DEFINITIONS OF UNIT NON-RESPONSE IN TRAVEL SURVEYS

A report emerging from Workshop A2
International Conference on Transport Survey Quality
August 5-10, 2001, Kruger National Park, South Africa

A.J. Richardson,

The Urban Transport Institute
PO Box 363, Alexandra, Victoria, 3714, Australia
email: tony.richardson@tuti.com.au

A.H. Meyburg

School of Civil and Environmental Engineering
Cornell University, Ithaca, New York, 14853, USA
email: ahm2@cornell.edu

ABSTRACT

In all types of travel survey, non-response is a growing problem worldwide. While non-response is not a new problem either in general surveys or in travel surveys, the extent and potential significance of non-response is beginning to be more appreciated as response rates continue to fall. Reducing response rates have two implications. Firstly, they reduce the achieved sample, thereby increasing the sampling error in the resulting parameter estimates. Secondly, and potentially more importantly, they increase the possibility of introducing significant non-response bias, if the characteristics of the non-respondents are not the same as those of the respondents. This paper seeks to address two questions: what is Unit Non-Response and how should we measure it; and what are the consequences of the definition of Unit Non-Response? While there is an understandable focus on trying to improve response rates, in response to an overall decline in response rates, this paper has suggested that such efforts may be counter-productive. Strict definitions of "acceptable response" may result in significant data loss and the introduction of biases. In addition, the specification of required response rates may lead to practices which achieve the desired response rate, but at the expense of data quality.

INTRODUCTION

In all types of travel survey, non-response is a growing problem worldwide. While non-response is not a new problem either in general surveys (Deming, 1953; Brownlee, 1957; Donald, 1960; Kanuk and Berenson, 1975; Steeh, 1981) or in travel surveys (Brög and Meyburg, 1980; Brög and Meyburg, 1981; Stopher and Sheskin, 1982; Wermuth, 1985; Richardson and Ampt, 1994), the extent and potential significance of non-response is beginning to be more appreciated as response rates continue to fall. Reducing response rates have two implications. Firstly, they reduce the achieved sample, thereby increasing the sampling error in the resulting parameter estimates. Secondly, and potentially more importantly, they increase the possibility of introducing significant non-response bias, if the characteristics of the non-respondents are not the same as those of the respondents.

In the transport field, most studies of non-response have concentrated on various aspects of non-response to mailback surveys (Bonsall and McKimm, 1993; Polak et al., 1995; Sen et al., 1995; Richardson et al., 1996; Richardson, 2000). This is perhaps because mailback surveys have traditionally been considered to have the highest rates of non-response. However, CATI surveys often have even higher rates of non-response when all stages of the response process are considered (non-coverage, non-contact, non-response to CATI, non-response to mailout diaries). In addition, even personal interview surveys are now obtaining relatively low response rates. For example, Battelino and Peachman (2001) report that the Sydney Household Travel Survey only obtained a 57% response rate in 1999. Therefore, all types of surveys must consider the issue of non-response in the future.

Response and non-response has been extensively researched in the past, with a number of major books having been published on the topic (Groves, 1989; Biemer et al., 1991; Lessler and Kalsbeek, 1992; Groves and Couper, 1998; Tourangeau, Rips and Rasinski, 2000). In addition, guidelines have been published by the Council of American Survey Research Organizations (CASRO, 1982) and Statistics Canada (Statistics Canada, 1998) which provide commentary on the issue of non-response. Therefore, rather than repeat this material, this paper will address two fundamental issues in dealing with Unit Non-Response (UNR), which are crucial to a thorough understanding of the issue:

- What is Unit Non-Response and how should we measure it?
- What are the consequences of the definition of Unit Non-Response?

The reason for concentrating on these two issues is twofold. Firstly, as noted by Lessler and Kalsbeek (1992), there is a wide variety of definitions of non-response which causes confusion and scope for mis-measurement and mis-use of the calculated response rate. Secondly, as noted by Groves (1989), “non-response rates are often used mistakenly as a measure of the quality of survey statistics”. Commissioning agencies often specify required response rates in the belief that this will assure quality in the responses. However, there are many ways to achieve high response rates that are in fact counter-productive in terms of data quality. This paper examines one relationship between the definition of response rate and the bias in the resulting sample.

WHAT IS UNIT NON-RESPONSE AND HOW SHOULD WE MEASURE IT?

A distinction is often made between Unit Non-Response (UNR) and Item Non-Response (INR). Indeed, the last two International Conferences on Transport Surveys (Grainau and Kruger Park) have had separate workshops on UNR and INR. UNR refers to the situation where no response is received from the sampling unit (e.g. the household) while INR refers to the situation where a response is received but there is data missing within that response. This missing data can refer to individual questions or it can refer to missing trips or missing people within the household. The extent of this INR, where some household members are missing, can be significant. For example, while Battelino and Peachman (2001) report that the Sydney Household Travel Survey obtained a 57% response rate in 1999, they also reported that the partially responding households accounted for another 11% of the sample.

The question must therefore be asked as to when a responding household with some INR becomes a non-responding household. What is the dividing line between INR and UNR? To provide some perspectives on this issue, the major reference books on non-response cited above were consulted to find the definitions of UNR and response rate used therein, in an attempt to further understand the meaning of UNR. Not unexpectedly, there were several variations in definition.

Groves (1989) defines UNR as "the failure to obtain any of the substantive measurements from the sample person or household". Lessler and Kalsbeek (1992) provide a range of definitions given by other researchers and practitioners, but recommend a definition of response rate as "the proportion of participating units among all eligible units in the sample". Tourangeau et al. (2000) define UNR as "the complete refusal to participate". CASRO(1982) acknowledges that many different definitions of response rate exist, and makes the distinction between "response rate" and "completion rate". They define "completion rate" as a "term that is used to designate how well a task has been accomplished", while "response rate" is defined as "a summary measure … used to designate the ratio of the number of interviews to the number of eligible units in the sample". They note that completion rates for various components should be combined to calculate the response rate.

