8th INTERNATIONAL CONFERENCE ON SURVEY METHODS IN TRANSPORT
ANNECY, FRANCE, MAY 25-31, 2008.
Resource Paper for Workshop on Surveys for Behavioural Experiments

MYTHS, (MIS)PERCEPTIONS AND REALITY IN MEASURING VOLUNTARY BEHAVIOUR CHANGE

AUTHORS:

Werner Brög, Socialdata
Ian Ker, CATALYST (Consulting in Applied Transport, Access and Land use sYSTems)

ABSTRACT:

Increasing interest in 'soft' policy approaches to demand management, initially in personal travel, now in water and power usage, poses the question of how to measure the effectiveness of interventions. Much of the focus has been on statistical reliability of measured change where sample surveys are the primary means of estimating change. Sample surveys also pose issues of non-sampling errors, especially when the 'measure' is the difference between two estimates ('before' and 'after').

In some cases, direct measurement of usage appears less difficult – both electricity and water consumption are metered in Australia. The challenge in these cases is to establish the extent to which changes in consumption are real and to what extent they are the result of factors such as seasonal differences (either between seasons or from year to year) or demographic change in the household.

This paper outlines the principles and pitfalls in measuring behaviour change,. It draws on voluntary travel behaviour change, using a number of approaches, including but not limited to Individualised Marketing. A key issue is the extent to which repeated experience can validate the effectiveness of voluntary behaviour change interventions in general, despite statistical errors of individual measurements.

Measurement is fundamental to evaluation of outcomes. It can also aid the selection of locations with high potential to achieve change through identification of key success factors. In the specific case of travel behaviour change, there is now a substantial body of research that potentially allows outcomes to be related to socio-demographic and other factors. To-date, no strong relationships have been identified, but this would a useful area for further research.

Experience does demonstrate, however, that the scale of the intervention is important. Interventions with more than 5000 households are consistently more successful than small ones, even allowing for the greater statistical variability of measurement for smaller projects. Large-scale also offers opportunities for intervention design to benefit from the potential for diffusion beyond those directly involved in the project.

INTRODUCTION

The focus of urban transport policy has shifted from roads to public transport and non-motorised transport, from infrastructure to services, and from demand satisfaction to demand management. This is a reflection of changing values about what represents viable and sustainable cities and communities as much as it is of the financial, physical and social realities of attempting to provide major road infrastructure to satisfy continually growing demand for car travel.

Demand management has been given further impetus by the twin pressures of global climate change (private motorised transport is a large and growing source of greenhouse gas emissions) and, more recently, the so-called 'Peak Oil' phenomenon. Peak Oil contends that conventional oil production has either peaked or will do so within the next decade, which would be a major challenge in its own right but is exacerbated by the continuing rapid growth in oil demand from developing economies, particularly China. This paper is not the place to debate the extent or consequences of Peak Oil (see, for example, http://www.peakoil.net), but there is widespread agreement that energy for transport (whether from conventional oil, unconventional oil (oil shale, tar sands) or alternative sources) will be more costly.

At the same time, we are more aware of and concerned by the adverse health effects of lower levels of physical activity, partly resulting from increasing car use.

There is a large amount of research over more than half a century to support the analysis and estimation of the impacts of 'demand satisfaction', although much of this is problematic when there are substantial impacts on land use. Demand management does not have the same history of application, monitoring and analysis. Voluntary travel behaviour change is relatively new in the context of demand management.

The primary issue with any new transport initiative is identifying and estimating the impacts on the amount and type (including mode) of travel. This paper outlines the state of current practice, the issues arising and some ways of resolving them for voluntary travel behaviour change through household-based programs, primarily in transport but also in the emerging areas of water and electricity.

VOLUNTARY BEHAVIOUR CHANGE

What Is Voluntary Travel Behaviour Change?

Voluntary travel behaviour change is one part of the travel demand management toolkit, not a single 'silver bullet'. It is the face of demand management that operates largely without 'carrots' (direct incentives, including system improvements) or 'sticks' (disincentives, such as additional charges or regulation). In most cases, voluntary travel behaviour change (VTBC) raises awareness, improves availability of information and support for people to try alternatives to driving their cars, working through empowerment and motivation.

VTBC, itself, has a range of tools that operate in the household, school and workplace contexts, each of which is almost a separate area of development. Household-based programs are of two principal types:

- Those that deal with the whole identified population (usually of an area), but with a structured program that focuses on those with identified sub-populations that have the greatest potential for change. The major program of this type is *Individualised Marketing* (*IndiMark*®). See, for example, http://www.travelsmart.gov.au/training/packaging_comm_indi.html.

- Those that deal with a sub-population, often self-selected. These are largely based on *Travel Blending*® and its derivative *Living Neighbourhoods*® (see, for example, http://www.travelsmart.gov.au/training/packaging_comm_blend.html). *Travel Blending®/Living Neighbourhoods®* may also include aspects of community development, but we are concerned only with the travel behaviour outcomes in this paper.

- A third group (such as the 'Households on the Move' project in the Australian Capital Territory) is aimed at people or households making changes such as moving house that will inevitably have an impact on travel behaviour. Such an approach is potentially very effective for individuals, but because the participating people or households would change their travel behaviour anyway (because they are moving house or changing jobs, for example) there is no way of estimating the extent of travel behaviour change or even demonstrating that it has occurred.

Typically, there are no changes to transport infrastructure or services, other than the provision of route and timetable information at bus stops (in the case of IndiMark, new stands with route maps and stop-specific timetable information). Whilst there is some evidence that service improvements implemented in conjunction with *IndiMark*® can have a compounding (positive) effect (UITP, 1998), service changes during an initiative can have a detrimental impact by invalidating information and confusing participants.[1]

What Is Voluntary Behaviour Change Trying to Achieve?

From a public policy perspective, it is important to be clear about the objectives that are being sought by an intervention, before we can establish how best to measure the extent to which they have been achieved by the intervention. This sounds trite or self-evident, but is often not clearly articulated in the professional debates on such topics. In the case of water and power consumption, the objective is usually expressed in terms of actual consumption of water or electricity. In the case of transport, there is

---

[1] The change in bus boardings in the City of Subiaco, Perth, Western Australia, as measured through the Transperth ticketing system, was much lower in those areas where bus services were re-organised during the Indimark® period (personal communication from Department for Planning and Infrastructure, Perth, Western Australia).

less convergence on a single outcome and this has sometimes led to confusion about the nature and validity of measurements.

There are many potential outcomes of travel behaviour change that are of interest, across the full range of economic, environmental and social outcomes, including:

- *Economic*
  - The economic cost of transport
  - Traffic congestion

- *Environmental*
  - Greenhouse gas emissions
  - Local air pollution
  - Noise
  - Water pollution

- *Social*
  - Accessibility
  - Social inclusion

In transport terms, these objectives are often reduced to the shorthand of mode share targets, although these are usually aspirational targets rather than achievable outcomes. There are good reasons for the aspirational nature of targets, including the fact that the primary strategic objective is to change from increasing personal car use. It is also important that, since most of the policy measures that could contribute to this objective are new or have not previously been subjected to rigorous analysis, we do not know how effective they are likely to be, individually or in combination.

At the strategic level, the extent of car driver travel is a direct measure of the amount of private motor vehicle travel, which is the principal determinant of most of the higher level outcomes. There is much less certainty about the achievability of targets of the non-car modes than for the car driver mode and measurement of change is also correspondingly more difficult as they represent a smaller initial quantum.

INDICATORS OF CHANGE

It is difficult to measure the impacts of behavioural changes as a result of marketing actions. There are several methods available and each has advantages and disadvantages. So when measuring the success of a household-based travel behaviour change project, it is preferable to combine these methods and if they all point in the same direction and are consistent with the same size of change, then it is most likely that there was real success.

There are three main measures that should be considered:

- *Marketing Indicators.* These are determined by the amount and type of information requests compared with the total target group, and the quantitative feedback from residents throughout the project. For example, one of the desired changes is to increase the use of public transport and stop-related timetables are sought after by many people. It is extremely unlikely that thousands of households order specific, address-based timetables that they are not interested in and will never use.

  In traditional direct marketing, these types of indicators are the only success factors used. That might be a little too courageous, but they are reliable, precise and easy to measure indicators which should not be ignored.

  Further to this, it has to be remembered that behaviour change tries to effect a mind or culture change and it achieves this quite often. Hundreds and thousands of comments from people are available which document exactly this. These are unprompted comments from real people and are, in our view, as important as changes measured in counts or surveys: ‘Your programme was a great success, because it was the first time in 25 years that my husband used public transport and left his beloved Jaguar in the garage’ (Nuernberg, 2001).

