan Planning Organization opted to use a one-day diary in their add-on sample to the statewide survey, mainly as a result of their concerns with the drop off in reporting on the second day. 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.

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 respondent burden and the associated tendency towards less accurate reporting as survey duration increases, any method that moderates self-reporting effort may facilitate the successful completion of multi-day surveys. Before moving into that discussion, however, it is useful to consider, and quantify, the potential merit in multi-day surveys. To do this, we examine some travel-distance results that have been obtained from self-report based multi-day surveys, such as those listed in the previous section.

Day-to-day Variability in Travel

One key driver of interest in multi-day surveys is sample size reductions that may result from a recognition of day-to-day variability in individuals' travel choices. While much of what we do is based on routine, there is considerable variability in travel behaviour. For instance, Pas and Sundar *[6]* analysed variations in trip rates, trip chaining, and daily travel times from three-day Seattle data. They found intra-person variability[1] in trip rates to be 38 percent of total variation (across all person-day trip counts), as compared to 50 percent in the 5-day Reading, UK data set, something they attributed to the longer reporting period (five days for the Reading data set). In a more recent study, Pendyala *[7]* compared intra-personal variability in trip rates and travel times with these earlier studies using GPS data collected for the Lexington, Kentucky pilot study. He reported that intra-personal variability in trip rates for the three-day (weekdays only) sample was 49 percent *[6]*. In 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 that it is better able to measure infrequent and irregular behaviours (including short trips) that tend to be missed in self-reported diary surveys.

To our knowledge, the longest duration survey completed is the six week MobiDrive survey *[8]*, which was completed in 1999 in the German cities of Halle and Karlsrhue. The extended nature of the survey enables unique analyses of behavioural variations. For instance, Richardson *[9]* analysed relative variation in a number of measures of travel behaviour, a summary of which are provided in Table 2. While his work does not highlight the impacts of extending the period to two, three, four, or more days, it is nevertheless interesting to note the reduction in

[1] Intra-person variability refers to variance in trip counts around individual means (within sum of squares), and is 38% of the total variance (total sum of squares) across all single-day recorded trip counts.

variability that occurs from (in effect) sampling people on the same day of the week for six weeks, rather than sampling them for one week.

Table 2: Day-to-Day Variability from the MobiDrive Data (adapted from Richardson 2003)

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%

Note: *CV* is the coefficient of variation, or the ratio of standard deviation to mean value.

The only other analysis of this type we have found is from the five-week 1971 Uppsala survey, where Hanson and Huff *[10]* found the average number of home-to-home journeys and stops for one selected week were (nearly) identical to those from the five-week period. Their findings underscore the presence of behavioural cycling and regularities across week-long time frames.

Sample Size Implications

While many studies of inter-day variability exist, few have dealt with sample size impacts of multi-day approaches. Savings in sample size and survey costs may be significant, even if one allows for the additional expense and attrition attributed to additional days of data collection. For instance, using the Reading data, Pas *[11]* showed how 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.

Clearly, there is potential value in collecting multi-day data to permit analysis of day-to-day behavioural patterns while reducing sample sizes and survey costs (and/or enhancing estimator precision). These reasons justify interest in procedures that facilitate multi-day surveys, particular where the roles of respondent burden and reporting fatigue are mitigated.

Technological Developments

The use of Global Positioning System (GPS) devices to track people's travel has emerged as a competitive method for collecting household travel survey data. A brief history of the use of GPS devices in travel monitoring can be found elsewhere *[e.g., 12, 13]*. Suffice it to say that GPS devices for monitoring travel have progressed rapidly in the past decade to the point where a personal, easy-to-carry GPS device is now available. This device can store weeks of data yet weighs less than 50 grams and is smaller than a mobile telephone. Using a device that predates this latest technologies, Stopher et al. *[14, 16]* found that it is quite feasible to ask people to carry such devices with them for periods as long as one month. While longer periods have not been tried, there is no obvious reason why such persons 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 personal (self and household) characteristics, 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 respondents simply carry the device throughout the requested study period, and are asked 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 the location and regular purposes of personal travel. In Stopher et al.'s *[14, 16]* use of these devices, they have, so far, found relatively few objections related to invasion of privacy (about 2 to 5 percent), perhaps because there is nothing displayed on the devices that indicates the data are being recorded and also because the data are not collected from the individual in real time. Thus, although the respondent carries the GPS device 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.