This sample of definitions of response rate raises several points. Clearly, response rate is the ratio of two terms; the numerator is some measure of "actual response" while the denominator is some measure of "possible response". The problem lies in determining the precise definitions of the numerator and the denominator. From the above definitions, it seems that the numerator could possibly be defined as:

- The number of complete responses obtained
- The obtaining of "substantive measurements"

From the above definitions, the denominator could be:

- The total initial number of sample units
- The number of eligible sample units

The definition of the numerator and the denominator in the calculation of response rate is discussed further below.

What is the Numerator in the Response Rate Calculation?

The choice of numerator for the response rate calculation in household travel surveys appears to lie between a “complete household response” and a response which provides the “substantive measurements” required from the survey. In the context of household travel surveys, a “complete household response” is regarded as one in which all members of the household complete and return their travel diaries.

A major problem in defining an acceptable or complete response is that "complete" can be defined on at least six levels in a travel survey:

- Household
- Person within household
- Person diary (whole day)
- Trip chain within a day
- Trip within a chain
- Stage within a trip

Requiring all persons to provide all information about all stages within all trips within all chains is a fairly high standard to achieve. Indeed, there is a real question as to whether the inclusion of the word “household” in “complete household response” is absolutely essential, or whether “substantive measurements” could be defined on a “person” basis rather than a “household” basis. This is especially the case if “complete household” data are not required in subsequent analyses. One reason for obtaining “complete household” information is where subsequent analysis and modelling takes account of the interactions between members of the same household (such as occurs in activity-based modelling of travel behaviour). However, the data from many travel surveys are intended to be used in relatively conventional 4-step modelling procedures. In such modelling, only the first step (trip generation) is done on a household basis, with the remaining steps (trip distribution, mode choice and trip assignment) performed on a person and trip basis. Indeed, there is no absolute reason why even the trip generation modelling is performed on a household basis. The modelling could just as well be done on a person basis, with a separate model being used to aggregate persons into households. Considering that most of the population forecasts used in forecasting are essentially person-based, the use of a person-based trip generation model would appear to be more consistent with the intended usage of the model for modelling and forecasting.

The use of “persons” rather than “households” as the underlying basis of data collection is not new, with France, Norway and Denmark having collected data from only one person per household, Switzerland collecting data from only one or two persons per household (depending on household size) while the US National Travel Survey required the collection of travel diaries from only 50% of the household members for a household to be counted as “responding”. It is not necessary that most travel surveys aim to collect data from only one person per household; the intention could still be to collect travel data from all household members, if possible. What is suggested is that the definition of “complete response” be based on “person response” rather than “household response” and that trip generation modelling be based on person-trips rather than household-trips.

The adoption of a person-based trip generation model would have significant implications for the definition of "substantive measurements" from household travel surveys. If "response" were defined in terms of "person response" and not "household response", then the definition of "response rate" could be relaxed somewhat, with consequent improvements in cost effectiveness and reduction in sampling bias. There would be improvements in cost effectiveness because responses where there was only one missing person would not have to be discarded as "incomplete". This would avoid wasting resources that were used to collect perfectly good data (when viewed from a "person" perspective). There would be a reduction in bias, since "incomplete returns" (using the current definition) are more likely to come from certain types of households (such as large households who make many trips). Removing them from the sample would result in a final sample which was over-represented in terms of smaller households making fewer trips, which would under-estimate total travel. This issue is investigated further in a later section of this paper.

If, however, it was desired to stay with "complete household response" as the numerator in the response rate calculation, then the issues of proxy reporting and imputation of missing data become very important.

With respect to a complete person diary, missing information could potentially be provided by asking another member of the household to complete the diary on behalf of the person who failed to respond (i.e. proxy reporting). Unfortunately, it is well-known that proxy reporting will tend to produce missing information at the lower levels, where complete trip chains are omitted because they are unknown to the person completing the diary. For example, a work-based trip chain at lunch-time is often unknown to other members of the household. While proxy reporting might satisfy the definition of an acceptable "complete" response (because the extent of any missing trips or trip chains would be unknown), it might not provide better data. It may well be better (i.e. less biased) to have a missing person diary than to have an incorrect person diary.

Imputation is commonly used to replace "missing data" with the best "statistical guess" about what that missing data might be. Such imputation can take on many forms (hot-deck, cold-deck, regression-based, stochastic, multiple) and is commonly used to replace missing data for individual variables at the household, person and trip level. If response rates are based on the number of "complete" response obtained, is this definition of "complete" to be considered before or after imputation. If it is to be considered after imputation, are there any limitations on how much data imputation is permitted in order to achieve these targets?

What emerges from the above discussion is that perhaps a more flexible, robust and achievable definition of "complete response" be adopted in order to improve the cost effectiveness of data collection and to reduce the potential bias in the "complete responses". Such a definition would need to be multi-dimensional and based on how the data will eventually be used. It would be based more on "substantive measurements" and less on "complete households", and may require a re-focussing of the modelling effort on person-based modelling and less on household-based modelling. A suggested definition is given below.

At the level of each household, an "acceptable household response" is defined as one where:

- at least a% of travel diaries are provided for each household;
- for responding persons and completed diaries, no missing data on Essential Items of information would be accepted. These Essential Items would be defined for household, person

and travel data, and would be limited to a small number of absolutely essential data items, without which the response would be essentially useless; and

- no more than b% of missing responses for all variables other than those defined as Essential Items (and for this test, missing diaries are not counted).

To maintain control at an aggregate level, across the entire sample of households there should also be:

- no more than c% of diaries missing from members of all households;
- for responding persons and completed diaries, no missing data on Essential Items of information; and
- no more than d% of missing responses for all variables other than those defined as Essential Items (except for the personal income question where e% missing data will be regarded as a maximum).

The values of a%, b%, c%, d% and e% in the above definition of "acceptable response" would need to be determined in conjunction with the client, and would depend on the objectives of the survey and the available resources. As an indication of the likely magnitude of these percentages, however, they could be in the order of 50%, 10%, 5%, 3% and 15%, respectively.

The above definition of "complete response" has combined various elements of UNR and INR. Indeed, a complete appreciation of UNR cannot be obtained without a concurrent consideration of INR.

