
DEVELOPING STANDARDS OF TRANSPORT SURVEY QUALITY

By

**Peter R. Stopher, Professor of Transport Planning
Institute of Transport Studies, The University of Sydney**

and

**Peter M. Jones, Professor of Transport Policy and Behavioural Analysis
Director, Transport Studies Group, University of Westminster**

**Keynote Paper prepared for the International Conference on Transport
Survey Quality and Innovation: How to Recognize It and How to Achieve It
Kruger Park, South Africa, August 2001**

July 2001

1. INTRODUCTION

Over the past 40 years, many millions of dollars have been spent on collecting household or person based data for transport planning. For most metropolitan areas around the world, the largest routine expenditure made from planning budgets is for the conduct of household or person travel surveys. In 1996, it was reported (Stopher and Metcalf, 1996) that the average survey cost in the U.S.A. was \$400,000 for consultant services for the conduct of household travel surveys. Even though the cost of conducting surveys may be significantly lower in many other countries, a conservative estimate might be that somewhere in excess of \$50 million is spent on household travel surveys each year around the world. To this must be added the costs of other surveys that would typically be undertaken as part of metropolitan planning. These may include freight and commercial vehicle surveys, on-board public transport surveys, cordon and screen line surveys, workplace surveys, special generator surveys, etc.

In spite of this huge level of activity and expenditure, there is no agreement within the transport profession as to what constitutes a good survey, nor how one would recognize whether a survey is good or not. Nevertheless, even larger sums of money are subsequently spent on developing and using travel-demand models that are based on these data, and in investments into major capital projects, implementation of far-reaching policies, and other related decisions.

In many cases, the agencies that commission the data collection do not have staff with in-depth knowledge and experience in surveys. As a result, many of these agencies are unable to make informed selections of consultants to perform surveys, and are also unable to assess whether or not a useful product was obtained. Subsequent work in using the data for descriptions and modelling often reveals serious flaws in the data that could have been avoided if there were either a sufficient availability of expertise at the agencies, or a clear definition of what constitutes a good quality survey that could be followed by an agency in guiding the process, selecting consultants and assessing the work that was done.

Some consultants that undertake such work are unaware of the difficulties involved in data collection, and lack knowledge and expertise in various aspects of collection and assessment of the data, which are neither apparent to them, nor to the agencies that may select them. They, too, could benefit from a definition of survey quality and an indication of the procedures and measures that are required to produce a quality survey. This would assist them in determining the type of survey to undertake, the methods to be implemented, and the means to assess whether or not the survey was being executed satisfactorily.

Because each transport survey that is undertaken is usually different from any other such survey, it is also frequently difficult, if not impossible, to compare among surveys to ascertain whether an appropriate level of quality has been attained. Differences in methodology, design, and protocols often obscure basic differences in quality among different surveys.

The primary goals of this conference are to move the practice of transport surveys forward, by determining how to measure and assess transport survey quality, to identify emerging issues that may affect the relevance of quality and performance measures in the next decade, and to consider the suitability of performance measures largely developed in a USA/European context for application in other cultural contexts. In part, a move towards higher quality requires greater comparability among surveys, and the adoption of a minimum level of what may be termed “good practice” in the conduct of transport surveys. What constitutes good practice in transport surveys has not been defined previously and is a key issue to be addressed at this conference.

The purpose of this keynote paper is to provide a general framework within which the assessment of survey quality can be progressed, and to which the fourteen workshops can contribute in a number of specific areas. This paper also seeks to set out some of the steps necessary to establish a definition of what constitutes good practice in transport surveys at the beginning of the 21st century. In short, the paper seeks to illuminate how to recognise transport survey quality, and to begin the process of defining what is needed in order to be able to achieve quality. At the same time, we recognise that it is desirable to allow innovation to take place in the field of transport surveys, so that the definition of what constitutes good quality will advance and change as new techniques and technologies are applied.

Following this introduction, the paper discusses some of the issues surrounding the definition of standards and the measurement of quality (section 2). Section 3 then identifies nine key stages in the design and execution of transport surveys, which make up a framework for the assessment of survey quality. Each of these nine stages is outlined in greater detail in section 4. Finally, section 5 sets out some of the key links between the fourteen conference workshops and the 66 elements of the framework.

2. DEFINING STANDARDS AND RECOGNISING QUALITY

If it is desired to establish some sort of benchmark of good practice, and to provide a means to assess quality, then some type of standard is implicitly required. Standards provide the means against which to assess a specific instance of the practice of a transport survey, thereby allowing one to determine if the survey has used current elements of good practice.

It is important to recognise the distinction between “standards” and “standardisation.” **Standards** are considered minimum thresholds of the properties of a product that must be attained in order for the product to be acceptable. In the context of travel surveys, the properties considered would typically be the quality of the data, the ethics employed in collecting the data, and the procedures used to evaluate, document, archive, and disseminate the information collected. **Standardised procedures**, on the other hand, are stipulated methods of conducting an activity (e.g., using agreed terminology and classifications). By fixing a process, ambiguity is reduced, standards are indirectly achieved, and assessment is promoted by clarity of concept and the opportunity to compare values from different sources. Thus, standardised procedures are an indirect application of standards, but they also enhance communication and understanding, promote efficiency, and facilitate assessment of the product.

According to the dictionary definition, the word “quality” has many meanings. The one with most applicability to this context is “degree of excellence,” “grade,” or “superiority in kind.” Implicit in these meanings of the word “quality” is the adjective “good,” or “high,” i.e., what is being sought here is a transport survey that is of good or high quality. From this, one could suggest that quality is a measure of suitability to purpose, or acceptability.

One of the main functions of standards is the assurance that a product a user or client plans to purchase carries the approval or certification of a reputable standards organization, and is of reliable quality. In the manufacturing and service sectors, international standards agencies such as the International Standards Organization (ISO) are increasingly being used to accredit organisations. The ISO requires that suppliers structure and operate their company according to eight quality management principles. The ISO defines a quality management principle as (ISO, 1997):

“... a comprehensive and fundamental rule or belief, for leading and operating an organization, aimed at continually improving performance over the long term by focusing on customers while addressing the needs of all other stakeholders.”

Richardson and Pisarski (1997) have translated the ISO guidelines into requirements for a travel survey company. They maintain that while it requires considerable commitment from the company to implement and maintain, its benefits in being able to deliver a quality product in a consistent manner are substantial.

While there are no generally accepted standards for transport surveys, several organisations have produced useful guidelines. For example, Statistics Canada has produced a comprehensive set of “good practices” in travel surveys in their document *Quality Guidelines* (Statistics Canada, 1998). Guidance is provided on how to conduct each step in a survey and how to structure and operate a survey company so as to collect quality data. The Council of American Survey Research Organizations (CASRO) has produced similar guidelines on “good practice” in its *Survey Research Quality Guidelines* document (CASRO, 1998).

We suggest that there might be up to eight ways in which standards and standardisation might contribute to the improvement of surveys:

- Improvement in survey *quality*
- Improvement in survey data *reliability*
- Improvement in survey data *usefulness*
- Improvement in *cost effectiveness* or value
- Improvement in *comparability* among surveys
- Improvement in ability to *measure survey quality*
- Improvement in *clarity*
- Improvement in *completeness*

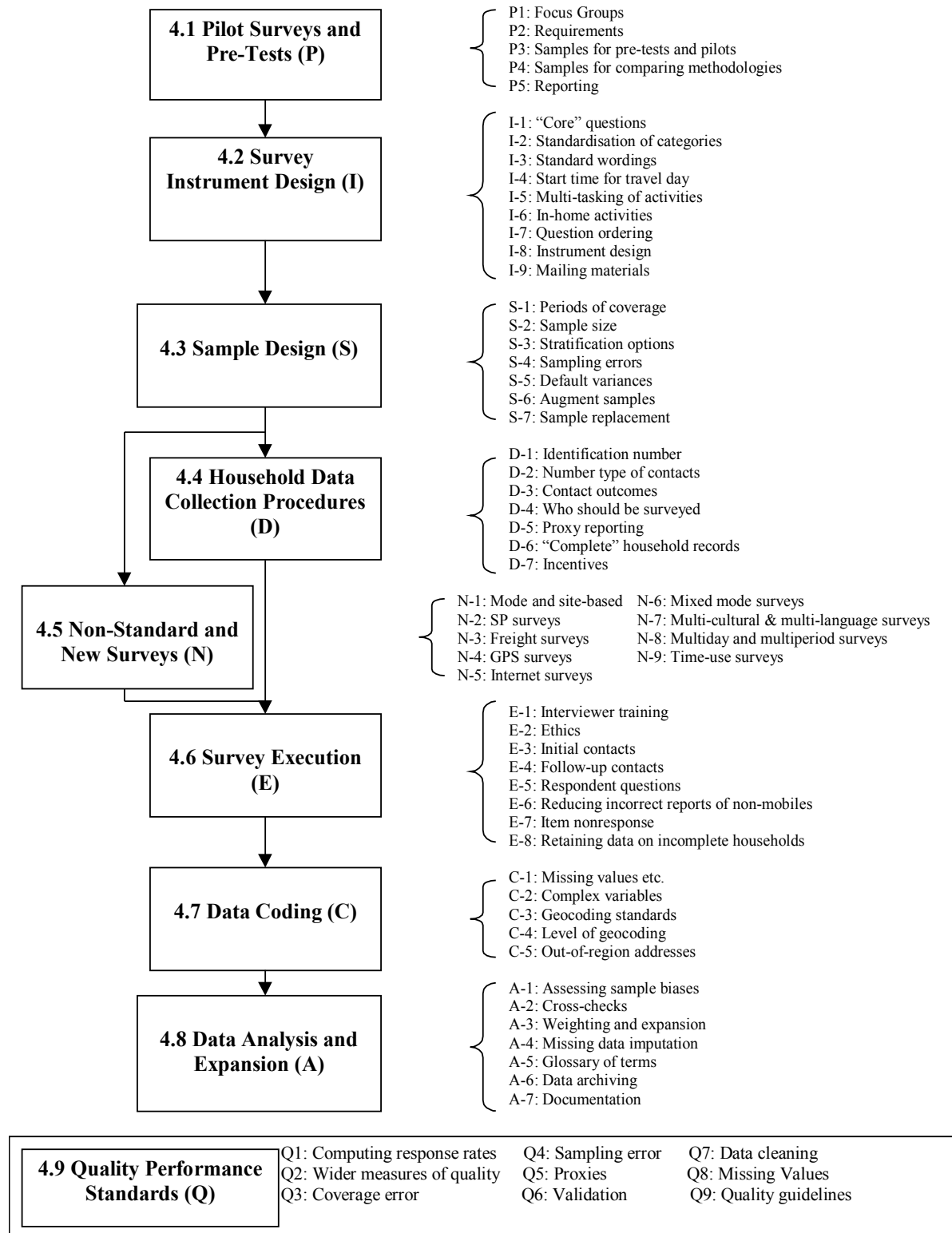
The relevance of each type of benefit will vary according to the aspect of the survey process that is being considered.