  And, when it comes to changes in the mind set which affect longer term planning, this type of indicator should not be dismissed: “We are moving house soon and have now disregarded the option with bad public transport connections” (Vancouver, 2006). After moving house, these people may no longer live in the TravelSmart area, but they are still part of the outcome.

- *External Indicators* include measuring public transport patronage. In the TravelSmart program in Western Australia, bus boardings are collected and analysed independently and the ability to do so will be enhanced by the recent successful introduction of a comprehensive Smart Card ticketing system (SmartRider). ITP (2007) supports the value of ‘robust corroborative data’.

- *Behavioural Indicators*. The effectiveness of travel behaviour change projects can also be evaluated by measuring changes in the mobility characteristics of residents, by conducting extensive ‘before’ and ‘after’ travel surveys. The analysis is detailed, based on mode share, activities and travel time and shows the mode shift from car-as-driver trips to environmentally friendly modes. The key measures are the reduction in greenhouse gas emissions.

Most attention has been focussed on behavioural indicators, often to the exclusion of the marketing and external indicators. For external and behavioural indicators, it is important to measure changes against a control group (ITP, 2007, para 9.48).

Whatever methods are used, acceptance of the results will be highly dependent upon comprehensive and consistent documentation of processes and outcomes.

ISSUES RAISED BY VOLUNTARY TRAVEL BEHAVIOUR CHANGE

Engagement of the Community

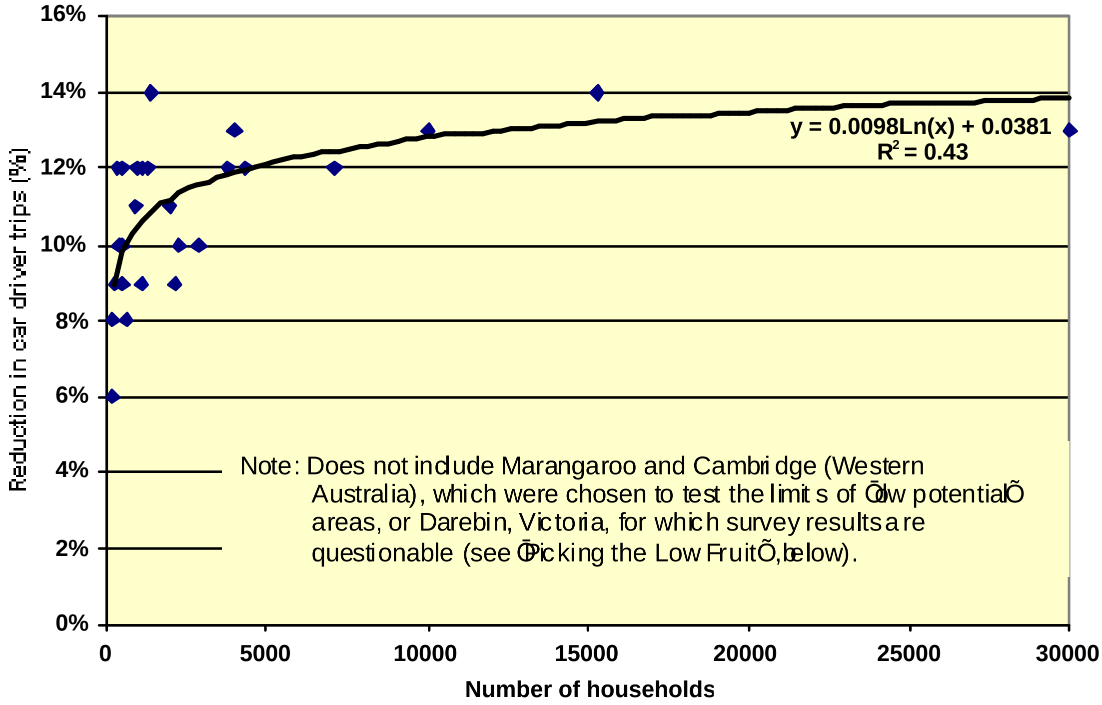
A key success factor for any program of VTBC is the engagement of the community. According to AGO (2006, p23):

There appears to be a correlation between changes in travel behaviour and:

- *personal engagement*
- *individualising materials to people's particular circumstances*
- *scale of the intervention – securing community support, as well as individual participation*
- *public visibility of the project (this is less important than the other points)*

The first two of these are difficult to measure, but there is clear evidence that the scale of travel behaviour change intervention has an impact on the outcome measured in terms of the change in car driver trips (Figure 1).

Figure 1 Travel behaviour change outcomes and scale of intervention (IndiMark®)



Public visibility will have a more substantial impact where diffusion is designed into the project (see 'Diffusion into the Broader Community', below).

Induced Demand?

Stopher (2003) argues that reduction in car driver trips from one part of the community will improve traffic conditions and induce additional car use. This is not a valid argument for car travel generated by households in the target area, as any such effect will already be included in the survey estimates. It might, however, have some validity for car trips generated beyond that area, primarily for trips undertaken in the most congested times and places in peak periods.

No evaluations have attempted to estimate any such effect largely because it is likely to be diffuse and could only be estimated through a detailed network-based traffic model. The authors are not aware of any successful attempt to incorporate travel behaviour change into a network-based travel demand and traffic model, which would be necessary to do this, although the requirements for doing so have been established (Ker, et al, 2005). It should be noted, however, that any such analysis would have to deal appropriately with statistical and modelling errors inherent in all traffic models.

The only reported measurement of VTBC impacts on actual road traffic volumes was for Darebin, Victoria, Australia. This appeared to show a reduction, relative to the metropolitan Melbourne average, of between 1% and 2% on roads outside the project area and between 2% and 3% on both main and local roads within the project area (Richardson, 2005). These data are consistent with diminishing impact of travel behaviour change by Darebin residents with distance from Darebin, but it is not possible to disentangle the extent of induced traffic from outside Darebin resulting from the lower traffic volumes. The same point is made by ITP (2007, p110).

The practical value of the Darebin data is significantly reduced by apparent problems with the other measurements of travel behaviour change for Darebin, including low response rates and sample losses from the panel.

Diffusion into the Broader Community

Diffusion is not a theoretical statistical concept, but a very real contributor to behaviour change across the whole population – including those outside the identified project population. It is well-known in many areas of commercial marketing. Diffusion relates to the extent to which awareness, attitudes and behaviours are transmitted within a population without external intervention (see box).

In the most common type of travel behaviour change applications, large-scale diffusion could only take place outside the area of application (as the total population was targeted). Since before and after surveys are only carried out in the identified target population, there has been no measure of any diffusion effect into the wider community. Indeed, to the extent that the diffusion effect might influence the control group (where one was established), diffusion would have led to under-estimation of car driver trip reduction in the identified target population.

Diffusion

There are three main models of innovation diffusion models, each arising from a different account of how innovations spread (Young, 2007):

- *Contagion*. People adopt an innovation when they come in contact with someone who has already adopted.
- *Social threshold*. People adopt when enough other people in the group have adopted.
- *Social learning*. People adopt once they see enough evidence among prior adopters to convince them that the innovation is worth adopting.

Contagion was deliberately avoided in the original South Perth pilot project, in order to test the effectiveness of the IndiMark® methodology itself. Subsequent larger-scale applications have not attempted to limit contagion, but have effectively kept its influence low by targeting a very large proportion of the population of an area – many of each person's contacts will already be part of the initiative.

Social threshold effects are most likely to occur where the innovation is clearly visible to the casual observer – as in the case of clothing fashion or the use of mobile phones. Travel behaviour will have some elements of this (eg if you see more people cycling), but in any area such visibility is limited by the extent to which travel in an area is undertaken by people from outside that area.

Social learning provides an explanation of why a person would adopt an innovation given that others have adopted it – the adoption decision flows directly from expected utility maximization. Specifically, the decision is based on reason to believe the innovation is better than what he is doing now, where the evidence comes from directly observing the outcomes among prior adopters. For example, when a new product becomes available, many people will want to see how it works for others over a of time before trying it themselves.

Working against diffusion is *inertia*, the fact that people delay in acting on new information. In the case of TravelSmart/IndiMark®, inertia is mitigated by its being known that the initiative is of a finite duration, even though the outcomes are expected to be durable and long-term.

In North Brisbane, Queensland, only 66% of households were contacted as part of the Household TravelSmart project. This approach simplifies and reduces the costs of making initial contact to establish a target population. At the same time, it leaves 'space' within total population of the area for diffusion to occur.

The measured travel behaviour change for the target population was consistent with expectation from previous IndiMark applications (13.4% reduction in car trips). There was no measurement of travel behaviour change in the rest of the area population.