What is the Denominator in the Response Rate Calculation?

The choice of denominator for the response rate calculation lies between the total number of units in the initial sample or the number of *eligible* units in the initial sample. A basic flaw with using the total number of units in the initial sample is that it does not take account of the number of *ineligible* records on the sampling frame, and confuses the quality of the sampling frame with the calculation of the response rate. Ineligible records on a sampling frame (such as a list of power connections or a similar list of addresses that might be used in a mail-back or personal interview survey) include addresses of vacant blocks, addresses of businesses, and addresses of houses which are vacant at the time of the survey. Similarly, telephone sampling frames contain phone numbers that are not eligible for the survey. The correct measure of response rate is the number of "acceptable responses" divided by the *net* sample size (i.e. the *gross* sample size minus *sample loss*) (Richardson, Ampt and Meyburg, 1995).

The identification of "sample loss" is not equally simple in different types of surveys. Vacant houses are relatively easily identified in a personal interview, since an interviewer actually visits the address and can see whether the house is occupied or unoccupied. In a mailout/mailback survey, such a checking process is not normally available, and vacant houses are only identified by the return of mail by the Post Office. Similarly, in a phone survey, it is sometimes not easy to identify whether a called number actually exists in an eligible household and is simply not being answered, or whether it does not belong to an eligible household.

The extent of ineligible records on the sampling frame, which give rise to so-called "sample loss", varies with the quality of the sampling frame. As an example, in the Victorian Activity and Travel Survey (VATS) (Richardson and Ampt, 1995) which used a list of water connections as the

sampling frame, the observed sample loss (as identified by the return of undelivered mail through the post to the survey office) was identified as being about 6% over the years 1994-98, as shown in Table 1.

Table 1: Observed Sample Loss in VATS 95

VATS Year	Observed Sample Loss
1994	7.7%
1995	6.0%
1996	5.4%
1997	5.3%
1998	4.2%

However, for mail-back surveys there is a special problem with respect to sample loss. In an interview survey, sample loss is clearly identified when the respondent visits an address to conduct an interview, only to find that an occupied household does not exist at that address. With mail-back surveys, sample loss is only partially identified by returns through the mail. Of the "households" that do not respond, it is generally unknown whether the household actually exists and has not responded, or whether the household does not exist at that address and hence could never respond.

In the VATS surveys, some estimate of this unobserved sample loss was obtained from the interviews conducted with non-responding households. These households were visited by an interviewer, and hence it was possible to see whether a household actually existed or not (at least within a sample of non-responding households). The proportion of sample loss among these households for the years 1994-98 are shown in Table 2, and can be seen to be of the same order of magnitude as the observed sample loss for a total sample loss of about 12%. This same effect was observed earlier in the 1992 South-East Queensland Household Travel Survey (SEQHTS) conducted for the Queensland Department of Transport.

Table 2: "Unobserved" Sample Loss in VATS 95

VATS Year	Unobserved Sample Loss
1994	10.4%
1995	5.8%
1996	6.9%
1997	unknown
1998	8.8%

The effect of properly allowing for sample loss in the calculation of response rate in a mailback travel survey is quite significant. A 60% response rate when sample loss is not allowed for is equivalent to a 68% response rate after properly allowing for sample loss of 12%. Clearly, the two measures of response rate would be equal if sample loss were reduced to zero. However, the effort

and resources required to edit the sample frame to reduce sample loss to zero would be substantial, and not worthwhile, if the correct method of calculating response rate were used.

It is therefore suggested that the calculation of response rate be based on "the number of *eligible* households in the initial sample", i.e. gross sample minus sample loss, and not on the "total number of units in the initial sample". In addition, it is suggested that unobserved, as well as observed, sample loss be used to calculate the net sample size. Observed sample loss could be measured in a sample used for non-response interviews, or preferably by other means, if possible, for the full sample of addresses drawn from the sampling frame.

WHAT ARE THE CONSEQUENCES OF THE DEFINITION OF UNIT NON-RESPONSE?

It was noted above that if the definition of "acceptable response" could be relaxed somewhat, this would have consequent improvements in cost effectiveness and a reduction in sampling bias. This section of the paper further explores this issue, and outlines the possible consequences of adopting various measures for the definition of an "acceptable response" in the context of a relatively state-of-the-art mailout/mailback travel survey.

The Underlying Hypothesis

The underlying hypothesis in exploring different options for the definition of acceptable response is that different assumptions will have consequences for the efficiency of data collection and that there may be biases created by the adoption of very tight definitions of response. This hypothesis is based on experience with the VATS survey, where overly rigid "cleaning" of the data set for modelling purposes (whereby any household with any missing data was excluded from the data set) led to the exclusion of 50% of households and to the creation of a very biased data set, consisting of small households that made relatively few trips.

The Method of Analysis

The analysis reported here is based on data from the 1995 VATS survey (it is expected that the use of any of the years of VATS data would lead to essentially the same conclusions). For the purposes of this investigation, the VATS data was reduced to a set of variables with essentially the same structure as the essential data items often required from many travel surveys. The variables remaining in the VATS 95 data for this analysis, and the Item Non-Response Rates for each variable in VATS 95 are as shown in Table 3.

Table 3: Item Non-Response Rates in VATS 95

VATS 95 Variable	Item Non-response Rate
Household	
Household size	0.0%
Dwelling type	1.9%
Ownership of dwelling	2.6%
Number of vehicles	0.8%
Travel Day date	0.0%
Home location	0.0%
Person	
Year of Birth	2.5%
Gender	0.5%
Licence holding	1.6%
Current activity	1.5%
Occupation	2.4%
Industry	2.9%
Personal income	7.1%
Starting place for day	1.9%
Reason for not travelling	0.2%
Travel (Trip Stages)	
Trip start time	2.1%
Trip end time	2.3%
Destination location	0.3%
Trip purpose	0.3%
Mode	0.3%
Number of passengers	1.4%

The analysis then examined every trip stage for every person for every household, and calculated the number of missing items of data within the household. For the purposes of this analysis, the VATS 95 data is based on the data files before any imputation algorithms were applied. Where a person within the household failed to provide any trip information, it was assumed that they would have made an average number of trip stages (4 per day), and that all the variables for these four trip stages were missing. Since there were six variables per trip stage, a missing diary therefore represented 24 missing items of data.

