Mixed Method Data Collection in Travel Surveys: Challenges and Opportunities

Elizabeth Ampt[1], Peter Stopher[2],
[1] Steer Davies Gleave, Adelaide, SA, Australia
[2] Institute of Transport and Logistic Studies, University of Sydney, Sydney, NSW, Australia

1 Introduction

The Transport and Population Data Centre (TPDC) has been running the Sydney Household Travel Survey (HTS) as a continuous survey since 1997/8. It collects information on day-to-day travel in the Greater Sydney Metropolitan Region including Newcastle, Wollongong and the Blue Mountains. Based on the face-to-face personal interview method, it has been seen internationally as a leading example of achieving high quality results in terms of response rates and accuracy of trip reporting.

Using this data collection method, the full response rate (all members in the household participated) and part response rate are currently about 55% and 8% respectively out of a net sample of approximately 4,800 households annually. In line with the trend in Sydney and internationally, the response rate has, however, declined significantly over the years since 1981 when the Sydney Household Travel Survey achieved a response rate of about 79% and even since the first wave of the current survey when the overall rates were still around 76%.

This paper examines the issue of decreasing response rates and sets out to find if using mixed methods of data collection would produce better trip estimates and would b more cost effective.

1.1 Intuitive Response

In trying to deal with the problem of decreasing response rates, there is one response that often springs to mind:

> "Adding new options for how people can respond is likely to get a higher response rate (i.e. more people responding) – so that must give a better trip estimate"

While intuitively appealing, as will be shown below, this is fraught with danger, and *can lead to more costs with no more (and sometimes even less) reliability of trip estimates*. In simple terms, this is because by offering different options to complete the survey to more people, we are often increasing uncertainty about the validity of our data – and getting more bias, i.e., worse data.

1.2 The Correct Question

This means that the question, "How can we increase response rates?" is not the correct one, because increasing response rates alone has the possibility of giving us worse data. We need, therefore to answer three much more important questions:

- Will adding new methods of data collection (e.g., telephone interviews or the option of postal responses) achieve better response rates, and perhaps more importantly,
- Will adding new methods of data collection give better trip estimates?
- Will the effort of adding the new methods be cost efficient?

2 Error and Bias

To begin answering these questions, it is important to understand in more detail the characteristics of a sample survey that lead us to assert that simply increasing sample size and response rate does not necessarily improve the quality of the data. Any sample survey suffers from two sources of inaccuracy. The first of these is error, and the second is called bias.

2.1 Error

Error, or sampling error, is present because a sample survey cannot measure every person or every household in the population. Error, however, is a known function of the variability of a measure in the population and the size of the sample. It decreases with sample size. Therefore, we generally strive to reduce sampling error by increasing the sample size. Fortunately, sampling error, while being quite large for very small samples (say less than 100 persons or households), diminishes rapidly, until we reach sample sizes in the thousands, at which point, very large increases are needed in sample size to make significant reductions in error. **Error is predictable and controllable.**

2.2 Bias

Bias, on the other hand, results from either a problem in the measurement device, or a problem of non-response.

2.2.1 Measurement device bias

If we were measuring people's weights in a survey, and we had a scale that consistently gave weights that were 2 kilos too heavy, then all weights measured would be biased, and, no matter how many people were included in the survey, it would not reduce the bias. Indeed, the bias would still be present even if we weighed everyone in the population.

2.2.2 Non-Response bias

Another source of bias can arise from those people who will not respond to a survey. If those who do not respond are identical in all relevant respects to those who do respond, then there is no bias from non-response. However, this is not usually the case. Rather, it is more usually the case that those who do not respond are dissimilar in some relevant ways to those who do respond.

For example, if, in a travel survey, those who do not respond tend to be those who travel a great deal, then estimates of trip rates will be too low from the respondents. No matter how we might increase the sample size, if the frequent travellers still do not respond, we do nothing to reduce the bias.

This is, then, the major difference between error and bias – error decreases with increasing sample size, and is predictable and controllable, while bias can exist even in a census, is not necessarily reduced by increasing sample size, is not predictable, and is only controllable to a very limited degree.

2.3 Error, Bias, and Response

When response rates fall, as is happening almost around the world in most types of surveys in the early 21st century, there are usually two possible reactions by the agencies responsible for collecting data.

2.3.1 Approach 1: Recruit more respondents

The first reaction is to maintain the sample size by attempting to recruit more respondents, so that the sample size does not fall, even though response rates may be falling. Thus, if a response rate to a household survey was 60 percent two years ago, and a sample of 3,500 households was required, then approximately 5,835 households would have been approached to produce the desired sample. If, today, the response rate has fallen to 55 percent, then this strategy would require that, instead of approaching 5,385 households, we would now approach 6,365 households, which should still allow us to maintain the sample size of 3,500 households. What this strategy has done is to enable us to retain the same level of sampling error, because the sample size is still the same.

