Evaluating changes associated with workplace and school travel plans – something old, something borrowed, something new

Charles Sullivan\textsuperscript{1} and Anna Percy\textsuperscript{2}

\textsuperscript{1} Capital Research, Wellington, New Zealand
\textsuperscript{2} Auckland Regional Transport Authority, New Zealand

1 Introduction

Travel plans can be part of travel demand management efforts to reduce problems associated with urban traffic growth. They have been widely used in North America and Europe (e.g. Rye, 2002), and more recently Australasia (e.g. Meiklejohn and Wake, 2007).

In New Zealand, the Auckland Regional Transport Authority (ARTA) has taken the lead in implementing travel plans. As seen in Table 1, by early 2008, ARTA’s TravelWise programme included 178 schools, 31 workplace travel plans covering over 15,000 staff, and 2 tertiary institutes with 45,000 staff and students (excluding the more experimental site and area travel plans).

<table>
<thead>
<tr>
<th>School year</th>
<th>Number of TravelWise schools</th>
<th>Total students at TravelWise schools</th>
<th>Walking school buses</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005 (Nov.)</td>
<td>51</td>
<td>31,941</td>
<td>179</td>
</tr>
<tr>
<td>2006 (Nov.)</td>
<td>92</td>
<td>57,199</td>
<td>200</td>
</tr>
<tr>
<td>2007 (July)</td>
<td>138</td>
<td>75,000</td>
<td>218</td>
</tr>
<tr>
<td>2008 (April)</td>
<td>178</td>
<td>98,000</td>
<td>245</td>
</tr>
</tbody>
</table>

Note: Not all walking school buses are at TravelWise schools.

Internationally, figures as to the impact of workplace and school travel plans are commonly mentioned. For example, a literature review found about 50 studies with an average diversion rate for car as driver with workplace travel plans of 7.2 percentage points (Maunsell \textit{et al.}, 2004). But when one has to answer practical questions about measuring impact in programmes as large as ARTA’s (not to mention extensions to other cities), detailed advice on good practice about data collection and analysis is remarkably difficult to find (beyond examples of simple questionnaires and tips at Stats 101 level). The technical reports available on the Western Australian Department for Planning and Infrastructure website\textsuperscript{1} are a welcome exception.

In particular, despite the greater reason in recent years to assess impact in terms of reductions in kilometres travelled by mode (so as to estimate impact on greenhouse gas emissions), good guidance on how to estimate kilometres travelled for participants in travel plans is exceptionally hard to find. Indeed, the \textit{Smarter Choices} report in the UK identified ‘a number of quite serious measurement problems’ with monitoring of travel behaviour change initiatives in the UK, and the first problem they then examined in detail relates to kilometres travelled:

\textsuperscript{1} www.dpi.wa.gov.au/travelsmart/14974.asp
 Many of the policy objectives, reviews and individual studies have expressed their targets and outcomes in the form of shifts in the proportion of trips by each mode, then converting these into traffic impacts by re-calculating the figures as vehicle kilometres. However, they have done so in different ways, in some cases simply by assuming that all origins, destinations and average journey distances stay the same, others by allowing for differential impact on journeys of different lengths, and a few by allowing for an effect on wider dimensions of choice such as destination, number of journeys, time of day, and patterns of trip tours at the household level. (Cairns et al., 2005, p. 369)

In a paper presented to the 2007 Australasian Transport Research Forum, Kota and Percy (2007) focused on estimating emissions (via kilometres travelled). Here we focus more narrowly on the underlying basics of estimating mode shift and change in kilometres travelled.

This paper contributes to building good practice guidelines for evaluating travel plans, especially where data on changes in kilometres travelled is an objective (e.g. to estimate changes in emissions from the changes in kilometres travelled) and before/after surveys are used for data collection. The lessons here include:

- **something old** (reminders of evaluation fundamentals that can easily get overlooked)
- **something borrowed** (relevant points from methods for evaluating personalised marketing that have been discussed in great detail at previous ATRFs)
- **something new** (innovations for collecting data about kilometres travelled).