3. A FRAMEWORK FOR THE ASSESSMENT OF SURVEY QUALITY

Figure 1 sets out our proposed framework for the assessment of survey quality. This is based on eight broad, largely sequential, stages covering the various processes involved in the design and execution of a transport survey, starting with “pilot surveys and pre-tests” and ending with “data analysis and expansion.” There is also a ninth category concerned with overall measures of survey quality performance standards.

Each of these nine stages is sub-divided into between 5 and 10 specific tasks, making a total of 66 sub-headings, dealing with specific elements of the framework. Each of these elements is briefly discussed in section 4. Some are already well documented, or are the subject of other current initiatives to define survey standards (e.g., a major NCHRP-funded initiative in the USA focussing particularly on household telephone and mailback surveys). The workshops in this conference focus on recent developments that are less well covered by other initiatives, as set out in section 5.

Figure 1: Proposed Framework for the Assessment of Survey Quality



4. Elements of the Framework

In this section we briefly describe each of the elements of the framework presented in Figure 1, in each case defining the key terms and suggesting features that might benefit from the setting of standards or the preparation of guidelines.

4.1 P: Pilot Surveys and Pre-Tests

It is generally essential to carry out preliminary work before launching a major survey, in order to test and refine the sampling methodology, the survey instrument and the procedures for data checking and analysis.

4.1.1 P-1: FOCUS GROUPS

Some preliminary qualitative research using focus groups can be very valuable both in helping to define the issues of concern to respondents and to determine the kinds of terminology to be used in the questionnaire that will be simplest for respondents to understand. They are used to assist in determining key attributes to be included in stated preference surveys. Focus groups can also be used at a later stage, to help interpret the findings of quantitative surveys.

Focus groups usually consist of a small number of people (typically 6-12), who are chosen to represent a specific subgroup of the population. They meet together and the group facilitator introduces the topics for discussion, e.g., starting with the general subject of the survey and then discussing how specific information might be collected. The discussion is usually recorded, and the facilitator plays the role primarily of keeping the group focused on the issues of concern, and making sure that all views are freely expressed. A group may often last for two or three hours. A pre-meeting activity of completing draft survey instruments may also take place, so that the group can discuss how each person found the task of completing the survey and whether there was difficulty in understanding any parts of the task.

Among the issues that need to be addressed are the following:

- How can focus groups best contribute to survey definition and design?
- How many focus groups are needed?
- How should focus groups be used to test a travel survey?
- How should focus groups be recruited?
- How much is it necessary or desirable to pay focus group members to participate?
- Should focus group members receive survey materials prior to meeting?
- What qualifications are needed to facilitate a focus group?
- Are there “do’s” and “don’ts” in conducting focus groups?
- What benefits arise from using focus groups?

Both standards and guidelines may be appropriate for Focus Groups.

4.1.2 P-2: REQUIREMENTS FOR PRETESTS OR PILOT SURVEYS

It would be useful to distinguish between a pilot survey and a pretest; currently the two terms are used interchangeably in the transport profession. However, the survey literature distinguishes between them as follows:

- A **pretest** is a test of one or a few elements of the survey, usually without testing other elements
- A **pilot survey** is a complete run through of the survey (i.e., a dress rehearsal), including analysis of the results.

In general, a pretest is necessary either when one element of a survey has been changed from an earlier version that has been applied to essentially the same population, or when a survey developed on one population is to be used on a different population. A pilot survey should usually be done each time a new survey is designed, or a survey performed on one population is to be performed on another population. When the pilot survey reveals the need for significant changes to one or more elements of the survey, the changed elements should be pretested again before full fielding of the survey.

The first standard is to correct the mislabelling of these test surveys in the transport profession. Because pilot surveys and pretests add to the time requirements of a survey, as well as to the cost, it will also be necessary to provide some way to demonstrate the cost-effectiveness of conducting these preliminary tests of the survey.

The second standard is to specify the circumstances under which a pilot survey or pretest could be considered unnecessary. We prefer to specify the conditions when they are not necessary, because these are far more limited than the conditions under which one or other is necessary.

4.1.3 P-3: SAMPLE SIZE REQUIREMENTS FOR PRETESTS AND PILOT SURVEYS

It seems likely that true pretests – that is, the testing of a single element of a survey, such as a redesign of certain questions – would be able to be done effectively with a very small sample of households, such as 25 to 50 households. However, a rationale needs to be found for specifying the sample size. For pilot surveys, it would seem likely that a larger sample should normally be necessary. Again, however, a rationale for this is needed.

There is a need to consider not only the size of the samples required for pretests and pilot surveys, but also the composition of the sample and how the sample should be drawn. There are no clear statistical procedures for determining the sizes of samples for pretests and pilot surveys. Clearly, the first issue must be one of what it is desired to learn from conducting the pretest or pilot survey. Since this will vary from survey to survey, it is possible that no standard can be set, but only guidance offered.

4.1.4 P-4: SAMPLE SIZES FOR COMPARING METHODOLOGIES

There are particular requirements that are placed on a pilot survey that is comparing two or more methodologies. In such a case, there is a need for a sample that will be large enough to draw a statistically valid conclusion that the results of one version of the survey are significantly different or better than the other, or to conclude that there is no statistically significant difference between them.

It is thus necessary that minimum sample sizes for testing alternative versions of a survey should be specified. This will need to be done by developing some general rules about the appropriate measures of what is different between two designs. For example, is response rate the principal statistic of concern, or is a statistic such as item non-response, or trip rate, or some other measure the appropriate one to be used to compare the effectiveness of the two survey designs?

4.1.5 P-5: REPORTING OF PRETESTS AND PILOT SURVEYS

There should be a minimum standard that the final report of a survey should document whether or not a pilot survey or any pretests were conducted. Second is the issue of what should be reported from a pretest or pilot survey. The latter should include:

- Sample sizes and methods of drawing the samples for any pretests and pilot surveys
- Nature of the design that was tested
- Results of the tests, including response rate(s) and other measures of quality
- Conclusions drawn from any pretests and pilot surveys and changes implemented as a result of the pretests or pilot surveys.

The documentation should include any statistical test performed to establish whether or not to make changes to the final survey, and anecdotal information should also be included that may have led to changes in the design of the survey and its protocols.

4.2 I: Survey Instrument Design

This section deals with the standardisation of key questions in transport surveys and the broader design of the survey instrument.

4.2.1 I-1: CORE QUESTION SPECIFICATIONS

Setting minimum specifications for 'core' questions would enable survey results to be compared between areas and over time. Table 1 provides examples of candidate questions that could be specified as a minimum for a personal travel survey. There is considerable debate on whether race and ethnicity should be asked, whether industry and occupation should be included, and whether personal income should be asked either in place of or in addition to household income.

Another issue that should be addressed in this area is that of whether or not personal travel surveys need to collect data on the alternative transport modes, destinations, or other elements that were in the decision set of the respondent, but not chosen.

A related issue is whether or not travel times, distances, and costs should be asked for in the minimum set of questions. This is related to a larger issue, which is the use of network-derived times, costs, and distances in travel-demand modelling.

One additional question that should almost certainly be included is to ascertain if the person would be willing to be contacted again for other transport study projects. This question would allow qualification of households that could be used for validation activities, as well as for a variety of other purposes. It would also provide a potential pool of respondents for use in focus groups of various types (see also item P-1).

4.2.2 I-2: STANDARDISATION OF CATEGORIES

There appears to be considerable merit in going further than defining the core questions in I-1 by setting standards for these core questions, as well possibly as also considering standard categories for some of the questions that are not specified within the minimum, but which may be included in many surveys.