However, because the response rate has dropped, there will be a concern that this has also meant that the response bias has increased. This is because it must generally be assumed that the 45 percent of households who are now not responding are different in ways that are important for the survey from the 55 percent who *do* respond. Furthermore, it is likely that the additional 5 percent of non-respondents are not the same as the remaining 55 percent of respondents, so that the bias is now worse than before.

2.3.2 Approach 2: Use additional survey methods

The second approach that might be taken is to consider using additional survey methods or modes. This might be argued on the grounds that the people who respond to one type of survey differ in some way from the people who respond to another type of survey. In household travel surveys, it is generally thought that people who do not respond to telephone and face-to-face surveys are more likely to include those who travel more than the average (they are more rarely at home and, therefore, less able either to be interviewed at the door, or contacted by telephone). Therefore, it may be decided to try to make up the sample by introducing a second survey method, such as a postal survey, in an attempt to maintain the sample size and increase the response rate.

Thus, this strategy is seen as one in which response rate may be increased, and the bias reduced to some degree. However, in this simple example, this strategy will only work to reduce non-response bias if the postal survey is actually more likely to be responded to by those who travel a great deal. If this is not the case, and there are other biases in the postal survey, then, while response rate might be increased, the bias may either be unchanged, or even increased. The following example might help to illustrate this.

Suppose that a face-to-face survey is being undertaken, and it is known that this method of survey tends to be biased against frequent travellers. Suppose, now, that a telephone survey option is added to the overall survey procedure, such that those who do not respond to the face-to-face survey are given the option of responding to a telephone survey. Suppose that the telephone survey is known to be biased against the elderly and the young, and is also biased against frequent travellers. In other words, those who are in their teens and twenties, and those who are in their seventies and beyond are more likely not to respond to the telephone survey, as are those who travel a great deal. What is happening is that we have simply introduced another, different bias, into the respondent set, without potentially affecting the original non-response bias in the face-to-face survey. Indeed, the new survey responses are still biased against frequent travellers, but have also introduced a bias against the young and the old.

In fact, it is important to note here that, even if the face-to-face survey in the example just cited, was collecting data from the elderly and the young, when the data from the telephone survey are combined with the face-to-face survey, a new bias has been introduced. This is because the overall sample now has under-reporting by the elderly and the young, and thus the expectation that the overall sample has become more biased than it was. If it is simply a bias of under-representation, this can be corrected by weighting the resulting sample.

However, it is much more likely that the bias would be such that those elderly who do respond to the telephone survey are more mobile than those who do not, and that those younger adults who respond are less mobile than those who do not. In this case, the added bias from the telephone survey will cause estimates of trip-making from elderly households to be biased upwards, and estimates for young households to be biased downwards.

As noted earlier, there are two principal sources of bias in a survey. One is non-response bias and the other is measurement bias. We know, from anecdotal evidence at least, that there is non-response bias in travel surveys. This is known to take the form, generally, of biases against single people living alone, those who travel very little and those who travel a great deal, the elderly, and those with lower incomes and lower education levels. Table 1 gives an example of different responses (and, hence, probably different levels of non-response bias) for different types of surveys.

Table 1 Response Rates

Survey	Face-to-Face	Telephone	Postal
HTS, Sydney 2002[1]	67%		
VATS Pilot, Melbourne 1994[2]	66%		52%
Grenoble, France 1985[3]	73%	65%	
Switzerland, Microcensus, 1994[3]		70%	
Switzerland, Microcensus, 1989[3]			63%
Dutch National Travel Survey, 2000[5]		44%	74%

However, there are also measurement biases, in that different survey methods tend to produce different results (see Table 2). We know, for example, that people tend to underreport their travel in a survey, partly because the amount of travel they report is directly correlated with how long the interview takes or how much time it takes to fill out a self-completion form. Because people are generally very sensitive about their time, they will tend to leave out reporting of what they consider to be less significant travel, if they perceive that this will shorten the time of the interview. As a result, face-to-face interviews tend to collect the most trips (face-to-face interviewers who are skilled in their task know that people will try to skip some trips, so probe for missed trips), telephone interviews will collect the next highest (probing over the telephone is more difficult to do), and postal surveys probably collect the least trips (there is no interviewer to probe, and the respondent is left to do as she or he chooses). This represents a measurement bias. Therefore, if we introduce a different survey method, as a means to increase response rate, we may also now introduce a measurement bias, in addition to the potential of adding a new response bias.