## 2 Some evaluation fundamentals

### 2.1 Beware: Surveys that are good enough to help establish travel plans are not necessarily good enough for monitoring their impact

Travel plans often include a baseline survey which is helpful to estimate baseline mode share and to provide input useful for travel planning. These include gathering feedback about current transport problems perceived, and/or collect reactions to possible initiatives such as car pooling scheme, fleet cycles. It is relatively easy to design and implement such a survey and get helpful results.

But the ease of creating such baseline surveys can be a deceptive trap:

- There may be an illusion that impact can be measured later even more easily (by just repeating a few key questions from the baseline survey).
- Measuring change well requires higher data quality than is required for the immediate uses of baseline surveys for designing travel plans.

This trap arises because different objectives can require different levels of data quality. For example, in a baseline survey, inaccuracies of a few percentage points (e.g. describing car mode share as 65% when it is really 60%) will often be of little practical significance in relation to the objective of designing a travel plan. However, a year or two later in a follow-up survey, the very same inaccuracy of 5% may be fatal to the objective of detecting impact (because realistically the short-term impact of many travel plans may be only a change of a few percentage points in car mode share). Small changes of around this size are clearly of interest in Australasia. For example, for driving the Victorian standard diversion rate for workplaces is 6.7% (Maunsell AECOM, 2006, p. 34) and the New Zealand default diversion rate (‘medium’) is 5.0% (Maunsell et al., 2004).
Unfortunately, without good statistical planning at the very beginning, travel planners are most likely to notice the need for higher data quality/precision at the time of the follow-up survey. By then, of course, it is too late: baseline and follow-up surveys have to be done in a comparable way so as to accurately estimate impact.

A logical implication of noting the difficulties when different objectives (i.e. monitoring change versus input into travel planning) are handled by one survey is to have two separate surveys rather than one. Splitting into two surveys was recently recommended to TravelSmart in Western Australia (SKM, 2007).

We expected to find guidance for data collection and data quality from the UK because of the large numbers of travel plans being done there. But we were sharply disappointed. For example, the annual spend by on school travel plans of around £13 million by Transport for London alone seemed likely to justify establishing sound evaluation. However, last year the politicians of the London Assembly were clearly underwhelmed by data quality:

*We found that the data collected by Transport for London on school travel plans were not sufficiently comprehensive to give us a clear answer to these questions. The figures that were available suggested that far from being ahead of the rest of the country in terms of the proportion of school with approved plans, London was lagging slightly behind. Furthermore, there appears to be little detailed research to back up Transport for London’s claims that school travel plans have led to a 5.5% reduction in single occupancy car trips.*

(London Assembly, 2007, pp. 1-2)

Nor do they believe the problem is specific to London:

*The problems of data collection and the paucity of evidence available to assess the effectiveness of school travel plans are not confined to London. Researchers commissioned by the Department for Transport found that the data that were available for the work were “generally of poor quality and coverage”.*

(London Assembly, 2007, p. 8)

Given that the difficulties obtaining fairly precise results from surveys are probably greater than many expect beforehand, it is important for organisations to consider alternatives that may be cheaper or more precise. Depending on the sites for travel plans and the scale of the intervention, alternative indicators such as direct measurement (e.g. public transport sales or boardings, traffic counts, cycle counts) may be preferable.

### 2.2 Accurately measuring impact requires confidence that bias is tolerable

Major things that affect data quality and precision include fundamentals of statistics (e.g. bias and sampling error). Travel plan surveys require special attention to data quality and precision because results of interest concern fairly small changes. Furthermore, others have quite specifically warned about bias in such surveys (e.g. “I suggest that the collection of data on sustainable travel behaviour is particularly prone to response bias and reporting bias”, Bonsall, 2008, p. 5).