In such work, response rates have been quite insensitive to the length of time for which respondents are asked to carry the devices. In work undertaken in South Australia, there is little evidence of a difference in either recruitment rates or the failure-to-comply rates among recruited households asked to carry GPS devices for periods from one week to as many as four weeks *[14]*. Beyond MobiDrive, researchers have not experimented with periods longer than four weeks. In an initial panel asked to undertake four weeks of data collection, Stopher et al. *[14]* 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. 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 offers 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 already attempted successfully *[14]*.

With the exception of the 1991 MobiDrive project and the much earlier 1971 Uppsala Project, 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, Stopher et al. *[15]* have documented much about the variability of personal travel, as derived from their 28 days of data. In this paper, to the focus likes on issues of sample size reductions that could be achieved through a multi-day survey, while maintaining estimator efficiency.

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 study unit is generally an individual person, whilst the sampling unit is the household. Within a household, travel decisions are not usually made independently, so the assumption of independence breaks down at the person level. And, of course, multiple trips by a single individual exhibit significant correlations. This means that sampling errors are routinely underestimated when calculated at the level of a person or a trip.[2]

When collecting multiple days of data from the same individual, the issue of independence of the observations is clearly apparent. If one were to treat all observations as though they are independent, one would underestimate the actual population variance. Underestimation appears intuitively obvious because one person's behaviour (e.g., trip rates or miles travelled) over multiple days is likely to be more similar than would be the behaviour of a number of people on any given day. The true variance needs to be understood, in order to estimate true sampling error , as well sample sizes required to achieve desired levels of estimator precision.

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 for each individual on each day with two (latent) components, as shown in equation (1):

$$y_{id} = \mu_i + \varepsilon_{id} = \mu + \delta_i + \varepsilon_{id} \tag{1}$$

where y_{id} is individual *i*'s PKT on day *d*, μ_i is person *i*'s true mean distance of travel, ε_{id} is the random error term that accompanies a single day's data point, μ is the population's mean travel distance, and δ_i is the difference between individual *i*'s mean travel distance and that of the population (serving as a fixed or random effect). Of course, with repeat observations of various individuals' behaviour, one can estimate each individual's mean PKT, his/her distance from the population mean, and the variance of his/her PKT, as follows:

$$\hat{\mu}_i = \sum_{d=1:D} y_{id} \Big/ D \tag{2}$$

$$\hat{\delta}_i = \sum_{i=1:N} \sum_{d=1:D} y_{id} \Big/ ND \tag{3}$$

$$\hat{\sigma}_i^2 = \sum_{d=1:D} \left(y_{id} - \hat{\mu}_i\right)^2 \Big/ (D-1) \tag{4}$$

where σ_i^2 is the variance of ε_{id}.

One also can estimate the variation in travel distances across all persons and all days, which is equivalent to summing the across-persons and across-days variations. The concepts are equivalent to an analysis of variance, where each respondent is a "treatment", and one has data for computing both between and within sums of squares (BSS and WSS). In other words, total variation in the data is the result of inter-personal plus inter-temporal variations, as follows:

[2] It also is possible to overestimate sampling errors, when behaviors are negatively correlated across trips by a single individual (e.g., start times) or across members of the same household (e.g., use of a particular vehicle).

$$\hat{\sigma}_{y_{id}}{}^2 = \hat{V}(Y_{id}) \approx \frac{TSS}{ND}$$
$$= \frac{\sum_{i=1:N}\sum_{d=1:D}(y_{id} - \hat{\mu})^2}{ND} = \frac{\sum_{i=1:N}\sum_{d=1:D}(y_{id} - \hat{\mu}_i)^2}{ND} + \frac{\sum_{i=1:N}(\hat{\mu}_i - \hat{\mu})^2}{N} = \frac{WSS + BSS}{ND} \tag{5}$$

where *TSS* stands for total sum of squares and $\hat{\mu}$ is the sample's "grand average". This relationship is shown as approximate simply because an unbiased estimate of $\hat{\sigma}_{y_{id}}{}^2$ requires some accounting for losses in degrees of freedom (by subtracting off 1 and N in various denominator terms). The use of sample estimates in the last line of the equation implies that such adjustments are necessary, and these can become significant when sample periods are short (e.g., $D = 3$ days).