Table 2 Trips (stages) per person

Survey	Face-to-Face	Telephone	Postal
HTS, Sydney 2002[1]	4.2		
VATS Pilot, Melbourne 1994[2]	5.3		3.5
Grenoble, France 1985[3]	4.2	4.1	
The Netherlands, 1983[4]	3.3	3.7	
The Netherlands, 1984[4]	3.5	3.5	
Dutch National Travel Survey, 2000[5]		3.6	3.1

[1] Transport Data Centre (2002)

[2] Ampt and Richardson (1994)

[3] CETE (1986)

[4] Karsten and Konig (1985)

[5] Van Evert. and Moritz, G. (2000)

3 What have others done?

To summarise the discussion so far:

- We need to ask the right questions: will adding new methods give better response rates, better trip estimates, and be more cost efficient?
- There are usually one of two approaches to getting better response:
  Recruit more respondents, or
  Add more survey methods
- Both have associated risks.

The question arises if anyone has a) asked the right questions (and got answers) and b) dealt with the risks associated with the approaches to getting a better response.

We have found that there is very little literature on either of these two questions. In addition, there does not seem to be a good description of the options for mixing methods.

3.1 Will adding methods give better response, trip rates and cost efficiency?

Several sources suggest that adding one method to an existing method will give higher response rates. A non-transport example is that reported by Dillman et al (2004). This work suggests that there was an improved response rate when the recruitment strategies of telephone, mail, telephone/internet and telephone/interactive voice response (IVR) were used. However, these authors argued that the improved response does not reduce non-response error especially if done at the collection phase and evaluated comparing demographics only.

The HTS Non-Response survey showed that, using a second method (telephone) resulted in a wave of people who had higher trip rates than the average of people responding to the first survey method. If these results are indicative, it would mean that 'last respondents' differ from their early responding household members and that by adding their data to the total, nonresponse error is being reduced. This is not proven, however.

We found no work on cost-efficiency of mixed method surveys.

3.2 Has anyone looked at the effects of mixed method surveys?

We found no evidence that anyone had examined the biases and were hence able to correct the data from surveys using a second method. Bonnel (2003) looked at some method comparisons, although none of the studies he cited really set out to measure the effects of mixed method surveys as such. Bonnel suggests, however, that postal surveys collect fewer trips than telephone – 3.1 compared to 3.6 trips per person per day (using the Dutch National Transportation Survey as an example), and also showed in a survey in Grenoble that face-to-face collected more trips than a telephone survey (4.19 compared to 4.06 trips per person per day). However, this was not a case of using a mixed method survey, but rather a change of method from one year to the next.

Perhaps the closest assessment is work that has been conducted recently in the United States, involving GPS measurement to assess the possibility of under-reporting in CATI surveys (Zmud and Wolf 2003). In this work, it has been found that CATI surveys appear to result in trip underreporting of the order of 20 to 25 percent, but with one case as high as 60 percent. Again, this is not a comparison of mixed methods, but rather an assessment of the accuracy of one method.

To understand the effects of mixed method surveys, it is essential to understand the nature of non-respondents. Only in this way will it be possible to make clear if the addition of respondents is creating more reliable data. To assist in understanding non-response, we have gathered the results of non-response surveys.

3.2.1 Non-response – Denver 1997

A non-response survey was devised to gain some insights of non-responding households to the 1997 Denver Region Travel Behavior Inventory Household Travel Survey. The survey concentrated on quick-refusal and no-contact households.

It was found that the refusal and non-contact households were:

- People in very low and very high income households
- Unemployed persons
- Persons with less than a high school education
- High and low mileage drivers
- Persons who do not use a motor vehicle for travel
- People in no-auto and high-auto owner households
- Older persons
- Young single people
- People in one and four or more person households

Households that responded only partially were also found to be statistically significantly different to responding households on many demographic characteristics. Interestingly, in the Denver survey, the refusers and non-contacts did not appear to exhibit statistically significantly different trip rates overall, although differences in distribution of trips by purpose were not investigated (DRCOG, 2000). On the other hand, the partial responses showed very significant differences in trip rates.

3.2.2 Non-response survey – Sydney 2001

Another non-response survey was conducted in Sydney, 2001, by the Transport Data Centre, NSW Department of Transport, to investigate non-response and its effects on data quality, in relation to the Sydney Household Travel Survey, as well as to test the telephone as an alternative data collection method to the costly face-to-face interview (TDC, 2002). Households that could not be contacted after at least five visits (non-contacts) and those that still refused after refusal conversion was attempted, were moved into the Non-Response Study. A full HTS telephone interview was offered first, if the main reason for non-response was unavailability for a face-to-face interview. If the non-respondents still declined, a shorter Person Non-Response Interview was offered. This only collected core demographic and trip information. If the non-respondent did not want to complete the Person Non-Response Interview form, a Person Non-Interview form was offered; information was collected by proxy. From the results of this study, TDC was unable to state with any confidence the relative accuracy of the telephone interview data to that of the personal interview (regular HTS), due to the insufficient sample size (TDC, 2002). However, the results of the non-response study conducted by TDC are useful for providing some insight into the characteristics of non-respondents to a face-to-face interview.