Probably, the most important of these are: income, race and ethnicity, job classification, type of dwelling/building, relationships among household members, modes of travel, mobility handicaps, education levels, and activities both within the home and outside. The grouping of

activities into common trip purpose-related categories would also be a useful element of this aspect of standardisation.

Table 1: Candidate Subjects for Inclusion as ‘Core’ Questions in a Survey

Category	Ref	Item	Description
Household	H1	Location	Home address or home position in geographic terms
	H2	Type of Building	Detached, semi-detached, terraced, flat, etc.
	H3	Number of Members	
	H4	Relationships	Matrix of relationships between all members of the household
	H5	Income	Indication of total household income (gross, annual) from all sources
	H6	Number of Vehicles	Summary of number of vehicles from vehicle data
Personal	P1	Gender	
	P2	Year of Birth	(Preferable to requesting age)
	P3	Commitments	Work and/or student status for each person
	P4	Paid Jobs	Number of paid positions and hours worked at each in the past week
	P5	Occupation	Type of work done
	P6	Job Classification	Employee, self-employed, etc.
	P7	Driving License	Whether or not a current drivers license is held
	P8	Nonmobility	Indication of why no out-of-home activity was performed on a survey day including work-at-home days
	P9	Start Location	Location at the beginning of the first survey day
	P10	Education Level	Highest level of education achieved
	P11	Handicap	Types of mobility handicap, both temporary and permanent
Vehicle	V1	Make	
	V2	Model	
	V3	Body Type	E.g., car, van, RV, SUV, etc.
	V4	Year of Production	
	V5	Ownership of Vehicle	Household/person, lease, institution
	V6	Use of Vehicle	Main user of vehicle and possible list of other users
Activity	A1	Start Time ¹	
	A2	Activity or Purpose	
	A3	Location	Where the activity was performed, unless travelling
	A4	Means of Travel	If activity is travel, what mode(s) was used
	A5	Mode Sequence	Unless collected as fully segmented data
	A6	Party Size	Number of persons travelling with respondent as a group
	A7	Household Membership of Party Members	Number of persons in the group who live in respondent’s household
	A8	Costs	Total amount spent on tolls, fares and respondent’s share
	A9	Parking	Amount spent to park
	A10	Public Transit Route	Name or number of bus route, train, or other transit used on each segment

There is a need to determine which are the candidate questions for standisation of categories, and what these might be.

4.2.3 I-3: STANDARD QUESTION WORDINGS

¹ Only start time needs to be ascertained in a time-use or activity survey, because, by definition, the start time of an activity is the end time of the previous activity. Only the last activity should need an end time.

There are several questions in most household travel surveys that could be assisted by standardising the wording. In particular, questions of income need standardising so that all surveys are measuring the same definition of income. If questions such as occupation and industry are asked, these also require some standardisation to ensure that all surveys are asking for the same information. Other issues pertain to whether vehicles should be those owned or available for use by the household, how activity information is requested, how education status is asked, work at home, and other similar variables.

Most questions for which standard categories are defined will need to have standard wordings defined, so that the wording relates correctly to the answer categories.

4.2.4 I-4: STANDARDISING AND CODING OF TIME OF DAY TO BEGIN AND END TRAVEL REPORTING PERIOD

A standard for the time of day to begin and end reporting would make surveys clearly compatible. One possibility would be to start the standard diary period at 3 a.m. and end at 2:59 a.m. For a 24-hour diary, this will mean that it would start at 3 a.m. on the diary day and end at 3 a.m. on the following day. Longer period diaries would start at 3 a.m. on the first day and end at 3 a.m. on the day following the n th day of the n -day diary period.

It would be useful to set a consistent standard for coding times of day. Because the interest is usually in both determining at what time of day a trip occurred and also its duration (which involves subtracting the starting time from the ending time), it is easiest for subsequent analysis if all clock times are recorded initially on a 24-hour clock or its multiple, and subsequently are converted to minutes from the start of the survey (e.g., at 3 a.m. on the first day).

4.2.5 I-5: MULTI-TASKING OF ACTIVITIES

People perform various multi-tasked activities throughout the day. These include such activities as driving and talking on a cellular phone, eating and watching TV, travelling on public transport and performing work activities such as reading, reviewing, or using a laptop, etc. Almost all survey instruments in transport continue to ask questions as though people only undertake a single activity at a time. As a result, much information is missing from the typical survey, and purposes are probably misstated by this simplification.

From the viewpoint of the blurring of work and other activities, the increasing ability of people to multi-task as a result of technological advances, and the potential impacts of these on daily travel and activity patterns, this would appear to be an important area for the development of standards.

4.2.6 I-6: COLLECTION OF IN-HOME ACTIVITIES

This is an area in which there are probably the greatest inconsistencies in current practice, and where there is a significant potential for standardisation that would set minimum guidance on what should be collected.

There appears to be general agreement among most travel-demand modellers that more detail needs to be collected about in-home activities. Furthermore, there is a need to standardise on how to collect this information so as to avoid prying into personal and private things that go on in the home, while also ensuring that the detail needed for transport planning purposes is provided. The way in which the collection of in-home data is standardised will necessarily depend on the instrument used in the survey.

4.2.7 I-7: ORDERING OF QUESTIONS

Standards or guidance here might cover a range of issues, from the placing of sensitive questions, such as income, towards the end of the survey, to providing guidance on the order in which travel and personal questions may be asked. Question ordering needs to appear logical to respondents, making it as easy as possible for them to work their way through. A well structured questionnaire makes it easy to provide ‘skips’ where there are sets of questions that are not applicable to certain types of respondent.

Another issue here relates to certain sequences of questions, such as occupation and working at home. There are many occupations, such as retail clerk, air-traffic controller, sanitation worker, etc. that do not permit working at home. Therefore, care needs to be taken not to ask a question about working at home following a question on occupation.

4.2.8 I-8: QUESTIONNAIRE INSTRUMENT DESIGN

Instrument design includes the overall format of the instrument (booklet, leaflet, two-sided card, etc.); uses of varying typefaces, bold, underline, and italics; use of colour, arrows, boxes, and other devices to direct the respondent; placement of instructions; survey instrument length; etc. In this area, it will be particularly important to specify minimum standards, but not to be prescriptive about the way in which instruments should be designed, because instrument design is an area that should see much innovation in the future. Included in the standards that might be considered here are such items as not using too many fonts and font enhancements; using consistent fonts and enhancements to distinguish instructions, questions, and responses; avoidance of clutter in the instrument design; avoidance of coding or other “office use” elements in the design; etc.

Another aspect in this area is that of the design of instructions for respondents. Many past transport surveys have included extensive written instructions, which a review of the survey results show were either not read, or at least not understood and applied by respondents. There is a need to move to more graphic instructions, simplification of instruments so that fewer instructions are needed, and other devices that avoid the need (and expectation) for respondents to read lengthy explanations and instructions.

4.2.9 I-9: MAILING MATERIALS

There is some evidence that the materials used to mail questionnaires to households, as well as materials for households to mail back their responses, have an effect on response rates. Some survey practitioners maintain that the appearance of the mailing envelope is of considerable importance for households to take a survey seriously. This is particularly considered to be an issue in North America, where the amounts of “junk mail” received by most households has become excessive, and anything that appears to be another item of such junk mail is likely to be discarded without even opening it.

Among the suggestions for mailing materials that may have an impact on response rates are the following:

- Use of a white rather than a brown envelope
- Use of postage stamps, especially commemorative stamps, rather than a franking machine or preprinted bulk mail
- Printing of addresses directly on the envelope, rather than use of address labels
- Provision of a recognizable return address on the envelope
- A clear indication on the outside of the envelope of the importance of the contents, or identification of the survey name
- Provision of a postage-paid return envelope when survey materials are to be mailed back

Because some of these involve added expense, either in time, or materials, it would be helpful to explore the extent to which any or all of these items can affect the response rate to a survey.

4.3 S: Sample Design

There are a number of issues relating to the definition and selection of the survey sample, some of which are primarily statistical in nature, and others that relate more broadly to the purpose of the survey.

4.3.1 S-1: PERIODS OF COVERAGE FOR DATA COLLECTION

Most household travel surveys are conducted in the spring and fall/autumn seasons, although there are a few that also include periods of the summer and winter, particularly where there may be air quality issues of severity in those seasons. A smaller number are all-year (e.g., the KONTIV survey in Germany, the Sydney Household Travel Survey, etc.), enabling analysis of annual energy consumption, links with accident analysis, etc.

Guidance needs to be provided on the suitability of conducting year-long or seasonally focused surveys, taking into account factors such as the use to which the data will be put, total sample sizes, etc. A second aspect relates to the issue of keeping a log of unusual events, weather, and other occurrences that may affect the travel of respondents during the survey period.

4.3.2 S-2: SAMPLE SIZE

This is probably the single most controversial item in household travel surveys, and one on which there is virtually no agreement, as evidenced by samples ranging from a few hundred to as much as 20,000 households. There is a need to develop minimum standards for sample size, depending on the purpose of the personal travel survey, e.g., model estimation, model updating, regional description, policy testing and formulation. This would be different from previous guidance which either offered formulas for calculating minimum samples, or provided some possible default values to use in sample size calculations.

The determination of an appropriate sample size will depend on stratification options (S-3) and acceptable error levels (S-4).

4.3.4 S-3: STRATIFICATION OPTIONS FOR SAMPLES

There are two reasons to stratify a sample. The first is to ensure inclusion in the sample of units that might otherwise be omitted, or that might occur in sufficiently small numbers as to be not useful (see S-6). The second is to increase the efficiency of the sample.