The analysis then simulated the effect of using different criteria to determine whether a response was “acceptable”. At the strictest level, any missing data item would result in an unacceptable response from that household. This criterion was then varied (on a percentage basis) to examine the consequences of allowing x% of the data items to be missing and still allow the household to be counted as acceptable. The calculation of the percent of data items missing was based on the number of data items actually missing divided by the maximum number of data items expected for that household, where there are 6 data items per household, 9 data items per person and 6 data items per trip stage (as shown in the table above).

The consequences of having different definitions of acceptability were measured in five ways:

- ***The percentage of households excluded*:** how many households that responded to VATS 95 would be classified as "unacceptable", given the criteria for acceptability.

- ***The percentage of data items provided in the retained households*:** for those households retained as acceptable using the specified criteria, what proportion of data items are provided across all these households.

- ***The percentage of data items provided from all households*:** for all households classified as acceptable or unacceptable using the specified criteria, what proportion of data items are provided across all these households.

- ***The Household Size Ratio*:** the ratio of the average household size for households retained as acceptable using the specified criteria, divided by the average household size for all households classified as acceptable or unacceptable using the specified criteria. This ratio is a measure of the bias in household size created by the exclusion of "unacceptable" responses.

- ***The Household Stops Ratio*:** the ratio of the average number of Stops (trip stages) for households retained as acceptable using the specified criteria, divided by the average number of Stops (trip stages) for all households classified as acceptable or unacceptable using the specified criteria. This ratio is a measure of the bias in mobility created by the exclusion of "unacceptable" responses.

Results from Two Ends of the Spectrum

The results for the five criteria for a very strict criterion for acceptability (no missing data items in a household) and a relatively lenient criterion for acceptability (20% missing data items in any household) are shown in Table 4.

Table 4: Observed Sample Loss in VATS 95

	% missing data allowed	
	0%	**20%**
% HH excluded	30.0%	3.2%
%data items in retained HHs	100.0%	99.0%
%data items from all HHs	67.2%	96.6%
HH size ratio (retained/total)	95.4%	100.0%
HH stops ratio (retained/total)	97.1%	102.3%

STRICT DEFINITION OF ACCEPTABILITY

When no data items are allowed to be missing in an acceptable household, it can be seen that 30% of households are excluded as unacceptable. That it, 30% of households have at least one item of data missing (where the average number of data items per household in VATS 95 was 115). This figure of 30% is lower than the 50% quoted earlier because the data cleaning described there was performed on the full VATS data set, which contained many more variables. For all households retained as acceptable, 100% of the data items were provided (by definition, since any households with missing data had already been excluded as unacceptable). However, this represents only 67.2% of the data items in all the households classified as acceptable or unacceptable. Thus, the

application of a very strict definition of acceptability has resulted in about one-third of the collected data being discarded.

The exclusion of households and travel data would only be a problem of economics and efficiency (caused by throwing away collected data) if the retained data was representative of the total data set. However, it can be argued that the households that are excluded are more likely to be larger households and more mobile households, since they have more opportunity to have some data missing (since they have more people and more trips and hence more data items in total). This hypothesis is supported by analysis of the VATS 95 data, as shown in Figures 1 and 2.

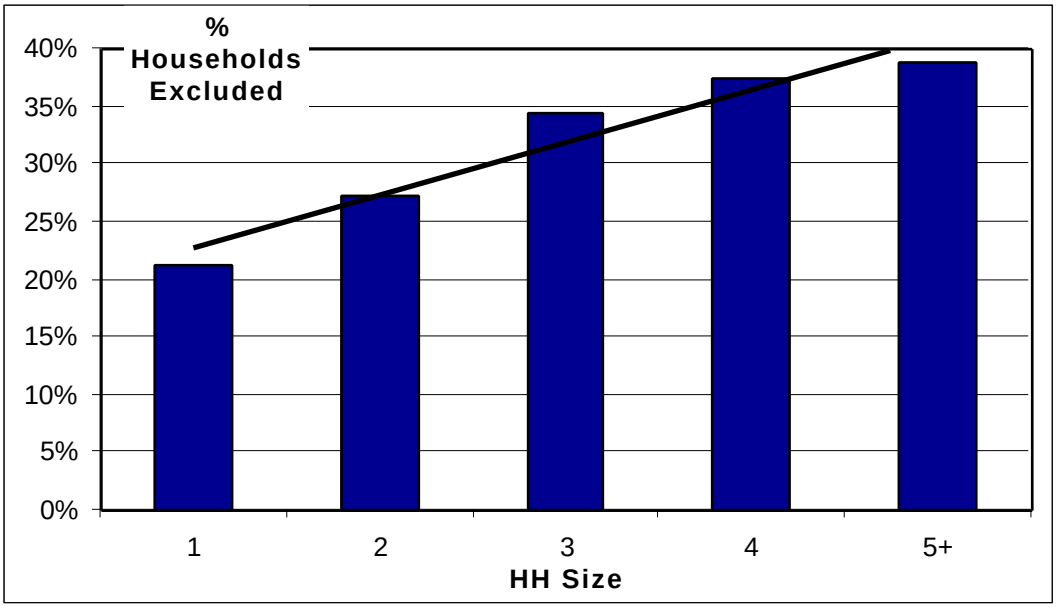

Figure 1: % of Households Excluded as a function of Household Size (0% INR)

Firstly, it can be seen in Figure 1 that while 30% of all households are excluded, this percentage is a function of household size. Only 21% of single-person households are excluded, compared to 39% of households with five or more people in the household. Similarly, this percentage is a function of the total number of stops (trip stages) made by the household, as shown in Figure 2. While the trend is not as strong, the average percent excluded varies from 26% for households with very low mobility up to 34% for households with very high mobility.