From the TDC survey, it was found that non-respondents were more likely to:

- Live in flats or rented properties (it was not indicated if they are in security buildings, but, if so, then face to face would not be effective);
- Be between 15 and 49 years of age; and
- Undertake more travel by train and walk.

A study done some years earlier for the Melbourne VATS could provide no information on socio-demographic characteristics or trip-making. Information was provided only by geographic area in Melbourne and Brisbane. Non-response in that survey appeared to be lowest in inner suburbs.

A very recent study by the Institute of Transport and Logistics Studies at the University of Sydney (ITLS, 2005), and undertaken in the USA also did not provide information on the comparative demographics of respondents and non-respondents or on comparative trip rates. However, it did show that major reasons given by non-respondents for not responding related to the time when they were called, and whether or not they perceived that they had sufficient time to complete the survey when asked to do it. In addition, in response to a Stated Preference experiment, it was found that there was a strong preference against e-mail as a recruitment strategy and a strong preference towards telephone and mail recruitment.

When respondents are allowed to choose not to respond, the *when to respond* parameter estimate becomes insignificant. The *how to respond* parameter however remains statistically significant such that respondents are more likely to respond when given the opportunity of selecting how they do so. Further, when respondents can choose not to reply there exists a strong preference for short surveys (less than 10 minutes) but an indifference to longer surveys, *ceteris paribus*.

Some people have argued that the lower trip rates obtained from postal surveys are a result of the high response rates obtained, where they argue that the higher response rates include more people with lower mobility (e.g., the results from the Dutch National Survey, quoted by Bonnel, 2003). On the other hand, the Grenoble results, reported by Bonnel (2003) seem to indicate the reverse, in that the trip rate is higher for the face-to-face survey that also had the higher response rate. It is, therefore, unclear at this point as to whether higher response rates will lead to increasing or decreasing trip rates. Arguments can be made strongly for both cases: that non-respondents tend to be more those who travel little, and therefore think their responses are of less value; that non-respondents tend to be those who travel a lot, and who are therefore harder to find for either a telephone or a face-to-face interview. The truth is probably that both effects are present, and that the proportion of frequent travellers and infrequent travellers in the population will impact the results in any particular case.

3.3 What options are there for mixing methods?

We did not find a clear list of the options for mixing methods. We have, therefore, summarised the types of survey method mixes that would be possible for travel surveys.

We take for granted the use of a prospective methodology (i.e., letting people know in advance which days will be surveyed) The recall survey (asking people about travel in the past) has already essentially been rejected as early as 1981 (Clarke et al., 1981) because of the high levels of underreporting inherent in this method.

To understand what mixed method means, it is useful to think of the travel survey process as having five components:

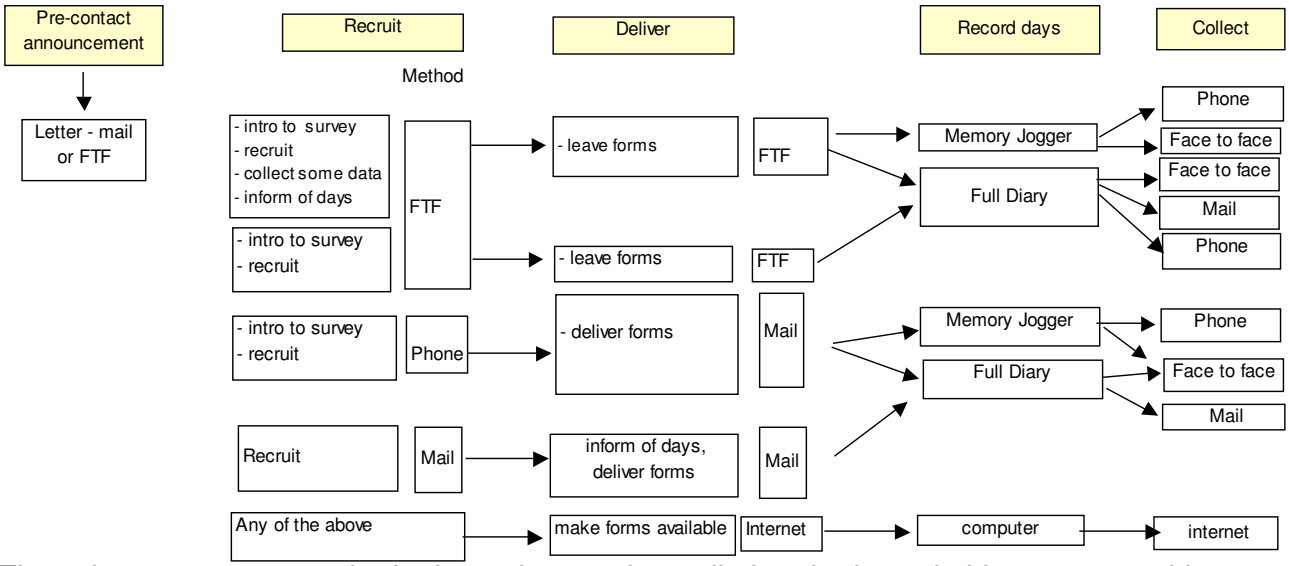
- Pre-contact announcement (pre-notification);
- A recruitment stage where people are informed of the project and asked to participate
- A delivery stage where people are informed of the travel days, the forms are delivered, and data (e.g., about the household) are sometimes collected;
- The time when people record travel; and
- The collection of the data