The North Brisbane project created space in the intervention communities in which *contagion, social threshold effects and social learning* could be manifest – only 66%

of households were directly contacted. The very large scale of the North Brisbane project will have enhanced the likelihood of *social threshold* effects. There is also likely to have been a diffusion impact from several component (sub-area) interventions being under way at any point in time.

The key indicators of the extent to which the social learning model is likely to apply are those relating to the satisfaction of participants with first the process and second the outcomes (ie their actual experience with the non-car alternatives they use).

The evidence regarding the process is largely anecdotal and overwhelmingly positive. It is extremely rare for people to express dissatisfaction with the TravelSmart/ IndiMark® process, whereas many positive responses are received.

The evidence regarding the experience of the non-car alternatives is, however, more tangible. In the North Brisbane project, the 'after' survey of travel behaviour was undertaken in March/April 2007, between 3 and 12 months after the TravelSmart/ IndiMark® intervention (the intervention was progressively rolled out across the area over a nine-month period, April-December 2006). It is unlikely that those who had a negative experience with the non-car alternatives would still be using them months afterwards.

In turn, this supports the contention that the impacts of voluntary travel behaviour change will continue as long as the people who make the change continue to receive the same quality of travel experience as they do initially. This will require continued resourcing of public transport and walking/cycling infrastructure to ensure that additional demands that factors such as population growth and higher fuel prices put on other transport systems do not reduce the level of service they offer.

There are sound reasons to believe that diffusion effects will work better with large-scale than smaller applications:

- Materials for supporting travel behaviour change are available for the target area only;
- The larger the target area, the greater is the proportion of participants' family, friends, work colleagues and others (to whom information and attitudes might be transferred though such means as general publicity and 'word of mouth') who are likely to live in the area covered by the intervention;
- A large-scale project takes place over a longer timeframe, creating additional opportunities for diffusion while the project is still in progress; and
- A large-scale application increases the likelihood that when a household moves to a new location it will be able to access comparable location-specific information such as local access maps and public transport timetables in a familiar form.

Picking the 'Low Fruit'?

Figure 1 does not include three interventions that produced very low car trip reduction. These are:

- Marangaroo (4%) and Cambridge (7%) in Western Australia, which were specifically chosen to test the lower bounds of achievable outcomes. Both had poor public transport, high car use and poor provision of local facilities (important for walking and cycling). Including these would have introduced substantial variation in a range of variables that were not represented in the regression.

- Darebin, Victoria, which had a low survey response rate (50% or 1346 responses out of 2772 sample) for the 'before' survey and substantial further losses for the 'after' survey (881 responses out of 1346 in the 'before' survey – Richardson, 2005). Differences in response rates have been shown to affect the reported travel activity (Brög & Erl, 1999) and the sample losses between the 'before' and 'after' surveys are likely to be statistically biased.

This prompts the question of whether these outcomes shown in Figure 1 are the result of 'picking the low fruit' or selecting target areas with high potential (Bonsall, 2007). The Marangaroo and Cambridge examples indicate that there is potential for this to occur, but the wide range of situations (for example, the pre-existing car driver mode share in IndiMark applications in the UK and Australia varies from less than 50% to 60%) in which comparable results have been achieved indicates wide applicability rather than 'low fruit', albeit in the sensible policy context of not choosing overtly unfavourable areas.

Limited analysis of Australian IndiMark interventions has not demonstrated any strong relationships between outcomes and socio-demographic parameters (Ker, 2007). However, this analysis was undertaken for the purposes of providing benchmarks for evaluation of a specific project. A more rigorous multi-variate analysis would potentially allow interventions from other countries to be included and issues of multi-colinearity to be dealt with.

The recent North Brisbane Household TravelSmart project has demonstrated that a wide range of values for key socio-demographic parameters does not preclude the achievement of car driver trip reduction at least comparable to those achieved in smaller, more homogenous, and apparently favourable locations. For example, the North Brisbane area included areas with:

- 1.8 to 3.1 (whole area 2.5) persons per household;
- 4% to 27% (whole area 12%) of households having no car;
- 42% to 77% (whole area 54%) of population in employment;
- 21% to 62% (whole area 38%) managerial/professional employment status; and
- 40% to 63% (whole area 53%) of employed people driving their car to work (ABS 2007).

Durability

There is limited evidence of the durability of travel behaviour change outcomes, although what there is tends to show sustained reductions in car use for 4-7 years (Roth, et al. 2003). Bus ticketing data in Cambridge, Western Australia, has shown a consistent (and, if anything, increasing) increase in public transport use (Figure 2).

The most likely time for 'disillusionment' to be experienced and reversion to previous behaviour to occur, is within the first few months of the changed behaviour. Since 'after' surveys are typically undertaken 6-9 months after the intervention, it is unlikely that those who had a negative experience with the non-car alternatives would still be using them months afterwards. According to Maunsell Australia (2004a, p57), *previous experience indicates that for household/community initiatives there appears to be some reversion to previous travel choices over the first nine months following the TBhC project but that people who have not reverted by this time tend to stay with their new travel choice*. On this basis, the 'after' survey will already incorporate the most substantial part of any 'reversion' to previous behaviour.

There are strong arguments to support at least medium-term durability, provided those who change continue to experience the same level of service in their newly-adopted modes. If there are doubts about the durability of behaviour change, it would be appropriate to use a shorter period for evaluation – for example, New Zealand uses a 10 year evaluation period for travel behaviour change and other 'soft' policies (Maunsell Australia, 2004a). At typical public sector discount rates, this would reduce the present value of benefits by 40% compared to a 25-year evaluation period. Whilst this is a substantial reduction, the impact on benefit-cost outcomes is unlikely to be critical, with benefit-cost ratios in excess of 30:1 typically being recorded for *IndiMark®* projects with 25 year evaluation periods.

MEASURING BEHAVIOUR

Direct Measurement?

The behaviours in which we try to measure change are often not directly measurable. Water , electricity and gas, in Australia, are directly metered at the consumer household level. However, we still do not know what the water or electricity is used for and the effect of exogenous factors can be difficult to identify. Some of the complexity of doing so, even at a highly aggregated level, is outlined in http://www.daa.com.au/case-studies/water-usage.

A common problem with direct metering is ensuring accurate and timely readings on a regular basis. These problems can be eliminated by using self-reporting meters, which are already available on a walk-by basis for water, electricity and gas (see, eg,http://www.neptunetg.com) with no need for access to premises or manual recording of meter reading) and could be adapted to automatically send readings by phone at specified times. Alternatively, data-loggers could be attached to meters for periodic automated walk-by download.

For the basic parameter of concern in VTBC, car use as driver, the extent of experience now available to support the robust effectiveness of the main travel behaviour change tools, has been interpreted (AGO, 2006) to support the use of simpler single measures such as odometer readings. However, this is not without its own problems, particularly in the case of households with multiple motor vehicles, where vehicles are used both for personal and business travel (a common issue in Australia with its Fringe Benefits Tax providing concessional taxation of cars as part of remuneration packages) and where vehicles have been bought or sold during the monitoring period.

Some Issues with Meters

In the recent behaviour change literature, it is often said that the use of meters (in transport: odometers) would be a promising tool to get better results. The main arguments to support the use of meters are that respondents in surveys are self reporting and might want to report project-friendly results and that meters are more precise and statistically valid. However:

- Even in water projects (where meter readings provide the only source for evaluation), the readings are reported readings. In the case of water, even by an 'independent' person. But painful experience from five water projects has shown that meter readers make mistakes, are sloppy or do not even read the meter and make the results up.

- In an odometer reading project, these problems get greater. It is self-reporting again, in this case by a household member. And might they – as is sometimes argued in survey-based evaluations – want to give project-friendly results? This risk alone would cut out odometer readings as a viable option, because there is nothing easier than to simply report 1000 km less, to pretend great success.

- Additionally, the common odometer reading projects achieve very low participating rates (recruitment rates Stopher, et al (2007a, Table 1) reports for newly recruited households vary between 12 % and 25 %) and need a strict regime to be kept by the households. Adding to the problems of low participation, *in any given wave, about 25% of households will fail to provide odometer readings* (Stopher, et al, 2007a). At best this increases the required sample size, but in conjunction with low recruitment rates it looks more like evidence of systematic bias that cannot be addressed by sample size. In a recent validation survey on odometer readings presently conducted by Socialdata Australia (recruitment rate 80 %), about two thirds of the readings were reported in doubt (collected at the wrong time, wrong day, wrong car, reading invented, later readings lower than earlier ones, distances travelled of several thousands kilometres a day, etc.)

Even if the meter reading is correct, it tells us nothing about the type of travel, its frequency (number of trips) or trip purpose.

For travel generally, and car use in particular, we are forced to less direct means of estimation, unless the outcome is so large that it can feasibly be measured through road traffic volumes. Even then, we face the problem of differentiating the traffic generated by the target population from that generated by the rest of the city.