The sampling error is a function of the standard deviation of a characteristic and the sample size. Thus, in simple random sampling - since one cannot control or affect the size of the standard deviation of a characteristic in the population - the only way to reduce the sampling error is to increase the sample size. Unfortunately, the error decreases only in inverse proportion to the square root of the sample size. Therefore, doubling the sample size only decreases the sampling error by 30%.

In stratified sampling, the sampling error is a function only of the standard deviation within a cell. Therefore, if the population is divided into strata in which the within-cell variance is small in comparison to the between-cell variance, the amount of variance that contributes to sampling error is reduced, and a more accurate sample is obtained without increasing the sample size. Or,

it could be turned around to say that the same sampling error could be achieved with a smaller sample, hence at lower cost.

Sample size (S-2) is affected directly by the design of the sampling methodology, i.e., simple random sample, proportionate sample (stratified sample with uniform sampling fraction), disproportionate sample (stratified sample with variable sampling fraction), cluster sample, or multistage sample.

Two other variants of sampling are also used in transport surveys: choice-based sampling and systematic sampling. Choice-based sampling can be considered as an extreme form of disproportionate sampling, in which those who do not make the choice(s) of concern are sampled at a zero rate. Systematic sampling can be considered to be an approximation to simple random sampling, provided that care is exercised in its execution.

There is little guidance on what is the appropriate sampling methodology in different circumstances, or on the sample sizes to choose for each cell of a stratification matrix. Because the use of stratified sampling generally requires contacting households to determine membership in a stratum and then qualifying or disqualifying the household on the basis of the required sample in a cell, there are cost implications that need to be taken into account.

4.3.4 S-4: SPECIFICATION OF SAMPLING ERROR REQUIREMENTS

Frequently, Requests for Proposals specify that the required sample must provide no more than, say, ± 10 percent error with 95 percent confidence in something such as a trip rate. Generally, this appears to be specified with little understanding of what it means. It would be reasonable to question whether 10 percent error is acceptable, compared to 5 percent error, and whether the significance level should be set to 90 or 95 percent. Also, the error is almost always specified for trip rates, while the data will be used for much more than trip-rate estimation. The implications of a particular error level for trip rates on estimation of such elements as mode choice or network volumes is largely unknown.

Among the issues that should be addressed in developing standards and guidance are:

- For error levels specified on trip rates, what is the appropriate level of error and significance to use, and what are the implications of choosing different error levels or significance levels?
- What are the implications of setting an error level on trip rates for other statistics that may be estimated from the data?
- How could error levels for other statistics, such as mode choice proportions, be specified and used in designing a sample?
- What are the implications on sample size and sampling method of different error specifications?

4.3.5 S-5: DEVELOPMENT OF DEFAULT VARIANCES

Estimation of error requires an estimate of the variance of critical variables. Probably one of the issues that has kept sampling strategies to relatively simplistic ones for household travel surveys is the lack of information on variances for those variables that are normally considered critical in transport planning analysis. Again, this has implications for all of the sampling issues, because the error levels are determined by the variance, hence sample size and stratification procedures are also determined by the variance. In the absence of information about the variance, survey designers either assume constant variances across all strata in a sampling scheme, or make some other working assumption that will allow sample size calculations to be made.

Default variances could be provided to help determine appropriate sample sizes and other issues in the absence of local data on these values. They could also be used subsequently to assist in assessing the quality of any given survey, by comparing the variance measured in a specific survey to the default value for each attribute of interest. Variances either much smaller or much larger might indicate potential problems in the survey.

4.3.6 S-6: SIZES AND COMPOSITION OF AUGMENT SAMPLES

It is usually the case that personal travel surveys require augmentation because of a lack of rare behaviours in the collected data, and the problems of sampling to include them. Rare behaviours in the U.S.A. include transit riding and bicycling. In many countries, carpooling may be a rare behaviour. Augmentation samples can be drawn in different ways, and several recent surveys have tried different methods to achieve the required enrichment of the data.

In Portland, OR, sampling was done more heavily in the vicinity of light rail lines and in areas designated as “bicycle friendly.” In Dallas-Fort Worth, some households were recruited through bus-stop intercepts, where a name, address, and telephone number were collected.

Two questions arise:

- When is an augment sample necessary?
- What size should the augment sample be?

4.3.7 S-7: SAMPLE REPLACEMENT

Procedures for sample replacement are critical in preserving the integrity of the initial sample – but should only be invoked once all reasonable efforts have been made to contact the original household (D-2) and obtain a complete response (D-6).

Most surveys are based on anticipated response rates and set up samples that provide sufficient oversampling to handle expected nonresponse. If the sample is not a simple random sample, but is a stratified or other more complex sample, the oversampling must account for varying nonresponse in different strata. Some surveys provide top-up samples “on the fly”, with new households being added to the sample as needed, whenever nonresponse drives the total sample below what is desired. On the face of it, this method has the probability of producing a very distorted sample, particularly when attempts to gain cooperation of respondents are not pursued aggressively.

A reasonable standard on the make up sample would be to provide guidance on expected rates of nonresponse by potential stratification (see item S-3) and to require that sufficient sample be drawn at the outset of sampling to allow for a worst-case scenario on nonresponse. Acceptance of this will require empirical research to demonstrate the effects of setting lower standards.

3.4 D: Design of Household Data-Collection Procedures

There are a number of issues that arise in relation to the procedures for data collection, for accepting proxy responses, and for defining a ‘complete’ response, that have major implications for survey quality.

4.4.1 D-1: CREATION OF UNIQUE IDENTIFICATION NUMBERS

In some CATI and CAPI software, a new ID number is assigned each time that a call attempt or reminder is made. Two aspects require standardisation here.

First, once a sampling unit is established, it would be assigned an unique identification number which would be retained for the duration of the survey.

The second part of this standard would be to propose how the ID number should be formulated. For example, some CATI programs automatically assign a six or seven digit ID number, in which the first two digits indicate the interviewer number and the remaining digits are assigned sequentially, as interviews are completed in the CATI system. This is probably not a useful procedure except where the same interviewer makes all contacts with a specific household or person.

The standard would ensure that the unique ID number contained useful information about the household, such as the pre-defined day of week for travel information, or the strata to which the household belonged.

4.4.2 D-2: NUMBER AND TYPE OF CONTACTS

The number and type of contacts can impact on response rates very significantly and improving on this is often overlooked as a cost-effective mechanism to obtain return of materials. As a rough rule of thumb, if the initial contact generates a response rate of R , then the first reminder will add $0.5R$, the second reminder will add $0.25R$, the third reminder $0.125R$, and so forth. Thus, three reminders can almost double the initial response rate.

Guidance is needed on the minimum number of contacts and the nature of the contacts that should be made in the survey. This will include any pre-recruitment contact, such as a letter or other prior contact, the recruitment contact itself, any contact prior to the diary day, such as an eve-of-diary-day contact, and the number and type of contacts to be made following the diary day, particularly for mail-back surveys.

Different guidance will be required according to the type of survey protocol, including: mail-out/mail-back surveys, telephone recruitment with mail-out/mail-back surveys, telephone recruitment with mail-out and CATI or CAPI retrieval, and face-to-face interviews. Reminders may involve a telephone call or a mail-out of a postcard, or re-mailing of the entire survey package. There has been some limited work done in tracking the effects of different reminders (Richardson, et al., 1995, Richardson, 2000), however, it is not extensive, and further corroborating research is required.

The standard should be specified by setting a minimum number of attempts to be made, specifying the spread over time at which such contacts should be made, and addressing the issue of trying to convert soft refusals.

In the case of telephone surveys, there is a need to establish the number of attempts that should be made for numbers that are frequently busy, not answered, or answered by answering machine, and also standards set for the calling patterns to get through to such numbers. There may be significantly different potentials to obtain a response from a phone number where there are repeated requests for a call back, compared to one that is simply busy on a number of consecutive attempts. Similarly, one where an answering machine is encountered on three or four attempts may have a different potential to yield an eventual good contact, than one where no answer is obtained on a sequence of attempts.

4.4.3 D-3: CLASSIFICATION OF CONTACT OUTCOMES

There is wide variability in reporting of contact outcomes. There are two issues at least under this topic:

- The actual classifications of the outcomes that need to be standardised.
- There needs to be a standard about tracking the history of contacts for a household address, telephone number, etc.

There is an interdependency here with response rates, in that the list of outcomes must be able to be readily grouped to allow computation of the response rate, in whatever way it is eventually determined to standardise that calculation.

4.4.4 D-4: WHO SHOULD BE SURVEYED

Traditionally, household travel surveys have collected data on all household members over the age of 5 years, on the assumption that any younger children would travel with the non-working mother, who would, therefore, provide complete data on the movements of any very young children. In many current societies, this can no longer be assumed, and consideration needs to be given to recording the travel patterns of all young children.

Another issue that arises in household travel surveys is whether to survey persons living in group quarters. In many instances, those living in group quarters do not travel, e.g., prison inmates, those in hospital, some types of elderly and infirm care facilities where residents are bedridden, etc. However, other types of group quarters may produce large amounts of travel, e.g., university dormitories, military facilities, etc. Guidance is needed on whether or not to survey these persons, or whether some group quarters should be included and others not. Cost of inclusion, effect on the data collected, and eventual use in travel-demand modelling must all be considered as pertinent issues for determination of this guidance.