Figure 2 shows the five stages and the options for different ways of collecting the data at each stage. There are basically two ways in which methods can be mixed:

- Provide respondents with a choice of collection method at the time of recruitment; or
- Provide respondents with an option for an alternative collection method if the arranged (core) method did not result in a response.

However, a further variant on either of these two methods is either to offer the alternatives at the level of the entire household, or to offer the alternatives to each member of the household. The Sydney Household Travel Survey has been using method 2 at the within-household level for the past one or two years. In this, the core method is a face-to-face interview. However, when responses are not obtained from one or two household members within a reasonable time, the alternative of telephone retrieval of data is offered to the non-responding members of the household.

Figure 1 Schematic of Mixed Method Survey Options



Thus, the pre-contact can be by letter that can be mailed to the household, or presented in an interviewer visit. Recruitment can be undertaken by face-to-face contact, telephone or mail, and may involve just the recruitment activity, or an introduction to the survey and recruitment, or both of these together with the collection of some data, and information about the diary days. Delivery is usually conditioned on the method of recruitment. If the recruitment is face-to-face, the delivery of forms will usually be face-to-face, at the time of recruitment. If recruitment is by telephone or by mail, the delivery of forms will usually be by mail. If the survey is to be conducted by internet, then the URL for the internet site can be provided by any of the above means, or by an e-mail.

For recording travel, there are two primary options – a memory jogger or a full diary. Either one of these can be used with any method of recruitment and delivery, except the internet option, where it normally must be a full diary provided on the web. The memory jogger requires an actual interview to collect full travel details, and may be accomplished by either face-to-face or telephone interview. The full diary can be collected by face-to-face interview, telephone interview, or mail. A full diary is an option for any method of recruitment and delivery. However, a memory jogger alone cannot be used with a mail out/mail back survey, because it will fail to collect the detailed travel information required.

By looking at Figure 1, it is possible to see how survey methods can be mixed. Essentially, a survey that is recruited by face-to-face contact can be collected by any of the survey methods. A survey where recruitment is done by telephone can also be collected by any survey method. A survey where recruitment is done by mail can also be collected by any

method, although it requires a full dairy and not a memory jogger. There is, therefore, extensive opportunity to mix methods in a survey, because there are few incompatibilities in the process, and the various collection methods are nearly independent of the method of recruitment and delivery of forms or other survey materials.

It should be noted that Figure 2 does not comment on whether these methods give better response or trip estimates, but is useful to understand options.

4 What Should be Done?

This paper shows that it is extremely difficult to answer the question as to whether or not it is worthwhile to add a second or third method to the Sydney HTS. Without better knowledge of the measurement bias in each method of surveying, and without better knowledge of the non-response bias, it is not possible to answer the question of whether or not it is worthwhile. There is also only relatively poor information on the relative costs of each survey method. Therefore, we propose that a test should be undertaken to ascertain the differences in the methods, and to document the relative costs of the procedures.

4.1 A Pilot Test

This section describes the way in which one might develop a pilot test to discover the differences in the methods. First, a specific suburb should be chosen for the test, so that there can be reasonable control of the samples that would be drawn. The suburb should ideally be one in which there are 25,000 or more households, in order to ensure that there is a sufficient population from which to draw the needed samples. We recommend that four independent cross-sectional samples should be drawn from the suburb, each of which would consist of about 400 households.

One of the problems that arises with a comparative test such as we propose in this section is that one does not know what the survey should have measured. However, there is now available a potential method to do that – the Global Positioning System (GPS) survey (Stopher, Greaves, and FitzGerald, 2005). In such a survey, respondents would carry “wearable” GPS devices with them for a week or so, which would record all places they visit during that time. This comes as near as modern technology will allow to being able to measure “true” travel.

The first sample would be drawn and households would be approached with the current method (face-to-face), followed by face-to-face completion of the survey, all using the standard Sydney HTS approach. One difference would be that 100 of these households would be recruited to undertake a GPS survey, in addition to the regular face-to-face survey. The GPS survey should span a period of one week, with the face-to-face diary day being at the end of the period. The aim would be to achieve a total sample of 200 households completing the face-to-face survey, with 100 of them also completing the GPS survey.