It has been argued that geographical positioning systems (GPS) technology can provide a means of directly measuring travel and, by inferential means, mode use (see, eg, Stopher 2007a). However, this is, as yet, unproven in at least two key areas:

- The robustness of the algorithms for inferring mode use; and
- The potential for the measurement actually to influence travel behaviour – for example, a person 'equipped' with a GPS device might be more aware of their travel behaviour (as distinct from its being habitual) and use transport more efficiently while carrying the device.

Unfortunately, there do not appear to have been any comparisons between GPS and surveys in respect of either error or data capture (AGO, 2006, p57).

GPS might have potential value in validating a travel diary or survey (for example in terms of places visited and times of travel), subject to the caveat of measurement influencing behaviour, above, but the reported difficulties of recruiting people to take part in trials (Stopher et al, 2005) and retaining them in subsequent waves of a GPS panel (Table 1) suggests that it might be difficult to achieve a sampling basis that is sufficiently representative of the community generally for useful results to be derived.

Table 1 Recruitment and retention in GPS trial (South Australian Panel)

Status	South Australian Panel (50 Households nominally)		
	Wave 1	Wave 2	Wave 3
Recruited (new to wave)	57	17	0
Completed	50 (88%)	14 (82%)	0
Continuing (Recruited)	--	35 (61% of 57)	44 (59% of (57+17)
Completed (Continuing)	--	32 (56% of 57)	36 (49% of (57+17)

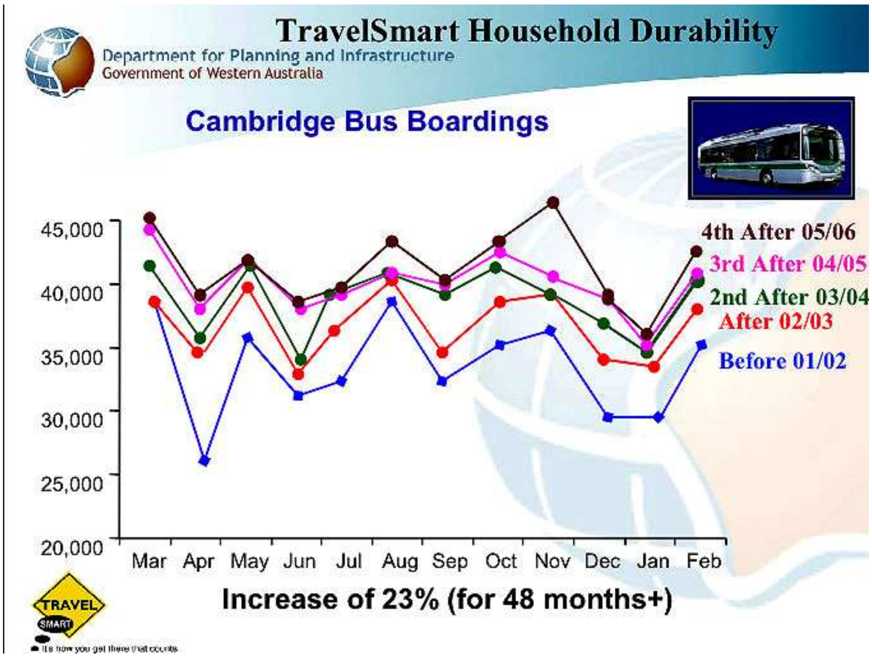
Source: Stopher et al (2007a, Table 3)

Alternatively, GPS might have a more robust role if attached to the vehicle rather than the person, subject to the caveats already noted about multi-vehicle households, vehicle changes and identifying business use of vehicles.

For public transport, ticketing systems may allow direct measurement of trips originating in an area. Where this was possible, ticketing data has been consistent with the change estimated from travel surveys (see, eg, Figure 2).

Implementation of smartcard ticketing systems will facilitate use of ticketing information to estimate changes in public transport patronage related to VTBC.

Figure 2 Public transport boardings and IndiMark® - Cambridge, WA



Source: http://www.dpi.wa.gov.au/mediaFiles/ts-Cambridgebusdata.pdf

Surveys

VTBC initiatives have most often been assessed by surveys of various types (see 'Cross Sections or Panels, below). We generally measure change as the difference between 'before' and 'after' surveys, with 'corrections' indicated by a control group (see 'Measuring Differences', below). It follows that is critical that the 'before', 'after' and 'control group' surveys are undertaken using the same survey design and methodology. Differences in design or methodology will affect response rates (see 'Response Rates', below) or have more subtle impacts on the accuracy of responses that make comparisons problematic.

Cross-Sections or Panels for Surveys

Many travel behaviour change interventions have adopted a cross-sectional survey approach to estimating 'before' and 'after' travel behaviour. For small projects, this can create issues of statistical reliability of the difference between the two surveys, but these are less severe for larger applications.

It has been suggested that a panel survey approach, using the same people for both surveys, would be more suitable. However, panel surveys have their own problems that are more systematic and less amenable to treatment by statistical analysis.

- Even an ageing population ages more slowly than the individuals that comprise it. At a more aggregated level, it ages more slowly than the households or other groups that comprise it. Even a panel that remains intact over a period of time, will not be the same at the end of the period as it was initially.

- Whilst this might not make very much difference for the typical ‘before-after’ time period of 12 months, it has an increasing impact over time and adversely affects the comparability over longer time periods and, hence, the estimation of the durability of behaviour change.

- In practice, there is attrition of panels. If no action is taken to replace, the integrity of the panel is compromised by the possibility of self-selection bias among those opting out. It also poses a difficult problem of identifying and recruiting replacements who have the same characteristics as those who leave.

- People also suffer from ‘survey fatigue’ and may drop out or become less reliable for that reason. Whilst this can be addressed through ‘refreshment’ (replacing a proportion of the panel on a periodic basis), this also raises the issue of how to ensure that replacements have the same characteristics, including travel, as those who are discarded.

Whilst panels have been argued to require lower samples sizes, some of the parameters enabling smaller sample sizes also give rise to survey designs which are more difficult to undertake. For example, panel survey data is more difficult to obtain (with full control of other biases) than repeated cross-sectional data (Richardson et al, 2003). Acquiring repeated data from the same respondents is challenging and a reduced response rate in the ‘after’ survey can lead to sampling bias (ITP, 2007, para 9.24).

One of the examples of a panel survey to estimate the effect of a VTBC initiative had a low initial response rate (49% - 1346 out of a sample of 2772 households) for the ‘before’ survey and substantial further loss between the ‘before’ and ‘after’ surveys. Furthermore, in the ‘after’ survey only 682 of the 881 households that responded to the ‘after’ survey had the same composition as in the ‘before’ survey (Richardson, 2005). At the individual level, people may exhibit substantially different travel behaviours at different times for reasons that are purely idiosyncratic and not related to the intervention being investigated.

In two cases where both panel and cross-section surveys were carried out (North Brisbane, Queensland, and Victoria Park, Western Australia), the panel survey recorded slightly smaller, but still very substantial, reductions in car driver trips (11%/13% and 12%/14%, respectively). This might be a result of people who are less likely to change residential location (and hence more likely to remain in the panel) being also less likely to change other behaviours, including travel (see Box).

Integral measurement of travel behaviour, as part of the intervention, is a form of panel survey that highlights some of these problems, most notably those of attrition and measurement actually influencing the outcome (travel behaviour).

Comparing Panels and Cross-Section Surveys: Two Case Studies

Two unpublished studies in Australia have evaluated a TravelSmart project (Brisbane North and Victoria Park, Perth) with a panel and repeated cross-sectional (random) surveys.

Sample sizes (respondents – persons)

	Before	After Panel	After Cross-Section
Brisbane North	1309	825	1381
Victoria Park	905	780	766

The response rates were in the seventies and eighties, the combined response rate for the panels in the mid-sixties. If we compare the results of the panels with the cross-section survey, we can see that the behaviour changes picked up in cross-section were always greater than in the panel.

Measured travel behaviour change – panel and cross-section (random) surveys

BRISBANE NORTH			**VICTORIA PARK**	
Relative change			**Relative change**	
Panel	**Random**		**Panel**	**Random**
+ 32 %	**+ 49 %**	**Walking**	**+ 12 %**	**+ 12 %**
+ 33 %	**+ 58 %**	**Bicycle**	**+ 60 %**	**+ 48 %**
n/a	**n/a**	**Motorcycle**	**n/a**	**n/a**
- 11 %	**- 13 %**	**Car driver**	**- 12 %**	**- 14 %**
+ 8 %	**+ 8 %**	**Car passenger**	**- 1 %**	**0 %**
+ 16 %	**+ 22 %**	**PT**	**+ 14 %**	**+ 17 %**

This supports the anecdotal evidence accumulated over more than hundred projects that people who are flexible, move more, change more easily and are underrepresented in panels, which tend to focus on stability, less change, status-quo. Behaviour changes measured in panels therefore seem to be smaller than in reality.