Equation (5) essentially implies that total variation in the data can be decomposed into variation around each traveller's average mean PKT and variation of these mean PKT values around the population mean. Assuming that all individuals exhibit the same variation in their day-to-day idiosyncratic effects (i.e., $\sigma_i^2 = \sigma_j^2 = \sigma_\varepsilon^2$ for all i,j), and recognizing that increases in sample size and sample duration reduce variation in final estimates, one has the following:

$$V(\hat{\mu}) = \frac{\sigma_\varepsilon^2}{ND} + \frac{\sigma_{i|\mu}^2}{N} \tag{6}$$

where $\sigma_{i|\mu}^2$ is the variation of each traveller's mean distance value around the population's grand mean.

It should be noted that Equation (6) will include some covariance terms if there is heteroskedasticity in idiosyncratic effects (i.e., $\sigma_i^2 \neq \sigma_j^2$), and correlation exists between an individual's mean and variance, such as can occur when longer-distance travelers exhibit higher variability in their day to day PKTs. For the time being, we ignore such correlations, and their associated covariance terms.

Rather clearly, one can estimate all terms in Equation (6) by simply computing sample variance across all travellers' recorded PKT values, averaging sample variances for PKT for all individuals, and computing sample variance across all travellers' average PKT values.

Clearly, at one extreme, if there is no day-to-day variation, then the first term on the right hand side of equation (6) becomes zero, and there is no gain to obtaining multi-day data. At the other extreme, if there is no inter-traveler variation in mean distances, but only day-to-day variability within a person, the second term will be zero, and there will be no point in observing more than one individual, although analysts would still 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; in other words: σ_ε^2 equals $0.5\,\sigma_{i|\mu}^2$. Moreover, define the constant "*K*" as shown here:

$$\sigma_\varepsilon^2 + \sigma_{i|\mu}^2 = K \tag{7}$$

Given these assumptions, the variance of the daily PKT across all data points can be written as follows (8):

$$V(\hat{\mu}) = \frac{K}{3ND} + \frac{K}{1.5N} \tag{8}$$

Suppose, for example, that D were to be 15 days. In this case, the variance and standard deviation would be given by equation (9):

$$\begin{aligned} V(\hat{\mu}) &= \frac{K}{45N} + \frac{K}{1.5N} = 0.689\frac{K}{N} \\ SD(\hat{\mu}) &= 0.830\sqrt{\frac{K}{N}} \end{aligned} \tag{9}$$

In this particular case, 15 days of data would reduce the sample size requirements by 17 percent over the situation of one day's data. However, if day-to-day variation equals inter-personal variation in mean PKT, Equation (7) now implies the following:

$$V(\hat{\mu}) = \frac{K}{2ND} + \frac{K}{2N} = \frac{K(1+D)}{2ND} \tag{10}$$

Now, if D is still 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 sample size, as compared to one-day data.

To put this theory to practical use, one needs to know the factors at play among the variance components, as well as test for the presence of correlation between μ_i and σ_i^2. This can be done in the following way.

First, we determine the variance in PKT by person day from one wave of a panel survey. Using Wave 1 data from Stopher et al.'s *[14]* GPS-based survey in South Australia, we have 1,554 person-days of data (with D ranging from 15 to 30 across the 79 respondents). The average daily PKT is 25.57 km and the variance in daily recorded values is 1088.17 km^2 (essentially $TSS/[ND\text{-}1]$). Second, in looking at average PKT values per respondent, we find that the sample variance across these person-level averages is 249.55 km^2 (or $BSS/[ND\text{-}1]$), so our estimate of σ_ε^2 (or $WSS/[ND\text{-}1]$) is therefore 838.62 km^2. In other words, the day-to-day variation in PKT 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 individuals' mean PKT values, then K equals $4.36\sigma_\varepsilon^2$. Hence, equation (7) will now imply the following:

$$V(\hat{\mu}) = \frac{3.36K}{4.36ND} + \frac{K}{4.36N} = \frac{K(3.36+D)}{4.36ND} \tag{11}$$