The second sample should be drawn and households in this sample should be recruited using the same face-to-face procedure, but this time with the survey to be completed by a computer aided telephone interview (CATI). As with the face-to-face survey, the goal would be to recruit a total of 200 households who would successfully complete the CATI survey, of which 100 will also complete a GPS survey for a period of a week.

The third sample should be drawn and households in this sample should be recruited using the same face-to-face procedure, but this time the survey would be completed by a postal, self-administered survey, which would be dropped off at the time of the recruitment visit. As with the previous two surveys, the goal would be to recruit a total of 200 households who

would successfully complete the postal survey, 100 of whom would also complete a GPS survey for a period of a week.

The fourth sample should be drawn and households in this sample should be recruited using the same face-to-face procedure, but this time with the survey to be completed by internet, where the URL for the internet survey would be provided at the time of recruitment. As with the previous three surveys, the goal would be to recruit a total of 200 households who would successfully complete the internet survey, 100 of whom would also complete a GPS survey for a period of a week.

With respect to sample size, we have found that the error level in a one-week GPS survey of 100 households is about equivalent to that of a one-day sample of 600-700 households. Therefore, in determining a correct trip rate, the GPS samples would give a relatively high level of accuracy. The issue with the overall sample sizes is to determine what level of error is acceptable with respect to differences in the results of the different methods. In addition, it cannot be stated with certainty what levels of error will arise with respect to distributions on household characteristics, until these are known for the suburb in question. However, the sample sizes of 200 per method would provide statistically sound information, but may require further exploration, depending on the results.

4.2 Purpose of the Pilot Survey

The purpose of this pilot survey would be twofold (see Figure 2). First, by examining the distributions of respondents on primary attributes, such as household size, car ownership, income, number of workers, etc., there would be a better idea of the non-response biases of each of the three methods. From this, it would be possible to see whether the adding of survey methods would tend to reduce or increase non-response bias. Second, by comparing the GPS and diary results, it would be possible to determine how complete the reporting is on trips, and posit correction factors for different survey methods that would produce comparable trip rates. This would answer the questions about measurement bias.

Ideally, there would also be a non-response survey for each of the four methods (see Figure 1). This would make it possible to give further information on the types of people being omitted and would give further clues on non-response bias. However, it is important to note that there is no existing example of a non-response survey in the transport field that has been sufficient to permit estimates of bias correction to be undertaken. The results of non-response surveys tend to be more in the direction of identifying the nature of the biases, some of which may then be correctable from secondary data sources, and others of which will not be correctable at all. The greatest difficulty is that it is usually not possible to obtain data on the actual trip making of non-respondent households, because it was often the attempt at collection of such information that originally caused the non-response action.

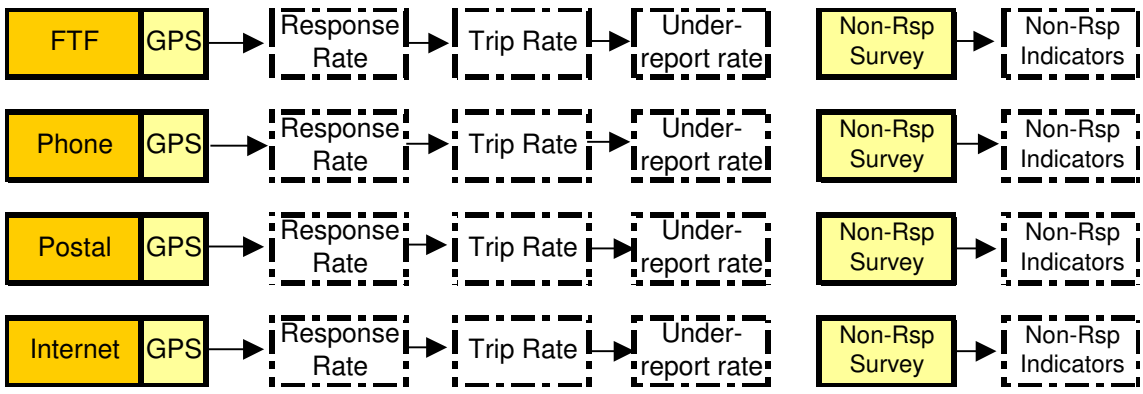


Figure 2 A model for a pilot survey

Once there are results from a pilot survey it would be possible to calculate whether increasing the responding sample using new methods will give better trip estimates in a cost efficient way. Table 3 gives a worked example to assist in understanding the process. Assume the following results:

Table 3 An hypothetical worked example to evaluate pilot results

Method	Response Rate	Trip Rate	Under report rate	Characteristics of Non-Respondents	Relevance of
FTF	65%	5	2%[6]	High car ownership	High travel
Phone	60%	4	20%[7]	High car ownership	High travel
Postal	50%	3	25%[8]	Low car ownership	1 person hhs

The under-reporting rate would be deduced by examining the difference between the trip rates given in the face-to-face, phone or postal surveys and that measured by the GPS surveys. The figures for face-to-face and phone surveys are ballpark from the RTA GPS surveys and results reported from the US (Kurth et al, 2001). Response rates are reflective of recent experience in Australian surveys. The current Sydney HTS is achieving about a 65 percent response rate. The postal survey being conducted by ITLS at present for the NSW DIPNR achieved better than 60 percent response rate. Experiences from the U.S. and VATS suggest that the telephone survey would be lower than face-to-face, but higher than the postal survey. Figures for the use of these as a supplementary method are not published anywhere in the literature, so the above rates are the only ones that can be used at present.