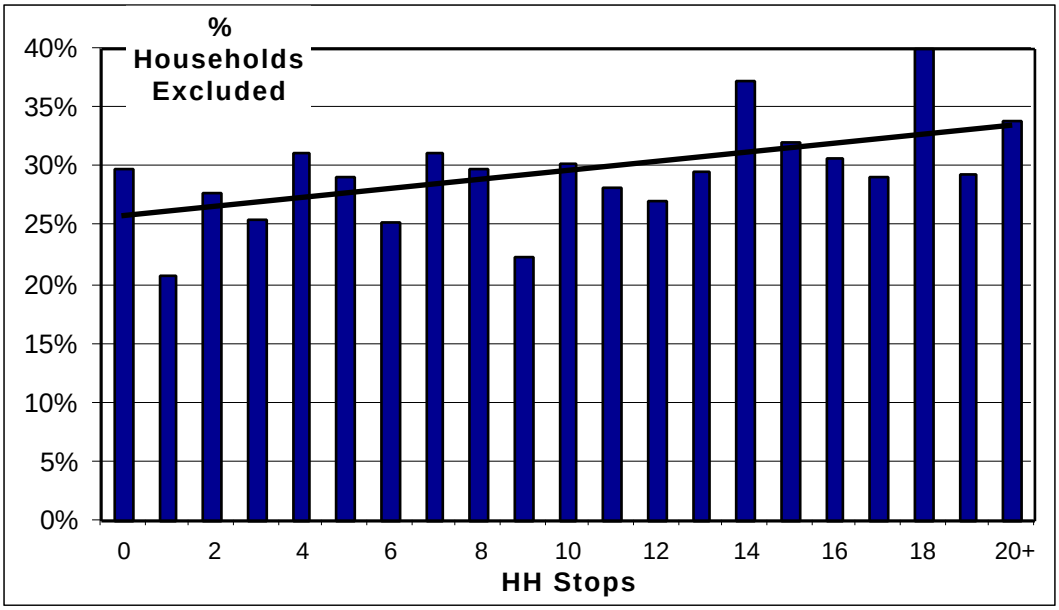

Figure 2: % of Households Excluded as a function of Household Stops (0% INR)

The overall effect of the percent of households excluded and the biases in household size and mobility between retained and excluded households is expressed in terms on the HH Size Ratio and the HH Stops Ratio. It can be seen that with a very strict definition of acceptability, the average household size of retained households is only 95% of the average household size of all households (i.e. a bias toward smaller households), while the average number of stops in retained households is only 97% of the average number of stops in all households (i.e. a bias toward less mobile households).

LENIENT DEFINITION OF ACCEPTABILITY

When up to 20% of data items are allowed to be missing in an acceptable household (i.e. 20% INR), it can be seen in Table 4 that only 3.2% of households are excluded as unacceptable. That it, 3.2% of households have more than 20% data items missing (where the average number of data items per household in VATS 95 was 115). Households with this much missing data often have a missing diary from one or more household members. For all households retained as acceptable, however, 99.0% of the data items were provided across all these households. That is, even with such a lenient definition of acceptability most of the retained households have almost complete data. Overall, these households represents 96.6% of the data items in all the households classified as acceptable or unacceptable. Thus, the application of a relatively lenient definition of acceptability has resulted in only 3.4% of the collected data being discarded. Thus while the "number of complete responses obtained" was relatively low (70%), the amount of "substantive measurements" was very high (96.6%).

The extent of bias in household size and household mobility in the retained households is shown in Figures 3 and 4.

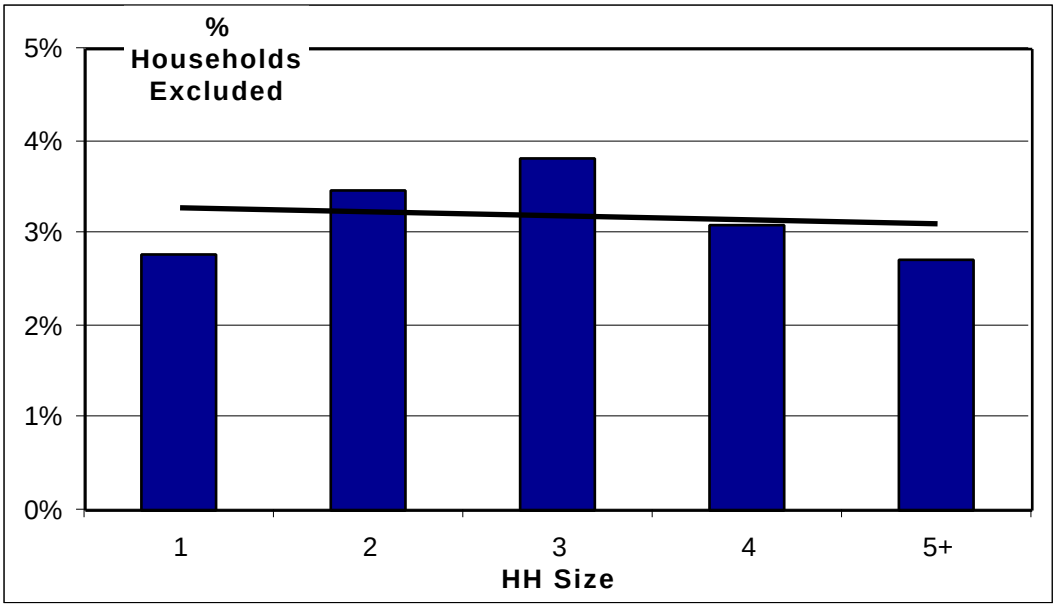

Figure 3: % of Households Excluded as a function of Household Size (20% INR)

Firstly, it can be seen in Figure 3 that in addition to only 3.4% of all households being excluded, this percentage appears not to be a direct function of household size. The percentage excluded varies between 2.5% and 3.5% for small, medium and large households. However, as shown in Figure 4, this percentage does appear to be a function of the total number of stops made by the household. Immobile households are much more likely to be excluded (13% excluded), while the percent excluded then falls with increasing mobility. The reason for this apparently counter-intuitive result is that mobile households have more travel data and the Item Non-Response is lower for travel questions than it is for Person Questions. Thus a household with low mobility will have proportionally more Person data, with higher levels of Item Non-Response, which will lead to an overall higher level of Item Non-Response leading to exclusion as an unacceptable response.