4.4.5 D-5: PROXY REPORTING

Proxy reporting is an issue in any type of personal interview survey, whether by telephone or face-to-face. It has substantial impacts on the completeness of the trip records obtained. When there is no completed diary, proxy reporting under-reports trip making very significantly. Even when a diary has been partially completed, proxy reporting has been shown to result in missing trips.

There are three significant problems relating to proxy reports. First, in countries such as the USA, protocols for human subjects studies usually require that no person under the age of 15 or 16 is involved in interviewing. Second, young people are hard to find at home for an in-person retrieval of data and may also be more reluctant to fill out survey forms. Third, adult men are more likely to decline to speak on the telephone and be interviewed for their data. In fact, anecdotally, it appears that the diaries for adult males are often filled out by one of the adult females in the home, with the result that even the diary record may be significantly incomplete for these individuals.

The issues to be addressed include:

- Under what age should proxy reporting be considered a requirement?
- What proportion of reports from people above this age may be accepted for proxy reporting when there is either no diary present, or when there is a diary present?

Reducing allowable levels of proxy reporting may lead to greater difficulty in obtaining completed households and therefore run counter to other efforts to increase overall response

rates, and will also increase the overall cost of the survey. It will be necessary to assess carefully the balance between increasing cost and difficulty of completing households and improving accuracy through reducing allowable proxy reporting.

Standards may be different according to whether or not a diary is available for proxy reporting. The standards will also need to address how to handle persons who are absent when reporting is done, sick, injured, mentally incompetent, or those who simply refuse to be interviewed when the reporting is done.

4.4.6 D-6: COMPLETE HOUSEHOLD RECORD DEFINITION

Surveys have been widely variable in what is considered acceptable as a ‘complete’ household record. The range appears to be from no missing persons to allowing as many as half of the household members to be missing from the record.

There are two aspects to complete household definition:

- The number of eligible household members from whom data are obtained
- The completeness of the information provided by each household member (this relates back to issues relating to ‘core’ questions – see I-1).

As with proxy reporting, setting standards that are overly strict on what constitutes a complete household will lead to increased survey costs, greater loss of sample, and the potential introduction of more bias, while at the same time reducing the potential errors in the sample obtained.

In examining this issue, it will be necessary to consider how well imputation (see A-4) can repair household-level data and keep it usable, in order that the costs of setting tight standards on complete households do not become too burdensome.

4.4.7 D-7: INCENTIVES

There is considerable difference of opinion among transport professionals – and in different countries - as to whether incentives should be offered or not. There is also a complete diversity of what is offered for an incentive. Incentives have ranged from a gift to a significant payment of money, and some are offered only to those completing the survey while others are offered to all potential respondents.

The standards should address whether or not incentives should be offered, whether incentives should be pre- or post-incentives, and what form incentives should take. There may be important cultural differences here, relating both to the need for and social acceptability of providing incentives and the forms that these might take. Standards would also be useful on how to present the incentive to prospective respondents, since Tooley (1996) points out that the wording used in offering a preincentive is almost as important as the incentive itself.

4.5 N: Non-Standard and New Surveys

Although most of this document deals with household questionnaire surveys (face-to-face, telephone or mailback) that collect information on daily travel behaviour, it is important to recognise that there are other types of survey that can contribute valuable data – and to recognise the growing interest in freight movements.

4.5.1 N-1: MODE AND SITE-BASED SURVEYS

Specialist surveys (e.g., of bus patronage) and augment samples for household surveys are often conducted away from the home, using location or choice-based samples. For example, a number of past household surveys have used an on-board bus survey to augment the sample for transit trips. However, the nature of the on-board survey is usually significantly different from the nature of the household travel survey. In other cases, there are special generator surveys and workplace surveys that are used to supplement the data from the household travel survey.

This topic area addresses issues of the collection of non-household samples and their suitability in different circumstances. It should also specify minimum requirements for the methods of collection and consider such concerns as specifying the minimum required questions.

An example of what may be considered here is to require that certain household characteristics be included in the on-board survey form, to ensure compatibility and the ability to link data from the different samples. Similarly, questions about access to transit and egress from transit are of considerable importance and need to be included in a format that matches the travel, activity, or time-use diary used in the household survey. Another area of concern is using the same categories on such questions as income, age, and other similar questions.

4.5.2 N-2: SP SURVEYS

Many travel surveys include the collection of Stated Preference (SP) or Stated Choice data. Standards and guidelines are needed for designing instruments. Among other things, these would need to consider:

- The need for contextual data to be collected at the same time
- The number of possible alternatives that respondents can be asked to handle (Hensher and Stopher, 2000)
- The number of attributes that can be included in the design
- The number of levels of each attribute that can be included (Stopher, 1998)
- How far the levels of the attributes can depart from current experience of the respondent
- The number of treatments that an individual respondent can be asked to handle
- Whether the order in which treatments are offered has an effect on choices
- The need for orthogonality in the design
- How best to administer the SP experiment – e.g., by paper and pencil, on laptop computer, etc.

4.5.3 N-3: FREIGHT SURVEYS

Much less has been written about the conduct of freight surveys, while the issues involved are often more complex than in the case of person travel surveys. For example, there is the question of the basic unit of analysis (varying from a parcel to a shipment of raw material) and identifying who is responsible for taking particular transport decisions (the customer, the supplier, the transport manager, the driver, etc.). Many areas are also experiencing a rapid growth in light van traffic, linked with the increasing out-sourcing of activities and an emphasis on just-in-time deliveries and e-commerce.

As part of setting standards in this area, there is a need to identify the basic issues that need to be addressed in designing and executing various kinds of freight surveys, and to consider the uses to which the data will be put.

Issues will include, but not be limited to:

- The case for vehicle versus commodity-based surveys

- Trip versus tour-based data collection
- Standard definitions of vehicle types, commodities, etc.
- Identifying key decision makers
- Sampling units and sampling strategies
- Appropriate survey instruments

4.5.4 N-4: GPS SURVEYS

There is growing interest in the use of GPS devices to collect data on a subsample of households in a household travel survey. The GPS is capable of providing very precise information about the locations to and from which people travel, the times of their travel, the routes used, and even the traffic conditions along the route of travel at the time of travel. At present, this is largely an experimental procedure, although it is moving rapidly forward as a mainstream activity in household travel surveys. It also has considerable potential in freight surveys for tracking vehicle movements.

There is scope for setting standards and providing guidance on a number of aspects of such surveys. This would include sample sizes and methods of drawing samples, geographic and sociodemographic distribution of the sample, number of days for which GPS data should be collected, minimum hardware specifications for the GPS devices, use of incentives, methods for deployment of the devices, methods of return of the devices, etc. However, at this time, it is probably too early in the development of such surveys to set standards, and the emphasis should be on early guidance.

4.5.5 N-5: INTERNET SURVEYS

This is another mechanism for obtaining data that is emerging at the present, and has yet to undergo extensive field testing. Nevertheless, like GPS surveys, it is likely to be a technique that will evolve rapidly and, if successful, be incorporated more and more frequently as a potential means for a household to provide the data from a household travel survey. Again, it is an experimental procedure that is not yet in the mainstream, although several current surveys are working to offer internet as an alternative means of response for a number of households.

There is enormous potential for setting minimum standards, and providing guidelines. These may address such issues of how to provide access to the web sites, the type of graphics and other materials to be provided, building in cross checks on data and cross-referencing travel of other household members, encryption, and a variety of ethical issues that will arise with internet surveys. As with GPS, however, it is considered that this is an area where guidance rather than standards is appropriate at this time.

4.5.6 N-6: MIXED MODE SURVEYS

Mixed mode surveys, using multiple survey instruments (e.g., telephone survey, mailback, internet) are becoming increasingly common. They raise a wide range of comparability issues, from sampling strategies (S) through instrument design (I) to data merging and data analysis (A).

These hybrid surveys are carried out in two main contexts. First, in parallel, where the use of different methods may be necessary to reach different population segments, but with the aim of obtaining similar information across all population segments. For example, a mailback survey may be needed to reach population segments not accessible by telephone; or an internet survey may be more effective at reaching certain target groups than a face-to-face interview. Mixed mode surveys may also be required where some population segments are illiterate (or have poor eyesight) and cannot take part in a written, self-completion survey.

The second situation involves the use of different survey instruments sequentially. For example, rail users might be asked to complete an on-mode, self-completion survey, with a sub-sample being asked to take part subsequently in a face-to-face SP survey. Or, a sub-sample of respondents completing a telephone survey might be asked to take part in qualitative research, to expand on the findings of the main survey.

The first priority here is to identify the kinds of compatibility problems that can arise, at all stages of the transport survey process identified in Figure 1. Following this, there is a need to provide guidance or define standards in relation to the key issues that have been identified.

4.5.7 N-7: MULTI-CULTURAL AND MULTI-LANGUAGE SURVEYS

With the increasing tendency for metropolitan areas throughout the world to house a more multi-cultural, cosmopolitan population, and the growing interest in pan-national surveys (e.g., European-level surveys), there is increasing diversity of first languages and cultures among samples selected for transport surveys. In addition, a growing emphasis on social issues in many countries places greater emphasis on fully representing hard-to-reach groups (e.g., socially excluded groups) within travel surveys.