With workplace surveys significant bias could easily result from low response rates. It is often difficult to achieve response rates above 50% in workplaces with standard survey approaches unless they have a particular interest in transport (e.g. public sector transport agencies where response rates of 90% have been achieved). Response rates as low as 20% to 30% are not uncommon.

Low response rates in the baseline survey may not hurt development of the travel plan much (e.g. because it may not matter much where the car mode share is 60% or 70%). But they do greatly reduce confidence that the typically small changes in mode share shown by follow-up surveys might be real. Without supporting evidence, it seems possible that bias of several
per cent could result from groups responding differently to the follow-up survey. Perhaps for example, staff who do not drive feel more motivated to deliver this ‘positive’ information in the follow-up survey than others who are (still) driving. Such bias could easily be driven by well-meaning attempts to increase response rate (e.g. emphasising the importance of doing the follow-up survey ‘to see how well we’re doing’).

We don’t have a magic low-cost way of achieving high response rates in all workplaces. But options to increase response rate in workplace surveys include:

**Personal assistance / reminders:** The travel plan co-ordinator or a hired interviewer can move around groups offering ‘help’ to complete the travel survey. With workplaces that are highly centralised into one or two sites, the cost of such face-to-face support is not all prohibitive like the cost of face-to-face household interviewing. The possible downside is that the ‘helper’ may be seen as someone inappropriately ‘standing over’ staff in some workplaces.

**Competition:** Reporting responses or response rates from different divisions or buildings can generate friendly rivalry in addition to simply being a tolerable way of keeping up awareness of the survey. Competition between classes can also be helpful to increase response rates with school surveys of parents.

### 2.3 Sampling error (margins of error) are relevant to travel plans

Next we deal with another statistical fundamental, leaving subtleties about sampling error to the next section. Even some quite experienced with surveys argue that sampling error is irrelevant to most school and workplace travel plans because in New Zealand the surveys typically aim to cover all pupils or all staff (that is, they aim to be a census not a sample). That argument is often sound for most staff satisfaction surveys, but not for travel plans. In other cases, it is not clear whether those reporting consider sampling error relevant or not – because they do not report margins of error, confidence intervals or significance tests.

Statisticians may like to think of this issue in terms of ‘enumerative’ versus ‘analytic’ studies (Deming, 1975). The rest of us may find it easier to think non-statistically from the perspective of public sector agencies like ARTA or Land Transport NZ as in the following thought experiment. Let’s imagine that perfect baseline and follow-up surveys (i.e. 100% response rate and 100% accurate answers) showed 50 fewer car trips for a particular workplace or school. From the perspective of the individual workplace/school, that might be seen as simple fact without any confusing effects of sampling error to worry about. However, from the perspective of a public sector agency or funding provider trying to detect the impact of their programme, sampling error is relevant. That school/workplace was just one of many such clusters of relevant citizens to whom the programme might have been applied. Hence the agency wants evidence that the 50 trip change is systematic and sustainable rather than perhaps merely being random fluctuation (e.g. a change of the size that might often happen just because people don’t travel the same way every day). That is, the agency needs to know that the change is statistically significant (and ideally, demonstrably caused by their programme). So margins of error (confidence intervals) or tests of statistical significance are relevant to travel plan evaluation.

To get sufficient sample size to reliably detect many differences of interest, evaluators may need to combine results from several workplaces or schools.

### 2.4 Other fundamentals

Because employer-based travel demand management started much earlier in the US, there are research articles from a decade or so ago about evaluating such programmes that still
have useful reminders of fundamental issues (e.g. Higgins, 1996). Weakness in such transport evaluations was noted in the 1980s and echoed in the 1990s; for example:

> Although the applications of these principles (good evaluation principles) are widespread in the literature (education, psychology and sociology), the use of these principles in transportation studies is limited.

(Hartgen and Brunso 1983; cited in Higgins and Johnson, 1999, p. 327).

We later present some innovations in data collection that can be useful, but we do not claim that these magically eliminate fundamental problems with these surveys. Rather we encourage users to have a realistic view accepting limits around precision of results from such surveys. We agree that:

- The overall conclusion ... is that the collection of information on the emergence of environmentally sustainable travel and activity patterns is fraught with difficulties ...