This conclusion is supported by current research into travel behaviour change projects for which panel evaluation has been selected. In all these surveys, the response rate of the ‘before’ survey was at least in the mid-seventies and the response rate of the ‘after’ always in the eighties. Thus the combined panel response was always over 60%.

Using the speed of response technique, it was possible to simulate to traditional before and after panel with a response rate in the low forties (before) and the low seventies (after) – a common combined response rate of about 30%. In terms of the analysis above, the selection worked even more in favour of stability and no change. This is clearly reflected in the results. Behaviour changes in a panel with a combined response rate of about 30% are smaller than in a panel with a combined response rate of 60+%.

This research will be presented in full at another occasion.

USING SURVEYS

Sample Size

With a sample survey, the reliability of a sample estimate (in this case, the mean proportion of trips by car as driver) is inversely related to the sample size, both in absolute terms and as a proportion of the population being sampled.

A great deal of attention has been paid to the theoretical requirements for sample sizes to identify relatively small changes at a reasonable level of confidence. Stopher, (2005) suggests that to identify a change of 2% at the 95% confidence level it might be necessary to gather data from around 8,000 people in a before and after study. Such a requirement would be beyond the resourcing likely to be available for monitoring any such project and would, in any case, substantially diminish the resources available for actually delivering the VTBC initiative.

The focus on establishing statistical confidence requirements for small levels of behaviour change appears to stem from an erroneous assumption that the behaviour change with *IndiMark®* was measured only over the actively-participating population (Stopher, 2003). The Australian Greenhouse Office has noted that: *In some of the reports evaluated, there appears to have been misunderstanding of statistical methods used to interpret results, and hence problems in the sample sizes, sampling methods and evaluation periods employed* (AGO, 2006, p58).

Much of the contention about sample sizes, however, is predicated on a desire for accurate estimation of the level of travel behaviour change, whereas the public policy imperative is for assurance that the level is greater than the level that would justify the initiative in socio-economic terms. Given the high benefit levels from travel behaviour change, this level is a very low bar – the prospective benefit-cost ratio for completion of the Perth TravelSmart Household program estimated a benefit-cost ratio of 67:1 on the basis of a 10.7% reduction in car driver trips (which was the weighted average of projects to-date in Perth, including two project areas (Marangaroo and Cambridge) specifically chosen for their 'low potential') (Ker, 2004). Typically, urban transport projects are regarded as 'justified' if they achieve a BCR of 3:1 or 4:1. On this basis, VTBC would be justified in benefit-cost terms with a car trip reduction as low as 1%.

There has also been a strong tendency to focus on the probability of the change being <u>less than</u> estimated through sample surveys, with little or no recognition of the corresponding probability of its being greater.

If the population mean is greater than 2% (as <u>all</u> the interventions except Darebin indicate), by a factor of up to seven times and an average of around four times, the sample size requirements become much less onerous.

Most of the debate on sample size has been on the basis of sampling error estimation. In practice, the error due to systematic factors (eg response rate, qv, below) can be many times larger than statistical error (Table 2). This is a serious problem, even if we do only focus on significance, because there are always two types of errors to be considered in empirical studies:

- Random errors (sampling-related) and
- Systematic errors (design related).

Since there are always systematic errors, the calculated random errors are always based on a wrong assumption (that there are no systematic errors).

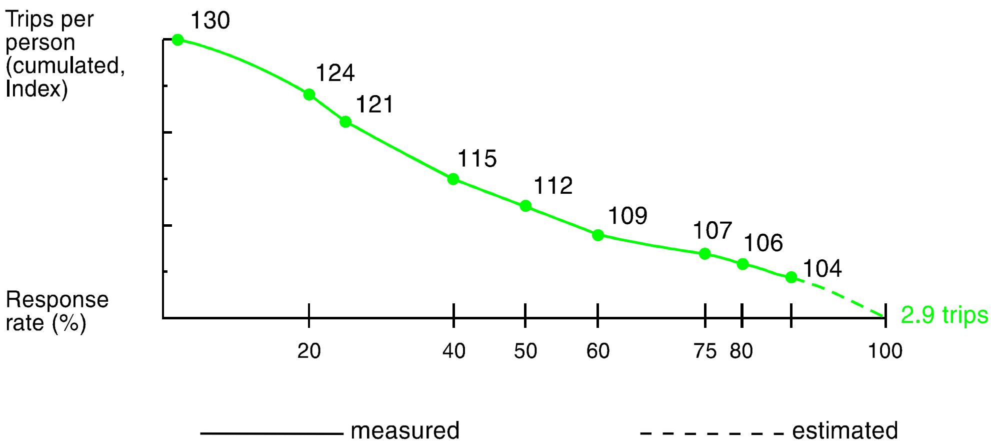
But the real problem of these two types of errors is their nature:

- Random errors can be calculated exactly, but not corrected, whereas
- Systematic errors can be corrected but not (precisely) calculated.

But in travel behaviour, the systematic errors outweigh the random errors in size significantly.

One of the most useful factors, to understand, calculate and correct systematic errors, is the response rate of a survey, and one of the best variables is the number of trips per person per day. In a mailback survey, the response can be analysed by speed of response. This has two advantages: lower response rates can be easily simulated, and the whole process can be analysed to estimate the final (unknown) result for the total. Both things have been done and this function has been used to analyse the likely shortcomings of a parallel survey that achieved only 25 % response (Figure 3).

Figure 3 Trip rate by response rate: Vienna 1993 (response rate 85%)

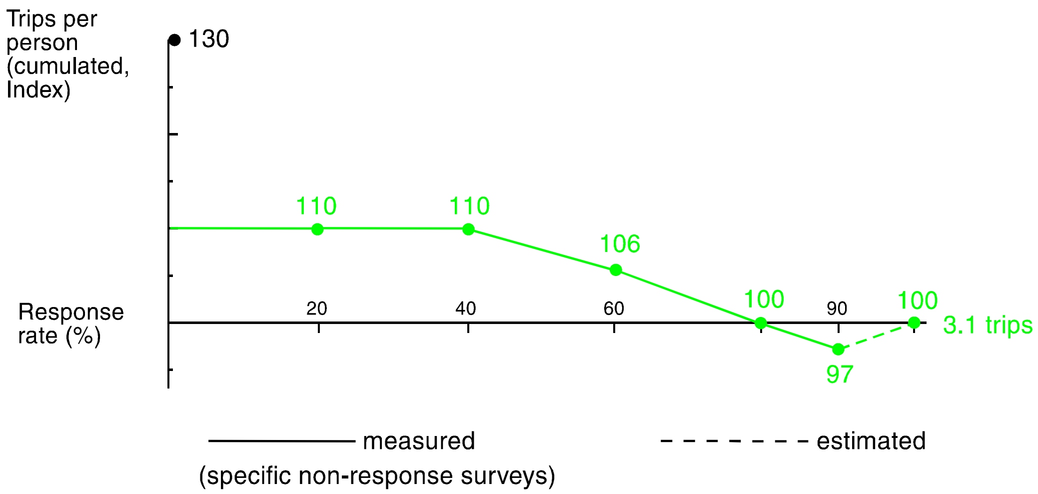


Source: Brög and Erl (1999)

The technique developed and used has been further progressed and applied in the Netherlands' National Travel Survey, the only long-running continuous travel survey in the world (MVRWAVV, 2007). As a consequence, a specific self-validating design has been developed and applied. This design can correct for non-reported items, non-reported trips and non-response and is not dependent on external data.

If we look at the respective response curve, we see that detailed analysis of reasons and effects with improved design of survey methodology has reduced the non-response-effect significantly, yet it is still there: With an estimate of the total trip rate to be 3.1 (100), the trip rate at a 40 % response rate would still be 10 % higher (Figure 4).

Figure 4 Trip rate by response rate: Netherlands (MON - Mobiliteitsonderzoek Nederland)



Many current designs applied to evaluate behavioural change projects have a much higher influence on systematic errors and the difference in the trip rate at a 40 %-level (which is normally regarded as good) would be considerably greater. However, if we use the MON data and simulate a situation of a 40 %-response rate we would have needed a gross sample of 166 250 people and achieved a trip rate of 3.4. The 'real' MON achieved an overall response rate of 72 % and a trip ate of 3.1.[2]

In the first case, the systematic error would be 0.3 trips per person and day, the random error only 0.02. This one type of systematic error, on its own, using only one variable shows already that the systematic error is fifteen times the (widely and commonly used) random error (Table 2).