These changing circumstances require the development of more sophisticated sampling frames and customised methods for engaging the interest of the relevant target groups. Survey instruments also need to be developed that are applicable to people from different cultural backgrounds, and that are capable of effective translation into several languages within the one sample. In addition, there are problems of low literacy levels among segments of the urban population that place additional language demands on the instruments that are used.

The first priority is to identify the issues that need to be addressed by those carrying out multi-cultural and multi-language surveys, from pilot surveys and pre-tests (P), through instrument design (I) and sample selection (S), to household data collection procedures (D) and data coding (C) and analysis (A). There is then a need to develop guidance or define standards in each of these areas, as appropriate.

4.5.8 N-8: MULTIDAY AND MULTIPERIOD SURVEYS

The limitations of one-day cross sectional travel surveys are well known, in relation to model formulation, behavioural understanding and policy assessments. In particular, issues relating to day-to-day variability in travel behaviour require richer data sets, as do studies that seek to assess the impacts of major policy interventions (e.g., a new subway line) on travel patterns. The former requires multiday data sets, and the latter multiperiod data.

Issues where guidance needs to be provided - and in some cases possibly standards – include:

- Clarification of the kinds of behavioural and methodological questions that require either multiday or multiperiod data to address them, rather than one-day or repeated cross sectional surveys.
- Any inherent differences in the design of the survey instrument.
- The appropriate number of days over which to collect multiday data, in different circumstances.
- The appropriate time periods between waves of panel surveys, and the number of waves that can be run successfully.
- Methodological issues relating to response and drop out rates in both types of survey
- Appropriate analytical tools for gaining the best from these richer data sources.

4.5.9 N-9: TIME-USE AND ACTIVITY SURVEYS

Recent developments in travel behaviour research and in the use of activity-based models have led to a growing interest in collecting activity/travel and time use diaries. These instruments are relatively new to travel researchers and practitioners, and guidance and standards are needed in respect to all aspects of instrument design, in particular. Implications for data coding and analysis also need to be addressed.

Two broad approaches to recording activities have been adopted. One is ‘episode-based’, whereby the respondent enters information each time a new activity or trip is undertaken, whether these occur one minute or several hours apart. The other tradition, drawing upon the time use literature, is ‘time interval-based’; here respondents enter details of their current activity(ies) at regular intervals (e.g., every 15 minutes). While the former approach may be better at picking up short duration trips, the latter is likely to reduce under reporting of activities and may better handle multiple activities (see I-5).

There is a need to identify the pros and cons of both approaches, and to provide guidance on which approach might be more or less appropriate under different circumstances. The potential of these methods for examining issues such as the use of the internet and other substitution of in-home for out-of-home activities should also be addressed. Guidance should also be provided on instrument design, with particular emphasis on how best to capture more conventional data requirements, including start and finish times of trips, travel modes and segments, and the attributes of modes. Attention also needs to be paid to the problems of ‘trip recovery’ from time-interval diaries.

4.6 E: Survey Execution

Here we deal with practical issues of administering the chosen survey instrument, with particular emphasis on situations where surveys are interviewer administered, either by telephone or in person.

4.6.1 E-1: INTERVIEWER TRAINING

It is axiomatic that the quality of the interviewers for a survey determines the quality of the resulting data. Poorly trained interviewers can effectively destroy the best survey design, simply by not adhering to procedures and by misunderstanding how the survey is to be conducted.

There are several elements to effective interviewer training:

- Explanation of the purposes of the survey;
- Clarification of any difficult-to-understand concepts involved in the survey (but, it is to be hoped, not in the survey instrument);
- Rehearsal of the interview process with other interviewers, with supervisors listening in remotely; and
- On-going listening in to interviews and re-training throughout the survey process.

Some of these items can probably be standardised. For example, that interviewers must receive at least minimal training in the specific purposes of the survey. And that each interviewer for a survey must conduct at least one interview of another survey interviewer, while being monitored by a supervisor, before being permitted to conduct any interviews of the selected sample.

Another possible element of a standard or guidance could be the development of a FAQ (frequently-asked-questions) listing for use by interviewers in responding to questions from potential respondents.

4.6.2 E-2: ETHICS

An extremely important area of survey design is that of ethics. Several survey research organizations have promulgated codes of practice and regulations that seek to set ethical standards in the conduct of surveys of various types. Among the organizations to do this are the Council of American Survey Research Organizations (CASRO, 1997), the Marketing Research Association (MRA, 2000a), and the European Society for Opinion and Marketing Research (ESOMAR, 1999). The MRA has also published a “Respondent Bill of Rights,” (MRA, 2000b).

In addition to general ethical issues that are covered in such items as the MRA Respondent Bill of Rights, specific issues that should be addressed in a code of ethics for personal and household travel surveys are the following:

- Minimum age for interviewing;
- Confidentiality of all information provided by respondents, together with names, addresses, telephone numbers;
- Respondent’s right of refusal to answer any question, to not participate, or to terminate the interview; and
- Definition of a hard refusal.

It will not be sufficient simply to list the various standards, but in a number of cases, more explicit guidance on the ethical standards will be required. For example, there is the question of confidentiality. Virtually every household and personal travel survey includes statements to potential respondents that all answers will be kept strictly confidential, and that only data on groups of households will be published. However, there is still a question of at what stage such information as telephone numbers, addresses, and names should be removed from the survey record, and how protection can be given that no traceable information will be released to third parties.

4.6.3 E-3: INITIAL CONTACTS

Many refusals take place either before anything is said, or before the interviewer has said more than a few words. McGuckin et al. (2001) have reported that many telephone survey refusals occur before any questions are asked and the interviewer has only begun an opening sentence. Similarly, when the first contact is a letter or other mailed item, the initial contact will often determine whether or not the household will proceed with the survey. There is a need to study those opening words that may avoid an up-front refusal, and recommend wording for face-to-face, telephone contact and mail pre-recruitment contact that may help to get past the initial refusals.

A second topic here is whether the initial contact should be in person or a telephone call, or whether there should be a pre-contact letter. There is no agreement as to whether it is better to contact households first by mail, or if a cold-contact personal visit or telephone call is just as effective.

With respect to using a pre-contact mailing, this is restricted when using Random Digit Dialing, because it will not be possible to obtain a street address for households with unlisted numbers. Similarly, if one were to start by sampling street addresses, it will often not be possible to obtain telephone numbers for the many households with unlisted numbers.

4.6.4 E-4: FOLLOW-UP CONTACTS

In most applications that use a telephone contact of any type, interviewers are assigned to undertake the work on a more or less random basis, in relation to the telephone numbers. When

there are multiple contacts to be made, each contact may be undertaken by a different interviewer, and the interviewers will not usually provide their names to potential respondents. This is likely to contribute to high refusal rates.

An alternative procedure has been pioneered by Socialdata in both European and other contexts, in which a household is assigned to a specific “motivator” from the first telephone contact and a complete contact history is maintained such that anyone can contact the household and be fully aware of past contact details. During that initial contact, the interviewer provides the name and telephone number of the motivator, where she or he can be reached. All subsequent contacts with the household are made with reference to this motivator. Household members are encouraged to call this person if they have any questions about the survey. It has been reported that this protocol has shown increases in response rates of almost double the rates from standard procedures of interviewing (Brög et al., 1983).

The second issue relates to handling refusals to telephone interviews. Refusals are often classified into “soft” and “hard” refusals. A “hard” refusal is usually defined as one in which the potential respondent tells the interviewer not to call again, or uses other very specific language to indicate that this person will not consider responding to the survey. All other refusals are considered to be soft refusals. In many recent surveys, soft refusals are assigned to certain interviewers who have a proven track record of being able to turn around such refusals. Because significant effort is often assigned to converting soft refusals, it would be beneficial to know in what proportion of cases the conversion results in a completed survey.

4.6.5 E-5: RESPONDENT QUESTIONS

In virtually any personal or household travel survey, respondents will have questions about the survey, about those conducting the survey, and will also require a means to check that this is a bona fide survey.

Standards and guidance in this area should include the establishment of the toll-free telephone number, the agency number for checking on the genuineness of the survey, the use of a FAQ leaflet, the use of a web site, and possible contents of each of the FAQ document and web site.

4.6.6 E-6: METHODS TO REDUCE INCORRECT REPORTING OF NON-MOBILE PERSONS AND HOUSEHOLDS

There is a level of non-mobility in the population at large on any given day, representing those individuals who, for a variety of reasons, do not leave their home on a specific day. However, non-mobility reporting can also be a method of subtle non-response.

In the case of trip-based surveys, there has been some success reported from asking for reasons for not travelling, which seem to reduce the incidence to some extent. Standards would include possible specification of questions that should be asked of claimed nonmobiles on the diary day, and also alternative ways of questioning about travel that make this an ineffective method of nonresponse.

An additional issue affecting this topic is the need to be sure that there is an indicator in the data set that distinguishes between cases where a person indicated that he or she did not travel and the case where a person refused to provide travel data.

4.6.7 E-7: ITEM NONRESPONSE

Item nonresponse relates to incompleteness within a response, as compared to a failure of the subject to respond at all (i.e. unit non-response). As such, there are varying levels at which

nonresponse can occur, including: Household, Person, Vehicle, Activity, Round trip, Trip, Trip stage, or Items describing any of the above elements.

It is affected by question wording, question order, and survey content. It also relates to training of interviewers on what to do when people refuse to answer a specific question. This is partly an issue to be dealt with through interviewer training procedures (E-1), and partly through the building-in of automatic checks in CATI, CAPI and internet surveys (see A-2).