The characteristics for non-respondents are fictitious and are based on anecdotal experience, but should be regarded in this case as illustrative, rather than factual. We have used car ownership as a bias characteristic here intentionally so that the illustrative example shows a potential for one survey to be biased in the opposite direction to another. Merits of the different survey methods cannot be deduced from this example. Unfortunately there is little factual information available from non-response surveys about the biases between respondents and non-respondents. Neither the VATS non-response survey, nor the ITLS non-response survey addressed this issue, so no results are forthcoming from them. The Denver survey (DRCOG, 2000) is the only one to report on this for a telephone survey, and is not considered a reliable source for this. For this reason, we have proposed that a non-response survey should be conducted as part of the recommended pilot test.

With these results, the following decisions **might** occur.

- There is no point in adding phone to the face-to-face methodology because it does not substantially increase response rate, it has a higher under-reporting of trips, and non-respondents are similar to those non-respondents of the face-to-face surveys. (Note, again, this is a fictitious example, so the conclusion is equally fictitious, but illustrative of how one should assess the use of mixed methods.)
- There would be reason to consider adding postal surveys to the survey because, although their response rate is lower, and their under-reporting rate higher (but known), they are gaining people with high car ownership who are being missed in the face-to-face survey. They are, of course, introducing a new bias (less response by 1-person households), but this level would also be known.

[6] Based on preliminary results from the RTA Project measuring the under-reporting of the Sydney HTS by the Institute of Transport Studies.

[7] Based on unpublished results from the U.S.A. for CATI surveys.

[8] Based on average trip rates compared between postal and CATI surveys (no GPS verification has been performed on postal surveys)

There is one final decision, however. The costs of the addition of another method (in this case a postal option) need to be considered. They will include:

- Redesign of the questionnaire to a format for self-completion (this would have been tested in the above pilot)
- Printing costs will be higher than for a face-to-face survey because
  - The questionnaires need to be in colour for a good response
  - There will need to be more pages printed (e.g., example pages)
  - The cost of envelopes, accompanying letters and stamps needs to be included.
- Data entry programs may need to be altered
- A weighting process will need to be added to the data analysis phase using the weights derived from the pilot.

Once these costs are estimated, it would be possible to decide if the additional cost is within a 'reasonable budget' and whether the additional methods will be cost efficient.

Table 4 Hypothetical evaluation of cost-efficiency

Method	Cost per household	No. of Households	Total Cost
FTF	$350	1000	$350,000
Postal	$175	200	$34,000

Is it worth a 10% increase in cost to get the high car owning households, and hence to improve the trip estimates?. For example, suppose that the average trip rate from the original face-to-face survey was 3.6 trips per person per day. Suppose that the postal survey produced an overall trip rate of 3.2 trips per person per day, which had been determined as needing to be weighted for underreporting by a factor of 1.3. The average trip rate (corrected) for the postal survey would now be 4.2 trips per person per day. Simply putting the two samples together with no further weighting would produce an average trip rate of about 3.7 trips per person per day. This represents an increase of 0.1 trips per person per day. The question is clearly whether such an increase is worth $34,000. Of course, there are also additional improvements to the accuracy of the resulting trip rates, because adding another 20 percent to the sample will also reduce the sampling error. However, to be able to estimate this requires knowledge of the variance of trip making from each of the two samples (face-to-face and telephone), and there are no reports of this information. To put together a hypothetical illustration using this, also, is highly speculative.

One could go further and suppose that the 1,000 household survey contained only 8 percent of high car-owning households, where the census indicated that it should be 14 percent. Suppose that the postal survey obtained 15 percent of high car owning households. Suppose also that those who were included in the face-to-face survey had an average trip rate of 5 trips per person per day, while those in the postal survey had an average of 5.3 trips per person per day, which, after factoring, is 6.89 trips per person per day. The weighted average trip rate of the two methods of the survey for high car owning households would now be estimated as 5.5 trips per person per day. If we had not used the mixed-method survey, then the total contribution of trips by high car-owning households to the average trip rate would have been 0.4 trips per person per day (8 percent of 5 trips per person per day). Knowing that these were under-reported, the contribution could have been corrected to 0.7 trips per person per day (14 percent of 5 trips per person per day). With the addition of the

postal survey, however, we know that the correct contribution should be closer to 0.77 trips per person per day (14 percent of 5.5 trips per person per day). Again, the question is whether it is worth the expenditure to get this correction.