Table 2 Random and systematic errors

Net responses	66,500 respondents (net)	
Response rate	40%	72%
Gross sample required	166,250	92,350
Trips per person per day	3.4	3.1
Random statistical error	±0.02 trips per person per day	
Systematic (response rate) error	±0.30 trips per person per day	

[2] Non-reported trips are very similar at all response rates (c5%), so the inverse relationship between trip rate and survey response rate is not a function of non-reporting.

It has been suggested that the use of GPS for measuring travel behaviour would permit a smaller sample size than for a conventional travel survey (Stopher, et al, 2007b). For example, we understand that the sample size proposed for long-term monitoring of travel behaviour change in North Brisbane, using GPS, is 300 households. It is not clear why this is any more so than would be the case for multi-day diaries, as the variability due to random sampling would be the same in both cases. The potential for non-sampling and response errors exists, in different ways, for both methods, but is not related to sample size. Indeed, the difficulties of recruiting and retaining households in GPS panel surveys (Stopher et al, 2005 and 2007a) suggest that non-sampling errors are likely to be a severe problem.

Daily or Weekly 'Diaries'?

Most travel surveys are based on a single day for each household, rather than a full week, despite the fact that travel behaviour for any individual will vary by day of week. This is most notable, on a systematic basis, between weekdays and weekends, with most trip purposes, other than leisure and recreation and some types of shopping trip, being focussed on weekdays. The rationale behind a daily diary is precisely that this differentiation is systematic and therefore can be picked up by appropriate sampling on each day of the week. Daily diaries should represent each day of the week equally.

A weekly travel diary will pick up the variation between days for individuals and individual households, but does so at the expense of:

- A lower response rate, as fewer people are willing to commit to the extended commitment;
- A higher attrition rate, as people cease to record their travel during the week; and/or
- Lower quality data towards the end of the survey week, as the novelty wears off.

Repetition

Experience with repeated application of *IndiMark®* (Figure 1) shows a high degree of consistency with:

- greater variability of outcomes for small applications (consistent with statistical expectations resulting from smaller survey sample sizes); and
- larger reductions in car driver trips (and lower variability) for larger applications.

The probability of all of these outcomes being over-estimates of the population value (let alone sufficiently large over-estimates to invalidate the principal evaluation conclusions for VTBC) is very small and gets smaller with every repeated application.

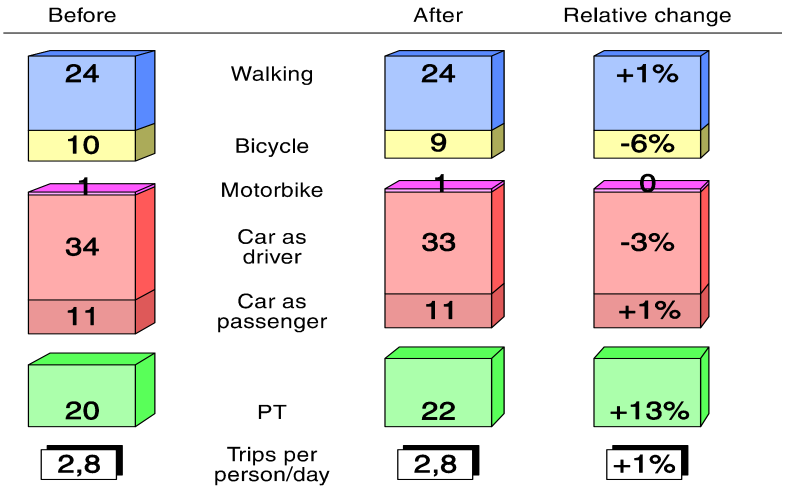
Alternatively, if the individual projects were to be aggregated into a single notional 'program' with a number of different spatial locations for evaluation purposes, the total sample size would now exceed the theoretical sample size required for measuring a level of change that is less than a quarter of that actually achieved. If these had been undertaken as one project, even this requirement would have been achieved. The consolidated result for Perth TravelSmart projects, including locations chosen for their likely poor performance (Marangaroo and Cambridge), is a 10% reduction in car driver trips (http://www.dpi.wa.gov.au/travelsmart/14960.asp).

This conclusion is supported by the Australian Greenhouse Office (AGO, 2006, p53): *For households, predicting the effectiveness of a Travelsmart project is now essentially a solved problem. The eleven Australian projects evaluated here join over 50 other formal evaluations conducted internationally and many other informal assessments (See Ker 2003, Maunsell Australia 2004 and UK Department of Transport 2004 for a list of some other evaluations.) While individual outcomes vary with geographic location, what can be said broadly is that community-based household projects will achieve a reduction in car travel of 5–15%, and this change appears to be sustained for several years without further intervention. Methods for achieving these results are now well understood, and further evaluations are unlikely to add much to existing knowledge.*

The practical value of repetition is evident from the City of Nürnberg, Germany (500,000 people), which uses Individualized Marketing (IndiMark®) to promote public transport (ie not in multi-modal mode). It started in 1996 with a small project of nearly 5000 people. From then on, every year several projects have been conducted. Since 2006, early areas are worked on for the second time and by the end of 2007 a total of 600 000 people has been part of the program (including repetition). All projects were successful and attitudes in the whole city have been changed.

Each project has been evaluated separately. In 2004 all evaluation surveys up to a certain point were amalgamated and analysed. The result is an increase in public transport by 13 % (fully supported by counts and calculated against control groups) and – although not targeted – a reduction in car driver trips of 3% (Figure 5). This is still continuing.

Figure 5 Consolidated Behaviour Change 1996-2004: Nürnberg, Germany



Source: Nürnberg (2006)

Response Bias

There is sometimes a danger that the very fact that a person is part of a program will influence that individual's actual or reported behaviour. In the case of VTBC, the influence on actual behaviour is precisely what is sought by the program, provided that this is an ongoing response and not a temporary one simply to be able to present in a better light for the program itself.

The greater danger lies with 'program-compliant' reported behaviour, although this is more likely to be found in panel surveys than in repeated cross-section surveys. Participants in a panel survey will know what the base travel behaviour was (the 'before' survey) and could, theoretically, report appropriately changed behaviour in the 'after' survey, even if actual behaviour had not changed. Participants in a cross-sectional 'after' survey will not have a 'base' to compare themselves against and will be less familiar with the detailed program objectives.

Misreporting travel behaviour in surveys happens for a variety of reasons, irrespective of any potential desire of the 'observed' to please the 'observer'. The best way to guard against the latter is exactly the same as for the former. Since even apparently simple activity and travel patterns can be quite complex, involving issues of location, duration, speed and mode, it is more difficult than might be expected to falsify a survey response in a realistic way. Detailed inspection of survey responses will show up unusual, inconsistent or infeasible travel behaviour that can then be followed up for clarification and, where necessary, correction.[3]

Bonsall (2007) draws attention to the possibility the repeated cross-section surveys will *by happenstance, result in some households being contacted more than once,* which he says could have negative (eg households feeling hassled by repeated surveys) or positive (the ability to explore changes in individual behaviour) consequences. In practice, unless sample sizes are larger than necessary or surveys are continually repeated to assess durability, multiple-contact households will be too few for any such issues or opportunities to be substantial.

Control Groups

Behaviour change initiatives do not operate in isolation. Almost by definition, such initiatives are funded because there is a sympathetic government policy environment. In recent times, the increasing price of petrol will also be having a systematic impact on travel behaviour. This requires that, wherever possible, we identify 'control groups' of people or households similar to those subject to the intervention, so that the impact of factors external to the intervention can be identified and appropriate adjustments made to the measured outcomes in the intervention group.

[3] The second author has direct experience of this in an *IndiMark*® project, having been part of the 'after' survey and contacted for clarification of his household's 'unusual' travel pattern – only one car and one licensed driver in a household of four adults; circular trips involving delivery on-foot of local newspapers; low level of travel activity; very low car driver trips.

This poses a problem for interventions that do not attempt to measure change over the whole target population, as there is a potentially high degree of self-selection in the intervention group that makes them different from any potential control group.

The key criteria for a control group include:

- Spatial location
- Demographics (age, gender, household size/structure, income, education, employment type and status)
- Opportunities in the relevant area (eg similar transport systems and activities)
- Exposure to the same exogenous factors such as petrol price, public transport fares and transport system development.

The only group that, in principle, meets all of these criteria (subject to sampling variability) would actually be a random sample from the population of the intervention area itself. This is particularly so with a very large-scale application that covers a spatially-large and demographically-diverse area. In practice, however, even those who are not part of the target population are highly likely to be influenced by the intervention itself (see 'Diffusion into the Broader Community', below).