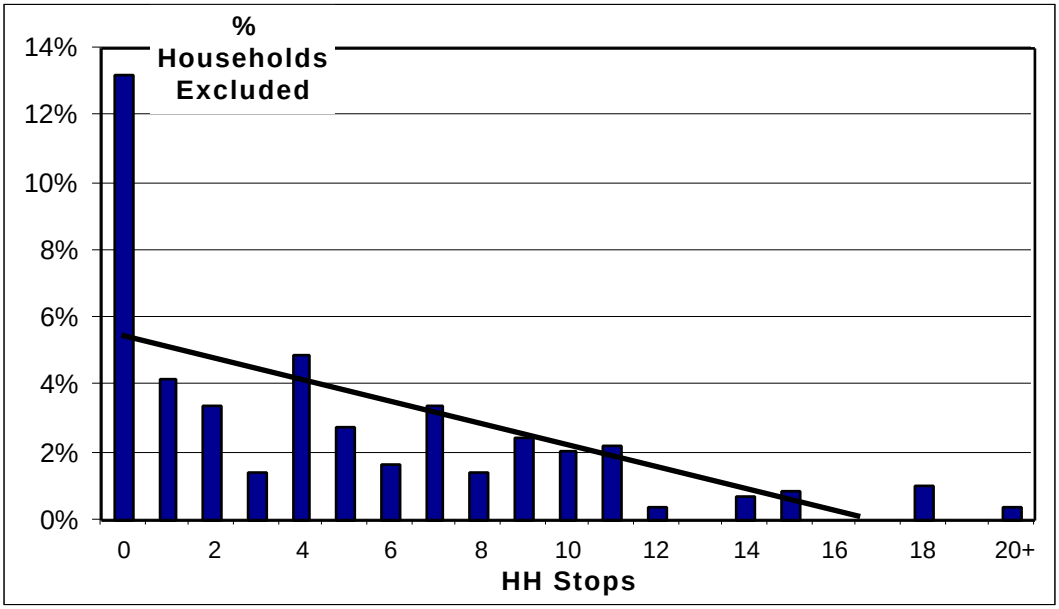

Figure 4: % of Households Excluded as a function of Household Stops (20% INR)

The overall effect of the percent of households excluded and the biases in household size and mobility between retained and excluded households, as expressed in terms on the HH Size Ratio and the HH Stops Ratio, is that with a relatively lenient definition of acceptability, the average household size of retained households is 100% of the average household size of all households (i.e. no bias toward smaller or larger households), while the average number of stops in retained households is 102.3% of the average number of stops in all households (i.e. a bias toward more mobile households).

The Full Spectrum of Acceptability Criteria

Given the above explanations of how the acceptability criterion influences the exclusion of households and the biases in household size and household mobility, the results will now be presented for the full spectrum of acceptability criteria from a very strict criterion for acceptability (no missing data items in a household) up to a very lenient criterion for acceptability (50% missing data items in any household). Figure 5 shows the effect of varying the allowable %INR on the quantity of data retained, while Figure 6 shows the effect of varying the allowable %INR on the quality of data retained.

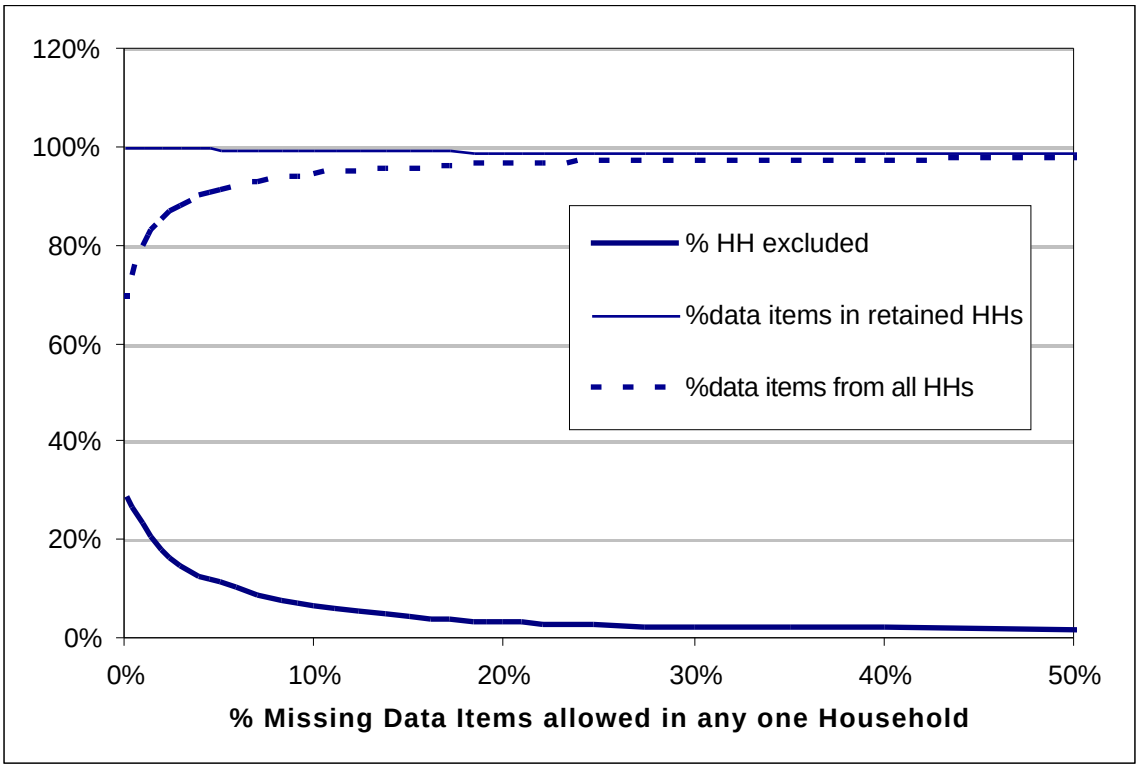

Figure 5: Effect of Allowable %INR on Quantity of Data Retained

Figure 5 shows that the percent of households excluded decreases as the acceptability criterion is relaxed (as might be expected), from 30% with 0% missing data in any household down to 1.5% with 50% missing data in any household. Importantly, however, while the percent of missing data in any one retained household increases from 0% to 50%, the percent of missing data across all retained household only decreases from 100% to 98.2%. As a result, the total proportion of data retained increases from 67.2% to 97.8% as the percent of missing data in any one retained household increases from 0% to 50%. That is, the more relaxed acceptability criterion at the level of the individual household does not significantly increase the proportion of missing data across all retained households, and, in fact, a more relaxed acceptability criterion at the level of the individual

household significantly increases the proportion of data obtained from all retained and discarded households.