- Analysts should proceed with care and, where possible try to use data from a variety of sources (preferably ones which are unlikely to suffer from the same biases!). Bonsall (2008, p. 21)

### 3 Sample size and ‘panel design’ issues dealt with for personalised marketing are highly relevant to travel plans

Having now established the importance of determining whether changes are statistically significant (section 2.3), it is useful to borrow some hard-won lessons about this from ‘personalised marketing’. Personalised marketing (also known as individualised marketing and personal travel planning) is used to describe a programme aimed at changing people's travel behaviour by a combination of education, persuasion and provision of personalised information to either individual households or individual people. There are large-scale applications in the UK (Sustrans, 2006) and Australia (e.g. the 30,000 households in the TravelSmart programme for Darebin, Melbourne; Richardson et al., 2005).

Although evaluating the impact of such programmes has often relied heavily on household surveys (rather than surveys of workplaces or schools), underlying statistical principles are the same because these also involve before-and-after measurements. For good reason (O'Fallon and Sullivan, 2004), there was discomfort about the quality of evidence underlying claims of success with early applications of personalised marketing in Australia (e.g. Stopher and Bullock, 2003), albeit sometimes based on misperceptions (Roth et al., 2003). It is important that such lessons not be re-learned the hard way with travel plans.

Two fundamental lessons about sample design used in later Australian evaluations emerged almost simultaneously in Australia (Richardson, 2003) and New Zealand (O'Fallon and Sullivan, 2003):

- sample sizes might have to be much larger than expected (e.g. thousands not hundreds)
- panel designs (i.e. using paired measurements on the same people/households) before and after) could greatly reduce the sample size needed to reliably estimate change.

The same conclusions apply to the before-and-after surveys with travel plans. With travel plans, needing sample sizes of several thousand is less of a problem because the programmes in a region quickly swell to that size. Where larger than expected sample sizes
can surprise with travel plans is by confounding expectations of precision. For example, here there have been unrealistic expectations that results would quickly show differences in impact between different types of workplace or school that would help guide future implementation. But once one begins to compare individual workplaces or schools, one often runs into problems of insufficient sample size preventing detection of small changes.

We are not aware of any local study using a panel design to measure change in a workplace travel plan yet. But the improvement in precision seems worth exploring in workplaces with stable workforces. With school travel plans, panel designs seem irrelevant to assessment of long-term change because children change so fast in their travel behaviour simply as a result of growing older.

3.1 Other statistical issues

Very briefly, some other statistical issues are shared with personalised marketing and others are more specific to travel plans. For example, changes in travel behaviour found might not be caused by the travel plan but by background changes such as fuel price increases or broader public transport improvements. Hence interpreting the extent to which travel plans may have contributed to change found may be helped by analysis of trends or consideration of control groups (e.g. similar to Richardson et al., 2005).

On the other hand, there are some issues specific to workplaces or schools. For example, analysis of workplace data over time can be dramatically affected by major personnel changes (e.g. contracting out of major services or changing the balance between full-time and part-time staff) or re-location of the major sites. As yet, we are not aware of any comprehensive set of papers dealing with such issues in relation to travel plans comparable to those that have been done in recent years about before-and-after surveys of households in Australia (and direct contact with several of those active in the field in the UK helped less than hoped).

4 Measuring distance directly doesn't have to in the ‘too hard basket': Some innovations

4.1 Estimating distances from mode

It is possible to approximate changes in kilometres driven from changes in mode shift. That seems reasonable for estimating impact of projects in advance, as suggested by the average trip lengths supplied for use in standard New Zealand procedures (Maunsell et al., 2004) and the similar Victorian ones (Maunsell AECOM, 2006).