There may be items to which no response is obtained in some cases, notwithstanding the best efforts of the interviewer. Two remedies are possible here. First, households can be called back by either the same or a different interviewer within a days or so, and the same question may be asked. In many cases, particularly if a different household member answers on the second attempt, the item will be completed. The second remedy is data imputation (see A-4).

The main items that require the setting of standards under this task relate to review of data and call backs for completion of the data. It is generally not possible to obtain reasonable correction and completion of responses when a significant amount of time has elapsed since the original interview. Therefore, there is a need for a standard concerning review of data, and checking for item nonresponse. The standard should stipulate the amount of time that is the maximum that should be allowed to elapse before checks are made and a re-contact can be initiated with the household to correct the erroneous or missing data. Second, there is a need to set standards as to how many attempts to correct the data should be made. Third, there should be guidance on what to do if different responses are obtained when the same or a different interviewer is used.

4.6.8 E-8: RETENTION OF DATA ON INCOMPLETE HOUSEHOLDS

In many recent surveys using CATI in the USA, either the software used for the CATI or the desires of the survey firm or client have resulted in destruction of data on incomplete households as soon as the full sample has been obtained. This can represent a serious loss of information on several elements of the survey, such as, the design of certain questions that may result in premature terminations of the interview, information on the biases in nonrespondents, and partial information that may be usable for some aspects of survey analysis.

It would thus be valuable to specify standards for retention of data on incomplete households. Such standards would need to define whether all incomplete household records should be retained or only those meeting some minimum criterion of completion.

4.7 C: Data Coding (C)

Once survey data has been collected there are a number of aspects relating to coding that can have a major influence on the quality and compatibility of the data.

4.7.1 C-1: CODING STANDARDS FOR MISSING VALUES, USE OF ZERO, ETC.

There is no agreement among recent household surveys on what to use for flagging missing values, and other aspects of setting coded values for non-numeric data.

Because computers tend to treat zeroes and blanks as the same, it appears that the first standard that should be set in this area is that blank should never be a legitimate code, and that all fields must contain numeric data. This will require additional data entry in locations where there is a legitimate skip. Given that such a standard is implemented, then standards for other aspects, such as codes indicating certain categories of missing information, can be set. In addition, if a

“no blanks” standard is set, then zero can be used as a legitimate value among the response codes, and will not lead to confusion.

For example, the codes for missing variables as currently used by the USA NPTS are to use ‘97’ to indicate that the variable was not applicable or a legal skip, ‘98’ to indicate that a response of “Don’t Know” was provided, and ‘99’ to indicate a refused response.

The next requirement is to specify as a standard that there should be correspondence between the numeric values of a categorical variable and the codes. For example, if there are 0 workers, this should be coded as 0, one worker as 1, 2 workers, as 2, etc. This coding is again contingent on adoption of the “no blanks” standard, since otherwise, the use of 0 could be confused with a blank, which might indicate a missing value.

A further issue is to establish standard codes for binary variables (e.g. that 0 be used always to indicate “No” and 1 be used always to indicate “Yes.”).

4.7.2 C-2: CODING COMPLEX VARIABLES

A closely-related issue is that of coding complex variables, such as industry, occupation, income, educational level, etc. Here, it is questioned whether it would be desirable for all household travel surveys to use similar or identical categories in the response sets and to code them identically. Standardisation of categories is dealt with elsewhere (under the I-codes).

Any such codes should be set up in such a way as to allow varying levels of aggregation, depending on the needs of any particular survey. In general, this could be done by setting up multi-digit codes, where appropriate, in which the first digit represents the coarsest level of aggregation that would be used, the second digit provides greater disaggregation, and the third digit (if applicable) could provide even greater disaggregation.

4.7.3 C-3: GEOCODING STANDARDS

Usually, about 60 to 70 percent of origins and destinations in a personal travel survey are home, a workplace, or a school and are fairly readily geocoded in most instances. The problem arises with the remaining 30 percent, which may include shops, stores, restaurants, friends’ homes, gyms, etc. Many of these are provided with too little information to allow automated address-matching procedures to provide an accurate geocode.

When geocoded information is missing from the final record, the trip will usually not be usable for any model estimation except trip generation. Hence, if a survey achieves 80 or 85 percent geocoding, this will imply that about 15 to 20 percent of the data are lost for most model estimation purposes, and particularly that trips apart from those between home and work and home and school will likely be under-represented for modelling purposes. Where tour-based modelling is applied, the problem is even more serious.

The geocoding standard would be one that specifies the proportion or percentage of trip ends that should be successfully geocoded in a household or personal travel survey. The geocoding standard is to some extent related also to the level of geocoding that is expected (see C-4).

4.7.4 C-4: LEVEL OF GEOCODING TO BE PERFORMED

There is currently a wide variety of levels of geocoding, ranging from TAZs (traffic analysis zones) and census tracts, to specific latitude and longitude. Unfortunately, once an aggregate level of geocoding is decided on, finer levels of disaggregation cannot be obtained. It therefore seems advisable that the standard, if one is to be set, should be towards the disaggregate end of

the scale. There will be a trade-off here between the level of disaggregation and what is a reasonable cost for the geocoding activity.

4.7.5 C-5: GEOCODING OUT-OF-REGION ADDRESSES

The final geocoding issue is that of what to do with addresses of origins and destinations that are outside the study region.

There are two common practices applied to trips with one end outside the region. The first is not to geocode the external trip end, but to use a single identifier that flags that this is an out-of-region trip. The second is to code the trip end to the external cordon station where the trip exits the region. Both of these may be used together in some surveys, where road-based trips will be coded to external cordon stations, and other trips (air, rail, etc.) would be coded by an identifier flag. With increasing attention being paid to long-distance modelling, and with the small number of intercity trips that are usually to be found in any travel survey, it may be more appropriate to attempt to geocode internal-external trips to a more useful location code.

4.8 A: Data Analysis and Expansion (A)

For most analytical purposes, survey data cannot be used in its raw form. It may have to be re-weighted to represent the population of interest, and questions will arise as to how to treat missing data. In addition, there will be requirements to fully document and archive the data set. Standards in all these areas would be very valuable.

4.8.1 A-1: ASSESSING SAMPLE BIASES

Whenever response rates are less than 100 percent, or if there is incomplete adherence to the sample design, there is potential for bias to exist in the sample. Generally, assessment of bias is done by comparing certain key variables that describe persons and households with recent census data and other possible benchmarking sources.

The standard will need to address which variables should be examined to determine bias, such as, household size, income, age, gender, vehicle ownership, working status, and education.

4.8.2 A-2: CROSS-CHECKS IN DATA COLLECTION AND DATA REVIEW

CATI and CAPI surveys offer enormous potential for cross checks on data quality in real time as the survey progresses. Almost every travel survey includes instances of people forgetting to report a trip back to home at the end of the day, or an activity at home after the last trip of the day. Work trips by people who report that they are not workers are another common occurrence in surveys. Another problem in activity and time-use diaries arises when people do not include activities at a place between trip segments, e.g., waiting at a bus stop, or parking a vehicle.

There would appear to be justification for the inclusion of standards relating to the cross-checks that should be built in to CATI and CAPI-type interviews, and the procedure of checking data after they have been retrieved and before they are considered complete enough to be free of further checks.

4.8.3 A-3: WEIGHTING AND EXPANSION OF DATA

Data expansion is usually only necessary if one of the uses of the data is to provide regional descriptions of households, persons, travel, and vehicle ownership and use patterns. It is generally not necessary for model estimation. There is a need to provide a simple explanation of how to calculate expansion factors, with easy examples.

Data weighting is necessary when biases are known or found to exist in the data, or when a stratified sampling design is used or other disproportionate sampling technique. In the latter case, weighting is needed for any activity that would combine data from different strata of the sample. This is important in most modelling, because the strata are not usually identical to subdivisions of the data for any modelling steps, with the possible exception of trip production estimates. Even trip attraction modelling should involve the use of weights, if a disproportionate sample has been drawn.

The standard for weighting would first identify the circumstances under which weighting is required. This should take account of such factors as households with multiple telephone lines, households with missing person records (depending on the definition of a complete household), as well as biases contributed by differential response rates. The second part of the standard will address how the weights should be calculated and applied.

4.8.4 A-4: MISSING DATA IMPUTATION

It is appropriate to distinguish between imputation and inference. Both are types of data repair. Inference has probably been used for many years in transport surveys, while imputation has been used rarely.

Inference is the deduction from other information of missing information. For example, if a person did not indicate whether or not he or she was a worker, but the person reported trips to and from work, it is a reasonable inference that the person is a worker. As part of the standardisation effort, some rules on inference, to ensure that it is fully objective, would probably be wise. This is particularly true where some variables may not be safe to infer, because of changing social structures and mores.

Imputation is the substitution of values for missing data, based on associative rules or other procedures, such as neural networks. The most common method of imputation is hot-deck imputation. In setting standards, it is probably appropriate to rule out some forms of imputation, such as imputing the mean value, cold-deck imputation, etc. It is probably not appropriate to specify a particular imputation method, because this appears to be an area in which new and better methods may be developed in the near future. However, there are issues of how much data can be imputed, and under what conditions imputation should be done. Examples here are questions of whether missing travel can be imputed, missing persons can be imputed, or only an occasional item should be imputed.