The effect of the acceptability criterion on the quality of data retained, as measured by the bias in household size and household mobility, is shown in Figure 6. It can be seen that a very strict criterion (0% data missing in acceptable households) produces a bias in the retained households towards smaller households and less mobile households. As the criterion is relaxed (up to about 5% data missing in acceptable households), the direction of the bias changes, and there is now a bias in the retained households towards larger households and more mobile households. As the criterion is further relaxed (up to 50% data missing in acceptable households), the bias in household size is effectively eliminated while the bias in household mobility is continually reduced.

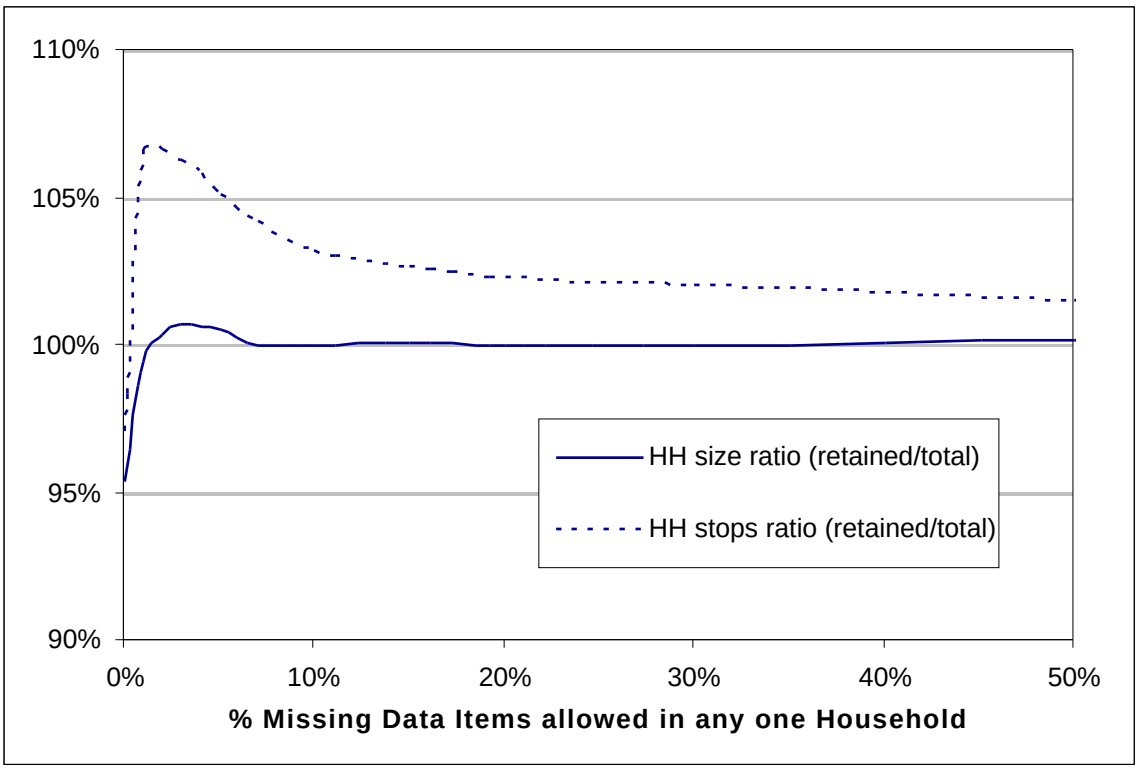

Figure 6: Effect of Allowable %INR on Quality of Data Retained

This section has considered the question of the effect of different criteria for the definition of acceptable responses, based on the maximum percent of missing data allowable in any acceptable household. As a result of an analysis based on VATS 95 data, the following conclusions may be drawn:

- A very strict definition (no missing data allowed in an acceptable household) results in 30% of households, and 32.8% of data items, being discarded;
- A very strict definition (no missing data allowed in an acceptable household) results in a bias in retained households towards smaller households and less mobile households;
- A more lenient definition (20%-50% missing data allowed in an acceptable household) results in less than 3% of households, and 3% of data items, being discarded;
- A more lenient definition (20%-50% missing data allowed in an acceptable household) results in no bias in retained households towards smaller or larger households and a slight bias towards more mobile households (mainly through the exclusion of totally immobile households).

On balance, it would appear that a more lenient definition of an “acceptable response” would be much more cost effective (by not discarding data already collected) and would produce less bias in the characteristics of the retained households.

CONCLUSIONS

This paper has sought to examine the meaning and measurement of Unit Non-Response, and to explore the implications of different definitions of UNR on the quantity and quality of data obtained from a typical mailout/mailback travel survey. The following conclusions have been reached:

- There is no clear dividing line between Unit Non-Response (UNR) and Item Non-Response (INR). High levels of INR eventually convert a household into UNR.

- All stages in a survey which lead to the exclusion of eligible respondents need to be considered in the calculation of a response rate, and a clear distinction needs to be made between completion rates for each stage and the response rate for the overall survey.

- Response rate is usefully defined as the number of acceptable responses divided by the maximum number of possible eligible responses.

- An “acceptable response” needs to be defined in terms of the objectives of the survey.

- Changing the emphasis from “acceptable household response” to “acceptable person response” can result in an increase in effective response rate, with consequent improvements in cost effectiveness and a reduction in sampling bias.