However, distance estimates specific to a particular site and the types of behaviour change occurring there promise more accurate estimates of impact than ones relying on broadly-based averages (although estimates using such averages do appear to be the most common method internationally still). In particular, there is often concern that trips changed by a travel plan may differ substantially in kilometres travelled from the average distance to work or school. As a simple example, consider mode shift resulting from implementing walking school buses. Walking school bus routes rarely extend more than 2 kilometres in length. Hence the vehicle kilometres reduced per trip because of the walking school bus are likely to be clearly less than the average distance children are driven to school.

The issue if one wants local measures of change (e.g. for a specific site or for a collection of sites that are part of a travel behaviour change programme supported by public funds) is then how best to collect estimates of distance. We consider three options below, two of them innovations we have not seen elsewhere. For brevity, several other methods considered or
used elsewhere (e.g. odometer surveys, GPS, respondents providing postcode or suburb, respondents tracing their routes use Google maps) are ignored here.

4.2 Subjective estimates of distance travelled: Not good enough for evaluating travel plan impact

Collecting distance information in surveys is a problem because people may not be able to report distances travelled with sufficient accuracy. Guidance on this point does vary however.

Some overseas sources suggested this could be a viable method:

- A major European Community mobility project involving collaboration between several countries included a direct question about distance as part of their highly detailed evaluation toolkit (MOST Consortium, 2003).

- The example travel survey questionnaire currently on the UK Department for Transport website simply asks: “How far do you usually travel to work?” (Department for Transport, 2007). Admittedly, this may be intended to meet objectives of helping to prepare a travel plan rather than to evaluate it (e.g. because promoting car sharing is more relevant if employees live further from work).

- A US report on estimating VKT reduction related to emissions did not use people’s estimates of distance in their methods because the majority of trips of interest to them were not commute trips. This implies that such estimates might have been acceptable to them for work or school travel plans that do involve ‘routine’ trips.

Research has shown that people have a hard time estimating mileage for all but the most routine of trips – such as commuting. (ESTC, 2003, p. 31)

On the other hand, other sources opposed such estimates quite directly for evaluating travel behaviour change:

Diaries and other similar surveys cannot provide accurate measurements of distance travelled because people are known to be poor at estimating travel distances... people report times and distances to within ±10 to ±25 percent. (Stopher et al., 2006, p. 7)

A local source of data was used to help clarify this issue. The Ongoing Household Travel Survey (Ministry of Transport, 2003) collects addresses for objective estimates of distance derived from geocoding and also subjective estimates (as a check on when geocoding fails to provide usable data). The Ministry of Transport ran comparisons for urban adults (age 18+) in the 2003-2006 NZ Travel Survey travelling from home to work (over 2000 trips). Even eliminating trips less than 3 kilometres in length (because rounding in responses can induce larger percentage errors for short trips), Figure 1 shows that about half the subjective estimates differed by more than 20% from the geocoded estimates (derived from addresses). On average, subjective estimates were systematically a little higher than the geocoded distances.
4.3 Online workplace surveys can estimate distances through user-friendly GIS

Estimates of distance travelled to workplaces or schools (albeit ignoring possible sidetrips) can be gathered systematically by getting employees’ addresses (with distances then calculated through geocoding and databases of the road network). Figure 2 has screenshot excerpts from a Greater Wellington survey showing:

1. How on-screen tips can be provided to help respondents through the process.

2. How information from databases can be shown quickly when only part of an address is typed (to reduce problems from misspelling, possible confusion over similar street names in different suburbs or nearby towns, typing of ‘Rd’ instead of ‘Road’, and so on); respondents can then select their address from the list rather than typing it in full.

3. How the final address to be used for estimating distances can be confirmed as correct by showing a map.

\[ \text{Discrepancy} = \frac{\text{Estimated distance} - \text{Digitised distance}}{\text{Digitised distance}} \]
Figure 2 – Three screenshot excerpts from an online workplace survey showing how links to address databases can help staff to reliably provide addresses

4.4 ‘Roll surveys’ enable collecting distance estimates from school students

Achieving high response rates (and hence minimal bias) with school students is very feasible. For example, ARTA circulates a detailed survey to secondary school pupils via class teachers, and the survey is completed in class time. Generally the only students missing from the sample set are those absent from class that day.