Once it is possible to give this level of information, it makes it relatively easy to decide whether there are enough funds, or whether it is considered worthwhile to spend the time and resources on increasing the reliability of the trip estimates.

4.3 How to Adjust for Underreporting

Assuming that the procedures described herein were applied, and from these procedures, a knowledge gained of the extent of underreporting, the next question would be how to use that knowledge to adjust the results from the survey. This is a far from trivial task, and one that has not been adequately researched or demonstrated to date. Indeed, the use of GPS to assess underreporting of conventional household travel survey methods is still a research issue awaiting resolution.

At one level, if it is established that certain subgroups of the population are underrepresented in a particular method of survey, and that this subgroup has a trip rate that is also incorrectly estimated, then one could propose to correct this. Correction would be done by first weighting the sample to represent the population subgroup correctly, and then adjusting the average trip rate of the group with the information obtained from secondary survey methods.

However, there is then another issue as to whether the subgroup of concern is sufficiently homogeneous for this to be an appropriate step. Using the earlier example of a bias against high car-owning households, it may be that these households are further split into average trip-making households, and high trip-making households. The former group are those who were present in the primary survey sample (the sample obtained using the primary survey method), and the latter are the ones that were present in the secondary survey sample (the sample obtained with one or more secondary method surveys). Perhaps, the trip-making propensity of these households is correlated with income and education. In that case, the correct method to adjust the trip rates would be through a cross-classification of these households by car ownership, income, and education.

However, without research into this issue, it is somewhat speculative to state how adjustments should be made. Our recommendation is that research into this issue would need to be carried out along with the collection of the data to support it. It probably is not sufficient just to develop rather simplistic factors to adjust and apply these to the data. In addition, this type of factoring approach will not be helpful for the estimation of disaggregate methodls, where correction is probably required in the disaggregate data. This may require the application of simulation or fuzzy logic to permit a proper correction of the database to be achieved.

5 References

Ampt, E. and Richardson, A. (1994) The Validity of Self-Completion Surveys for Collecting Travel Behaviour Data, Proceedings of the 22nd PTRC European Transport Forum, Transportation Planning Methods, Vol. 2 48-60

Bonnel, P. (2003) Postal*, Telephone and Face-to-Face Surveys: How Comparable are they? In* Stopher, P and Jones P. Transport Survey Quality and Innovation, Pergamon p.215-235.

CETE de Lyon and CETUR (1986) as cited in Bonnel, *Postal, Telephone and Face-to-Face Surveys: How Comparable are they? In* Stopher, P and Jones P. (2003) Transport Survey Quality and Innovation, Pergamon p.215-235.

Dillman, D., Phelps, G., Tortora, R., Swift, K, Kohrell, J. and Berck, J. (2004) *Response Rate and Measurement Differences in Mixed Mode Surveys: Using Mail, Telephone, Interactive Voice Response and the Internet, from website* http://survey.sesrc.wsu.edu/dillman/papers/MixedModeppr_withGallup_POQ.pdf

Denver Regional Council of Governments (DRCOG) and Parsons Transportation Group, Inc. (2000). *Denver Regional Travel Behavior Inventory: Describing and Reaching Nonresponding Populations, Analysis and Project Report*, Report prepared for the Federal Highway Administration, Washington, DC.

Karsten, H. and Konig, J (1985) Differences in Estimate Due to Change in Methods of Data Collection, Centraal Bureau voor de Statistiek, The Netherlands, as cited in Bonnel, *Postal, Telephone and Face-to-Face Surveys: How Comparable are they? In* Stopher, P and Jones P. (2003) Transport Survey Quality and Innovation, Pergamon p.215-235

Kurth, D.L., Coil, J.L., and Brown, M.J. 2001, "Assessment of Quick-Refusal and No-Contact Nonresponse in Household Travel Surveys", *Transportation Research Record 1768*, Paper No. 01-2761, pp 114-124.

Transport Data Centre (2002) *Household Travel Survey Nonresponse Study Report*, internal publication

Van Evert. and Moritz, G. (2000) The New Dutch Travel Survey, 9th International Association for Travel Behaviour Research Conference Proceedings, as cited in Bonnel, *Postal, Telephone and Face-to-Face Surveys: How Comparable are they? In* Stopher, P and Jones P. (2003) Transport Survey Quality and Innovation, Pergamon p.215-235.

Zmud, J. and Wolf, J. (2003) *Identifying the Correlates of Trip Misreporting – Results form the California State-wide Household Travel Survey GPS Study,* 10th International Conference on Travel Behaviour Research, Lucerne.