The factor by which standard deviation is now reduced is $\sqrt{(D+3.36)/4.36D}$, essentially implying that sample sizes can fall to just $(D+3.36)/4.36D$ of the 1-day sample size without increasing the variance of one's estimate of mean PKT, thanks to $D > 1$. Thus, if D is just 7 days, the expected reduction in the standard deviation of the estimate of mean PKT is a significant 58 percent (versus a one-day sample of same size)., If one wishes to achieve the same variance as before, one's sample size requirements essentially fall to just 34 percent of that required for a one-day sample. And, if D is 15, sample size needs are reduced by a striking 72 percent, to just 28 percent of the original sample size (assuming a one-day original sample).

It is useful to see if these relationships hold in a second wave of GPS data. In wave 2 of the same GPS panel, mean PKT per day (based on 1,986 person days of data for 73 persons) is

30.41 km, with a variance of 3999.04 km^2. Day-to-day variation within persons is estimated to be 3,781.08, making the estimate of day-to-day variation fully 17.348 times inter-personal variance.[3] The factor of interest is now (D+17.348)/18.348D, or 0.19 for 7 days, indicating that sample size for this case (assuming estimates of variance are true, population values) would be just 19 percent of that required for a one-day data set. For 15 days of data collection, these numbers would suggest needing a sample size that is only 11.8 percent of the one-day sample size.

These are striking reductions, and it must be noted that this Wave 2 case includes a major holiday period for several participants, resulting in over 600 km of travel on each of two or more days for 5 individuals With these five outliers removed, the mean PKT falls to 25.78 km and total variation falls to 1322.36 km^2. Inter-day variations average 1104.4 km^2, making these 5.07 times the inter-personal variation. This is not dissimilar to the result from the first wave, although it is a little higher. Based on these results, one might conclude that a reasonable assumption is that day-to-day variability is about 4.75 times person-to-person variability. This means that multi-day data can result in significant sample size reductions (65% for a 7-day survey and 72% for a 15-day survey) and potential cost savings. Interestingly, in the limit (as D goes to infinity), the marginal impacts of added survey duration disappear, so sample size reduction opportunities top out at 79% (or 21% of original sample size, in the case of a 4.75 factor).

Another point of interest, of course, is the issue of the neglected covariance terms. These can be determined by examining correlations between average PKT levels (per respondent) and day-to-day variability in PKT (per respondent). In the Australian data set described above, the correlations are significant, at roughly +0.73 (or an R^2 of about 0.5). The sample size impacts of such correlations should be the focus of further research.

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 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.

[3]

A second example is provided by the situation in which one wishes to evaluate a travel behaviour change project. Stopher and Montes *[16]* 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 km 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.

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 (in Australian dollars) 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. Moreover, this sample would produce about 198,000 trips, compared to the 27,000 from the conventional survey. In effect, then, thoughtful use of GPS devices may save agencies on the order of 50% of travel survey budgets while offering equivalent estimator efficiency gains (in terms of PKT estimates, anyhow) and providing far more detailed data on individuals and their trip-making (including the presence of habit and other interesting dynamics in day to day travel behaviors).

CONCLUSIONS AND EXTENSIONS

This paper describes why multi-day surveys of travel behaviour have been the exception rather than the rule in past transport studies -- despite the wonderful ability of multiday surveys to illuminate much more about person travel behaviour than can be obtained from examining one day of travel. This paper describes the capabilities now offered by recent technological developments, especially in the area of personal GPS devices while addressing the largely neglected issue of sample size implications when collecting multi-day data. Using GPS-based PKT data, this paper demonstrates how 7-day survey data may well diminish needed sample sizes by 65 percent (relative to a conventional, one-day diary survey), and a 15-day duration (which may be optimal for a number of reasons) would reduce sample sizes by over 70 percent, in the case of a focus on PKT. Further, we argue 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 the required sample size 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.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the University of Sydney for providing the funds to undertake this research. All errors in this paper remain the responsibility of the authors.

REFERENCES

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

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

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

4. 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.

5. 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.

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

7. 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.

8. 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.

9. 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.

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

11. 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.

12. 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.

13. 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.

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

15. 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.

16. 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.