However, for primary school children many questions (including getting estimates of distances) are too difficult for children to complete accurately and greater reliance has to be placed on surveys of parents, which suffer the usual survey response rate problems.

ARTA’s innovation in using the school roll rather than a survey is an exceptionally practical one that could well be useful to those implementing travel plans in other countries. For follow-up surveys, ARTA works with each school in advance to develop a survey in the
format of the school roll. On the designated survey day each teacher asks each student how they travelled to school and how they plan to travel home and records this mode data on the prepared form. The whole process takes only a few minutes longer than a normal roll call, and works at all levels from Year 1 to Year 13.

Using this innovation, ARTA successfully completed a survey of 35,153 students across 68 schools as part of the 2007 programme evaluation. Such large sample sizes lead to very sampling errors; at a 95% confidence level the margin of error in reporting aggregate modes from this survey was clearly less than 1% (Hinkson et al., 2008).

The ‘school roll survey’ approach does not require any explicit estimate of distance travelled, because the student’s home address is part of the school roll database. Returned survey forms are matched with the complete school dataset and the distance from home to school is calculated using the standard distance calculation features of the GIS. The addresses also enable delivery of maps to schools which help travel planning by clearly showing where students are travelling from (and by which mode).

ARTA’s 2007 School Travel Plan evaluation found a reduction in travel by “family car” of 3.4%, and calculated that this led to a reduction of 3.3 million kilometres of vehicle travel and a saving of 1,000 tonnes of CO₂.

5 Conclusions

5.1 Something old

Evaluating impact of travel plans is often non-trivial. Fundamental principles of sound evaluation and old lessons about basics from US experience evaluating travel plans appear to be overlooked. There is a deceptive trap here that may underlie some of the problems: it may seem relatively easy to implement a survey of mode use as part of setting up a travel plan, but we warn that travel planners should think ahead to follow-up surveys and consider if high data quality is required to reliably detect small changes in behaviour.

5.2 Something borrowed

Some recent lessons about design of before and after household surveys to assess travel behaviour change can be ‘borrowed’ to help evaluate travel plans. In particular, sample sizes to reliably detect the small changes expected can be much higher than many anticipate, and panel designs may occasionally be useful (for workplaces with stable workforces).

5.3 Something new

Given the importance of assessing impact on greenhouse gas emissions, measuring the impact of travel behaviour change in terms of kilometres travelled rather than simply mode shift is now more important to many public sector organisations funding travel behaviour change (and to some companies). ARTA has pioneered two techniques that may well prove useful elsewhere for estimating distance information directly relevant to travel plans (they are already being adapted and sometimes extended by Greater Wellington Regional Council):

- use of GIS data and maps within online workplace surveys so that home addresses are accurate recorded as a basis for geocoding distances.
- roll surveys for schools (which enable estimation of distance for children of all ages by using the home address information recorded on the school roll).
Acknowledgements

The authors wish to thank the following for their contributions to this paper:

David Wake (Department for Planning and Infrastructure, Perth): for informing us of recent changes to their workplace surveys, including SKM (2007).

Jill Beck (Greater Wellington Regional Council): for allowing use of screenshots from a workplace survey.

Lynne Batty, Jacqueline Blake and their staff (Land Transport New Zealand): LTNZ funded Charles Sullivan’s work on measuring kilometres travelled in particular and performance indicators for travel plans more generally.

Lynley Povey, Ministry of Transport: for providing comparisons of subjective and digitised distances from the Ongoing Household Travel Study.

References


3 On 1st July 2008, Land Transport New Zealand became New Zealand Transport Agency
Evaluating changes associated with workplace and school travel plans

Sullivan and Percy


Richardson, A. J. (2003). Temporal variability of car usage as an input to the design of before and after surveys. 82nd Annual Meeting of the Transportation Research Board. Washington, DC, United States.