Control groups, where used, are usually external to the intervention community and any measured travel behaviour changes in them have been small, as would be expected over a short period of 12 months. In recent years, the major systematic influence has been from increasing fuel prices and it is highly likely that the same impact would be found in both the control group and the identified target population. The effects of this can be unexpected, however – for example, in the recent North Brisbane project, the petrol price in March/April 2007 ('after') was very similar to the price in March/April 2006 ('before'), but in between there was an increase (followed by a decrease back to earlier price levels). The reduction in price leading up to March/April 2007, coming after a 2-year period of sustained price increases, could account for a measured increase in car use in the control group (Ker, 2008).

Response Rates

The response rate to a survey is an important measure of its reliability. It also affects the value of key parameter estimates from the survey. To obtain accurate data, it is important to interview a representative sample. When a survey's response rates are low, it's unlikely that the sample interviewed is completely representative. The data collected from a survey with low response rates can suffer from non-response bias and the decisions made using those data may be flawed. Response rates in all forms of travel surveys have been declining and telephone survey response rates are rarely above 60% and are often 20-30% (http://rns.trb.org/dproject.asp?n=14124).

With telephone surveys, non-respondents can include high-mobility people (who are often not at home) as well as those who don't consider themselves to have a sufficiently strong 'stake' in travel issues (eg because they travel very little).

With mailback surveys, it has been shown that the average number of trips per person per day varies inversely with the response rate (Brög & Erl, 1999). In other words, people who are less active in terms of travel are less likely to respond to surveys. This has a double importance when we are seeking to estimate the *difference* between two estimates:

- For any percentage change, the number of trips will depend upon the absolute values which in turn is inversely related to the response rate. It follows that a low response rate is likely to over-estimate not only the before and after trip making but also the difference between them.
- A difference in response rate between the 'before' and 'after' surveys will also affect the estimate of the difference between them. If the 'before' response rate is higher than the 'after' rate, the difference is likely to be under-estimated, and vice versa.

The best estimate of difference between two cross-sectional surveys of given sample size will come from surveys that achieve a similar high response rate. IndiMark applications typically have survey response rates of 70-80% for 'before' surveys and 70-85% for 'after' surveys. These surveys also show a consistency of frequency of out-of-home activities between the 'before' and 'after' surveys, indicating that the difference between them is unlikely to be a major influence on trip rates or mode usage.

Low and varying (between 'before' and 'after' surveys) response rates might be a partial explanation for differences in measured effectiveness between types of intervention.

MEASURING AND REPORTING BEHAVIOUR CHANGE

Measuring Differences in Behaviour

It is intrinsically more difficult to measure changes in behaviour than behaviour itself, largely because changes can only be measured as the difference between behaviour before and after. Estimation of differences is, therefore, affected by all of the issues associated with the measurement of both the before and after states.

In principle, this problem can be eliminated in the limited context of measuring travel behaviour across the whole participant population (as, for example, with early *Travel Blending* applications), with the before and after travel diaries being an integral part of the overall intervention. Unfortunately, this provides no information about behaviour change among those who do not participate right through the process and no generalisation to the whole population is possible.

For those interventions that do measure behaviour change across the whole identified target population, irrespective of their level of involvement (including zero), it is still possible to make some useful assessment of the reliability of estimates.

So, let's look at a specific example. The North Brisbane Household TravelSmart project achieved an estimated 13.4% reduction in car driver trips (8 percentage points, from 58.4% to 50.4%), which is significantly different from zero with a 99%-plus probability[4] (Socialdata, 2007). However, the tests do not estimate the potential range of the magnitude of the difference.

This result can be used inferentially to estimate the upper and lower confidence limits for the value of the difference between the means:

- The 'before' and 'after' surveys each have a sampling distribution for the estimated mean value of car driver trips;
- The sample size for each of the surveys is sufficiently close that we can assume, for practical purposes, that the sampling distribution for each is the same;
- The sampling distribution can be described in terms of the standard deviation;
- A 99% probability that the proportion of car driver trips was lower after IndiMark® means that there is less than 1% overlap between the two sampling distributions;
- With a normal distribution, 99% of the sampling distribution is within 2.5758 standard deviations of the mean;
- It follows that the difference between the estimated means is at least 5.15 standard deviations with 99% probability (Figure 6);
- Whilst the distributions of the 'before' and 'after' sample means indicates that the true value (99% probability) <u>could</u> be anywhere between zero and 15.6 percentage points (26.8% of previous car driver trips), the distribution is heavily weighted towards the mid-range;
- The difference between any two normal distributions is itself normally distributed. A normal distribution with a 99% probability range of 0 to 15.6 has a standard deviation of 3.0.

Since Figure 7 shows an 80% probability of the reduction in car driver trips being <u>between</u> 3.8% and 12.2% of all trips and the range outside that evenly distributed above and below that range, the probability of the reduction being greater than 3.8% is 90%. There is a 10% probability that the reduction exceeded 12.2% of trips.
In terms of the more commonly stated percentage reduction in car driver trips, these figures are equivalent to:

- An 80% probability of the reduction being between 6.4% and 20.4%;
- A 90% probability of the reduction being greater than 6.4%; and
- A 10% probability of the reduction being greater than 20.4%.

---

4 Target Group net sample size 1,309 ('before') and 1,381 ('after'), with 76% and 79% response rates, respectively. Total population: 277,000 people or 113,000 households.

Figure 6 Sampling distribution of the mean (Before (P1) & After (P2) IndiMark®)

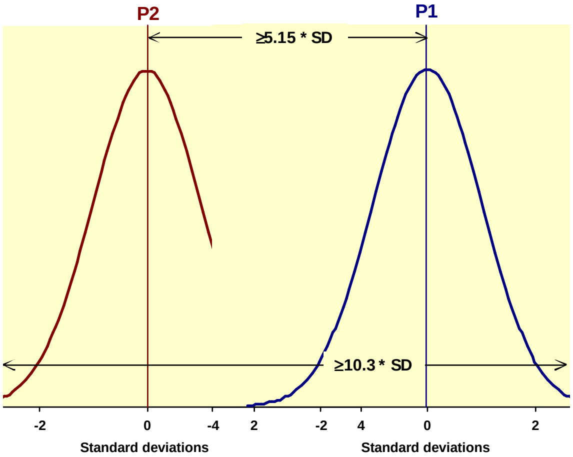
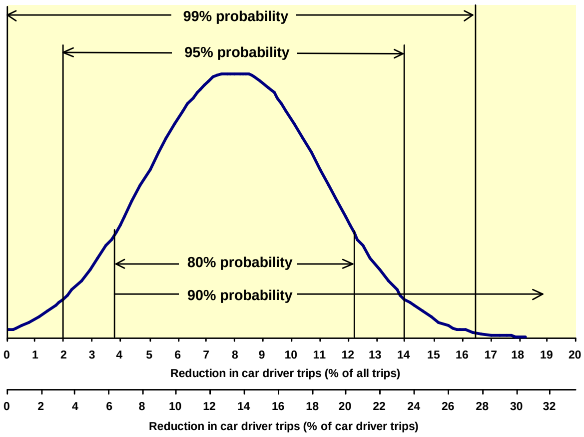


Figure 7 Sampling distribution of the difference between means



If we redo this for a 95% probability that the difference exceeds zero and for a more 'typical' measured reduction of six percentage points or 10% (from, say, 60% car driver trips to 54%), we get the following:

- The range of estimates will be between zero and 12 percentage points (20% of car driver trips);
- There is an 80% probability of the reduction being between 2.1 and 9.9 percentage points (3.5 to 16.5%);
- There is a 90% probability of the reduction being greater than 2.1 percentage points (3.5% of car driver trips);

There is a 10% probability of the reduction being greater than 9.9 percentage points (16.5% of car driver trips).

Population or Participants?

A specific issue that has received little attention is that intensive interventions may have substantial impacts on particular individuals or households but, because of their intensive or intrusive nature, may achieve lower participation rates. It is essential that comparative evaluations are on a like-for-like basis and, given that the strategic objectives for travel behaviour change are population-based, reflect the outcomes across the identified target population as whole, not just for the participating individuals or households.

IndiMark® reports travel behaviour change across the whole identified target population, including those who do not actively participate. This approach also facilitates the use of control groups (qv) not affected by the intervention. Although one would not normally expect external factors to have a major impact on travel behaviour over the typical 12-month period between the 'before' and 'after' surveys, the rapidity of petrol price changes in recent years is an exception to this.