- The measurement of “sample loss” needs to account for observed and unobserved sample loss.

- A very strict definition of “acceptable response” (no missing data allowed in an acceptable household) results in a significant loss of data and a bias in retained households towards smaller households and less mobile households.

- A more lenient definition of “acceptable response” (20%-50% missing data allowed in an acceptable household) results in a very small loss of data, no bias in retained households towards smaller or larger households and a slight bias towards more mobile households (mainly through the exclusion of totally immobile households).

- On balance, it would appear that a more lenient definition of an “acceptable response” would be much more cost effective (by not discarding data already collected) and would produce less bias in the characteristics of the retained households.

While there is an understandable focus on trying to improve response rates, in response to an overall decline in response rates, this paper has suggested that such efforts may be counter-productive. Strict definitions of “acceptable response” may result in significant data loss and the introduction of biases. In addition, the specification of required response rates may lead to practices which achieve the desired response rate, but at the expense of data quality. A cautionary note on this topic is sounded by Krosnick (1999) when he states that “it is not necessarily true that representativeness increases monotonically with increasing response rate” and that “recent research

has shown that surveys with very low response rates can be more accurate than surveys with much higher response rates". Improved response rates lessen the scope for non-response bias, but they do not eliminate it. It all depends on what has been done to improve the response rate. Changes in definitions, the use of non-neutral incentives and over-emphasis on the conversion of non-respondents can all worsen the quality of the data while increasing the quantity of the data.

ACKNOWLEDGMENTS

The authors acknowledge the contributions of Abraham, Barbara, Carsten, Chester, Henk, Jelle, Joy, Nancy, Phillip, Rajat and Susan to the production of this paper.

REFERENCES

Battellino, H. and Peachman, J. (2001). "The Joys and Tribulations of a Continuous Survey". Presented at the International Conference on Transport Survey Quality and Innovation, Kruger National Park, South Africa.

Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowitz, N.A. and Sudman, S. (1991*). Measurement Errors in Surveys.* John Wiley & Sons: New York.

Bonsall, P.W. and McKimm, J. (1993). "Non-Response Bias in Roadside Mailback Surveys". *Traffic Engineering and Control*, December , 582-591.

Brög, W. and Meyburg, A.H. (1980). "Nonresponse Problem in Travel Surveys: an Empirical Investigation". *Transportation Research Record*, 775, 34-38.

Brög, W. and Meyburg, A.H. (1981). "Consideration of Non-Response Effects in Large-Scale Mobility Surveys". *Transportation Research Record*, 807, 39-46.

Brownlee, K.A. (1957). "A Note of the Effects of Nonresponse on Surveys". *American Statistical Association Journal*, 29-32.

CASRO (1982). *On the Definition of Response Rates*. Special report of the CASRO Task Force on Completion Rates, Council of American Survey Research Organizations, New York.

Deming, W.E. (1953). "On the Probability Mechanism to Attain an Economic Balance Between the Resultant Error of Non-Response and the Bias of Non-Response". *Journal of the American Statistical Association*, 48, 743-772.

Donald, M.N. (1960). "Implications of Nonresponse for the Interpretation of Mail Questionnaire Data". *Public Opinion Quarterly*, 24, 99-114.

Groves, R.M. (1989). *Survey Errors and Survey Costs*. John Wiley & Sons: New York.

Groves, R.M. and Couper, M.P. (1998). *Nonresponse in Household Interview Surveys*. John Wiley & Sons: New York.

Kanuk, L. and Berenson, C. (1975). "Mail Surveys and Response Rates: A Literature Review". *Journal of Marketing Research*, 12, 454-473.

Krosnick, J.A. (1999). "Survey Methodology". *Annual Review of Psychology*, 50, pp 537-567

Lessler, J.T. and Kalsbeek, W.D. (1992). *Nonsampling Error in Surveys*. John Wiley & Sons: New York.

Polak, J.W., Ampt, E.S. and Richardson, A.J. (1995). "An Analysis of Non-Response in Travel Diary Surveys". Presented at *the 23rd PTRC European Transport Forum*, University of Warwick, UK.

Richardson, A.J. (2000). “Behavioural Mechanisms of Non-Response in Mailback Travel Surveys”. Presented at the *79th Annual Meeting of the Transportation Research Board*, Washington, DC.

Richardson, A.J. and Ampt, E.S. (1994). “Non-Response Effects in Mail-Back Travel Surveys”. Presented at the *7th International Conference on Travel Behaviour Research*, Santiago, Chile.

Richardson, A.J., Ampt, E.S. and Meyburg, A.H. (1996). “Nonresponse Issues in Household Travel Surveys”. *TRB Conference on Household Travel Surveys: New Concepts and Research Needs*, 79-114.

Richardson, A.J. and Ampt, E.S. (1995). "The Application of Total Design Principles in Mail-Back Travel Surveys", *7th World Conference of Transport Research*, Sydney.

Richardson, A.J., Ampt, E.S. and Meyburg, A.H. (1995). *Survey Methods for Transport Planning*. Eucalyptus Press: Melbourne.

Sen, A., Sööt, S., Yang, L. and Christopher, E. (1995). “Household Travel Survey Nonresponse Estimates: the Chicago Experience”. *Transportation Research Record*, 1493, 170-177.

Statistics Canada (1998). *Statistics Canada Quality Guidelines*. Statistics Canada, Catalogue 12-539-XIE.

Steeh, C. (1981). “Trends in Response Rates”. *Public Opinion Quarterly*, 45, 40-57.

Stopher, P.R. and Sheskin, I.M. (1982). “A Method for Determining and Reducing Non-Response Bias”. *Transportation Research Record*, 886, 35-41.

Tourangeau, R., Rips, L.J. and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge University Press: Cambridge, UK.

Wermuth, M.J. (1985). “Non-sampling Errors due to Non-response in Written Household Travel Surveys”. In ES Ampt, AJ Richardson & W Brög (Eds), *New Survey Methods in Transport*. VNU Science Press: Utrecht, The Netherlands.