Travel Blending® and other approaches that use travel diaries as an integral part of the process, are only able to report outcomes across the participating population – and specifically only those who complete both before and after travel diaries. Because the participants self-select through their choice to be part of the program, and there is a further stage of self-selection in the 'loss' of participants between the Stage 1 and Stage 2 diaries, the only measurement is of a demonstrably non-representative sample. Perkins (2002) estimated that the reported results of Travel Blending in South Australian pilot projects would have to be factored down by 0.6 to estimate population-wide outcomes on a basis comparable to *IndiMark®*.[5]

CONCLUSION

Voluntary travel behaviour change (VTBC) aims to reduce car driver trips without either specific investment in physical infrastructure or transport services or regulation of transport activity (including pricing). The principal techniques used in the household context differ in whether they deal with the whole identified target population, the means of identifying those who participate actively, the methods of participation and the way of assessing the extent of travel behaviour change.

Some interventions (such as *Travel Blending®* and *Living Neighbourhoods®*) work only with self-selected participants and measure the behaviour change for those participants only. Others (primarily *IndiMark®*) measure behaviour change across the whole identified target population. This difference has led to some confusion in the literature, with *IndiMark®* results being wrongly 'scaled back'.

VTBC is almost unique among transport initiatives in that it has been developed from a strong theoretical and observational basis, developed through a series of interventions with proper experimental design (including control groups) and has been subject to comprehensive monitoring and evaluation of outcomes with the process and outcomes being widely documented in the public domain. Despite (or perhaps because of) that, it has been subject to more intensive scrutiny than any comparable development. Much of this has focussed on the sample size requirements for measuring quite small changes with a satisfactory degree of statistical reliability (usually 95% probability).

In practice, measured estimates of travel behaviour change have consistently been in the range 5-15% reduction in car driver trips. This consistency, repetition of results from successive applications and the cumulative sample size now achieved is generally accepted to have successfully countered any doubts about effectiveness based on the method of measurement.

---

5 This is not a criticism of *Travel Blending®* alone. It is an observation of the difficulty of identifying policy-relevant outcomes from interventions that have the measurement tool as an integral part of the process. Behaviour change can only be estimated for those who participate right through the process and cannot be extrapolated to any larger group.

The most straightforward way of overcoming any remaining issues of statistical reliability in measuring behaviour change for an individual VTBC initiative is to undertake very large-scale interventions. This reduces the relative sample sizes required to achieve a specified level of statistical reliability and, hence, the cost of monitoring relative to delivery. A recent application in North Brisbane, to 74,500 households out of an area population of 113,000 households, has demonstrated that the change in car driver trips is significantly different from zero with 99% confidence. This is equivalent to a 90% probability that the reduction in car driver trips exceeds 6.4% (or 3.8 percentage points).

Very large-scale interventions, appropriately designed, also create substantial opportunities for achieving travel behaviour change through diffusion at little or no cost to the project. Further research into the extent of such diffusion would be useful.

Confidence in the estimates of outcomes is further enhanced by consistent high survey response rates, in the case of *IndiMark®*, which minimises the effects of non-response bias.

Further development is required of the potential for direct or indirect measurement of car driver trips/travel as an alternative to surveys, to provide robust estimates of the primary travel behaviour outcome. This might be achieved through measurement related to household vehicles rather than to individual members of households. However, such approaches do not automatically overcome issues related to sampling and non-sampling errors associated with surveys.

Further research would also be desirable into the extent of induced car travel resulting from the reduction car traffic generated by the population of the intervention area. In particular, this should focus on the circumstances in which such induced traffic might be significant.

REFERENCES

ABS (2007). *Census 2006: Community Profiles*. Australian Bureau of Statistics: Canberra, ACT. http://www.abs.gov.au/websitedbs/d3310114.nsf/Home/census

AGO (2006). *Evaluation of Australian TravelSmart Projects in the ACT, South Australia, Queensland, Victoria and Western Australia 2001-2005*. Australian Greenhouse Office, Department of Environment and Heritage: Canberra, ACT. http://www.travelsmart.gov.au/publications/pubs/evaluation-2005.pdf

Bonsall P (2007). *Does Individualised Travel Marketing Really Work?* Institute for Transport Studies: Leeds, UK.

Brög W and Erl E (1999). 'Systematic Errors in Mobility Surveys'. *Proc 23rd Australasian Transport Research Forum*: Perth, Western Australia. http://www.patrec.org/atrf/index.php?forum=12&showResults=Search&hdnShowResults=Yes&session=0&surname=brog&title=Systematic

ITP (2007). *Making Personal Travel Planning Work*. Integrated Transport Planning Ltd FOR Department for Transport: London, England.

Ker, I R (2003). *Travel Demand Management: Public Transport Business Case*. ARRB Transport Research FOR Department of Infrastructure: Melbourne, Victoria.

Ker, I R (2004). *TravelSmart Household (Individualised Marketing): Evaluation*. ARRB Transport Research FOR Department for Planning and Infrastructure: Perth, Western Australia.

Ker, I R (2007). *North Brisbane Household TravelSmart: Peer Review and Evaluation – Context, Methodology and Values*. CATALYST FOR Brisbane City Council, Queensland Transport and Australian Greenhouse Office: Brisbane, Queensland.

Ker, I R (2008). *North Brisbane Household TravelSmart: Peer Review and Evaluation – Appraisal of Outcomes*. CATALYST FOR Brisbane City Council, Queensland Transport and Australian Greenhouse Office: Brisbane, Queensland.

Ker, IR, Lloyd, R, Ringvall, K and Sidebottom, A (2005). *Effects of TravelSmart on Urban Congestion: Discussion Paper*. ARRB Consulting FOR Department of Infrastructure: Melbourne, Victoria.

Maunsell Australia (2004). *Travel Behaviour Change Evaluation Procedures and Guidelines: Literature Review*. Maunsell: Melbourne, Victoria.

Maunsell Australia (2004a). *Travel Behaviour Change Evaluation Procedures*. Maunsell Australia Pty Ltd, Pinnacle research and Booz Allen Hamilton FOR Transfund New Zealand/EECA. http://www.ltsa.govt.nz/funding/travel-behaviour/technical-report.pdf

MVWRAVV (2007). *Mobiliteitsonderzoek Nederland 2004 – 2007*. Ministerie van Verkeer en Waterstaat Rijkswaterstaat Adviesdienst Verkeer en Vervoer: Rotterdam, Netherlands

Nürnberg (2006). Die Bewertung eines funktionierenden ÖPNV für eine funktionierende Kommune im internationalen Vergleich, Nürnberg und seine VAG: Nürnberg 26.10.2006

Perkins, A (2002). *The Greenhouse Abatement Potential of Travel Behaviour Change Initiatives*. Transport SA and Transport WA FOR National Taskforce on National Greenhouse Strategy Measure 5.3 – Best Practice in Transport and Land Use Planning. Available form Transport Queensland: Brisbane, Queensland.

Richardson, A J (2005). *TravelSmart Communities Project 2004 – Evaluation Report*. The Urban Transport Institute FOR Department of Infrastructure: Melbourne, Victoria.

Richardson, A J, Seethaler, R K and Harbutt, P L (2003). 'Design Issues for Before and After Surveys of Travel Behaviour Change'. *Proc 26th Australasian Transport Research Forum*, Wellington, New Zealand. http://www.patrec.org/atrf/index.php?forum=35&showResults=Search&hdnShowResults=Yes&session=0&surname=richardson&title=Design

Roth M, Ker I, James B, Brög W, Ashton-Graham C,Ryle J, Goulias K and Richardson E (2003). 'A Dialogue on Individualised Marketing: Addressing Misperceptions'. *Proc. 26th Australasian Transport Research Forum*: Wellington, New Zealand.

Socialdata (2007). *Brisbane North TravelSmart Communities Project – Final Report*. FOR Queensland Transport: Brisbane, Queensland.

Stopher, P (2003). 'Travel behaviour modification: A critical appraisal'. *Proc. 26th Australasian Transport Research Forum*: Wellington, New Zealand. See also Roth, et al (2003) for rebuttal of many of the issues raised by Stopher.

Stopher P, Swann N and FitzGerald C (2007a). 'Using an Odometer and a GPS Panel to Evaluate Travel Behaviour Changes'. Paper presented to *12th TRB National Transportation Planning Applications Conference*. Transportation Research Board. http://www.trb-appcon.org/2007conf/program.html#s4

Stopher P, FitzGerald C, Greaves S and Khang J (2007b). 'Can GPS Measurement Replace Travel Diaries*?' Proc. Transport – the next 50 Years Conference*.

UITP (1998). *Switching to Public Transport*. Socialdata FOR International Association of Public Transport: Brussels, Belgium.

UK Department for Transport (2004). *A Review of the Effectiveness of Personalised Journey Planning Techniques*. Department for Transport: London, England.

Young, H P (2007). *Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning*. Center on Social and Economic Dynamics Working Paper No. 51, Center on Social and Economic Dynamics at Brookings, Brookings Institution: Washington DC, USA. http://www.brookings.edu/papers/2007/10_diffusion_young.aspx#_ftn1 (accessed 8 November 2007).