For imputation, first, it is necessary to set a standard that there must be a flag to indicate each and every instance of data imputation in the data set. Second, a minimum standard on imputation method appears to be warranted. It is likely that the minimum standard that should be set is for imputation to be done using random hot-deck imputation (Lessler and Kalsbeek, 1992). A third question that arises with imputation is whether or not hot-deck imputation should be done with or without replacement (other than in circumstances when there are more recipient observations than donor observations).

Some additional issues arise with imputation, because of its need for setting up categories of reasonably homogeneous units for classifying both donors and recipients. The first issue is what sort of sample size might be necessary to achieve reasonable homogeneity among households or persons to permit random hot-deck imputation to be undertaken. Another related issue to this is the specification of sociodemographic data that should be collected in traditional household travel surveys to allow categorization of households and imputation of missing data.

The remaining issues relating to standards of imputation concern how much can be imputed. Recently, some exploratory research has taken place in which imputation has been used to restore missing trips and missing persons in a household (Dudala and Stopher, 2001). This appears to take imputation far beyond the bounds of what has been done to date by statisticians working on imputation. It is necessary to consider whether this is an appropriate procedure to use, and to determine under what circumstances it can be used.

4.8.5 A-5: GLOSSARY OF TERMS

Many terms are used in surveys generally and in transport surveys specifically that have specialized meanings, which may not be fully understood by all persons working with the products of the surveys. There are many examples of terms that are incompletely defined or that require definition, such as activity diary, trip, pretest, classification or categorization, etc. A standard glossary of terms would provide definitions that, if adopted, would lead to greater clarity in the presentation of documentation on surveys and to understanding of the data recorded, and much greater consistency between surveys.

4.8.6 A-6: DATA ARCHIVING

Axhausen (2000) provides a succinct summary of the sorry state of data archiving in transport surveys (pages II-F/4-8). Essentially, there has been little attention paid to archiving past household and personal travel surveys, particularly in North America. Despite the high cost of collecting the data, and of processing it for reasonable use, most household travel surveys of the past remain inaccessible, and many are lost completely. Axhausen (2000) points out that the recent Synthesis on Household Travel Surveys (Stopher and Metcalf, 1996) does not even mention data archiving. This is because it has simply received no attention by most agencies in the U.S.A.

Standards for archiving are extensive and can be complex. It is not sufficient just to store the data on some accessible medium and place this record in a location where it can be maintained. There are needs to document what is archived, to provide explanations of the codes used, copies of the instruments used, details of editing performed, etc. It is possible that minimum standards would call for storage on at least two media types, with restoring of the data periodically to new media, storage of a data dictionary, coding guide, and data on edits performed in the data, and a copy of documentation describing how, when, and where the data were collected, including details of the sampling methodology. Archived data should be sufficient to allow any interested individual to access the data and use it in a statistical program of his or her choice, with complete interpretation of the data, and ability to replicate all edits performed, estimate weights and expansion factors, and reproduce any computations previously performed with the data.

4.8.7 A-7: DOCUMENTATION

This is related to the issue of archiving data, in that documentation of surveys should be sufficiently complete for the purposes of archiving, as well as for the purposes of comparing surveys one to another. The standards to be promulgated here refer to minimum documentation standards.

Topics for comprehensive documentation include:

- Sample design, sample size, and sampling procedure
- Sample frame information
- Copies of all survey instruments to be provided as an appendix
- Copies of telephone scripts used in the survey

- Data collection procedures, including the number and type of contacts, who was surveyed, what was permitted in proxy reporting, what constituted a complete household, etc.
- Conduct and results of any pretests and pilot surveys
- Implementation of the survey, including when calls were made, what days or times were avoided for data collection, how answering machines were handled, the period for diary recording, etc.
- How geocoding was done, how missing values were recorded, what data imputation and inference was undertaken, how such inferred and imputed values are flagged, etc.
- Complete reporting of the quality measures, such as standard response rate, non-mobiles, etc.

4.9 Q: Quality Performance Standards

In addition to specifying detailed standards for the various stages of the survey data collection and analysis process, there is a need to identify a few key measures of overall survey quality, that can be used to assess data quality.

4.9.1 Q-1: COMPUTING RESPONSE RATES

Response rate has not been consistently defined in the past and, therefore, reports of response rates have not been comparable. The most common interpretation of response rate is the proportion of eligible households that provide completed interviews (CASRO, 1982, p. 2, Richardson et al, 1995, p. 238, Zimkowski et al, 1997, p. 1-26). Ineligible households are those cases within the sample frame that relate to vacant or demolished dwellings, business addresses, non-residential telephones, or other cases in which no household is associated with the case in the sample frame. Ineligible households are subtracted from the total sample frame to establish the total number of eligible households in the survey. The response rate is established by dividing the number of completed interviews by the total number of eligible households.

One of the problems of implementing the above definition of response rate is to know whether a sample case is an eligible or ineligible household when no response is obtained from that household (Zimkowski, et. al, 1997). It is possible that a household that furnishes no response is an eligible household, but that no one in the household was present on all occasions when approaches were made, or that the household has caller I.D. on their telephone and they elected not to respond to the call. CASRO (1982) suggests that this situation be addressed by determining the proportion of eligible households within the population from those households whose eligibility was determined and assuming that the same proportion of eligible households exists among those whose eligibility is unknown.

Another problem with the implementation of the CASRO definition of response rate is the definition of a completed interview. Survey responses typically contain varying degrees of missing data items, and a standard needs to be set to distinguish between a complete and an incomplete interview, as is addressed in D-6.

Other possible ways of computing a response rate include assuming that all units for which eligibility was not established are ineligible - which undoubtedly over-inflates the response rate - and applying some other rate of non-eligibility to those units for whom eligibility is not determined.

Another concern arises when using survey procedures that involve a two-step or more process to gain completion of the survey. For example, in what has become the standard U.S. survey of a telephone contact for recruitment, mail-out of a survey package, and retrieval of the answers by either mail or CATI, the response rate often reported is that for the last step (retrieval) only, whereas it should be the compound of the response rates from recruitment and retrieval. Because recruitment rates are often fairly low, at say 60 percent or less, and retrieval rates are often between 40 and 60 percent, reporting only the latter rate is misleading, given that the real response rate varies between 24 and 36 percent, in this case.

4.9.2 Q-2: WIDER MEASURES OF QUALITY

It is inadequate to assess personal travel surveys on response rate alone. This ignores numerous other issues of data quality. Included among the possible transport-specific measures of quality are the following measures:

- Proportion of non-mobile households
- Proportion of non-mobile persons
- Average activity rate per household
- Average activity rate per person
- Average trip rate (overall) per household
- Average trip rate (overall) per person
- Average trip rates per household and person for specific trip purposes

In order for these to be effective, it will be necessary to provide clear definition of what is meant by a non-mobile household and person, what constitutes an activity, and how a trip is defined, including any trip linking involved. If trip rates are to be specified by purpose, then the purposes must also be defined clearly. Further, there may be good reason to define trip rates both to include non-mobile persons and households and also to exclude them, since variability in this statistic from survey to survey may make it more difficult to ascertain when a survey is producing comparable trip rate figures. While household-related figures are often useful for the modellers who work on household-based statistics, household-based figures are subject to variations in household size. For this reason, person-based statistics should probably always be used, in addition to any household-based statistics that might be desirable.

Consider should be given to whether there are any additional transport-specific measures that should be included. The standards should address how each of these measures is to be calculated. It will also be appropriate to provide default values for each measure, so that there is some way to know whether a particular value for a specific survey is good or bad.

4.9.3 Q-3: COVERAGE ERROR

Coverage error relates to the differences that appear between the population that was supposed to be represented by the sample survey, and the actual sample obtained. For example, in telephone contact surveys, there is a coverage error of those households without telephones. Similarly, households with multiple lines may be over-represented in the sample, because there is a higher potential for them to be sampled. Each of these represents a source of bias and is part of the coverage error.

4.9.4 Q-4: SAMPLING ERROR

The sampling error of individual variable estimates is measured by the Standard Error of the Estimate (SEE). However, the magnitude of the measure is affected by the units of measurement of the variable under consideration, making interpretation of the value and comparison of values among data sets difficult. To eliminate this effect, the coefficient of

variation (SEE divided by the estimate) provides a dimensionless measure of variation of the estimate about the mean and allows meaningful comparison among data sets.

To obtain a measure of sampling error for a data set as a whole, a procedure must be devised which draws information on sampling error in the individual variables in the data set and combines them into a single, composite value. One approach could be to identify the sampling error of a predefined set of key variables and take the average, or weighted average, of the values as a measure of the overall sampling error in the data set. The 'key' variables could be selected from among the 'core' variables required in any survey. An alternative approach could be to identify those variables most relevant to the purpose, or purposes, of the survey and measure the sampling error on each variable (Cambridge Systematics, 1996, p. 5-9). Consistent with the idea of "total design" promoted by Dillman (1978) in which the idea is promoted that the quality of a process is only as good as the weakest link in the process, the sampling error of the data set would be related to the highest sampling error among the measured variables.

4.9.5 Q-5: PROXIES

It is generally accepted that proxy reporting produces data that are inferior in quality to those collected directly from respondents (see also D-5). Since cases involving proxy reporting are usually identified, it would be easy to determine the proportion of cases in the data set that relied on proxy data and use that as a measure of data quality. The measure may also be used as a standard in a RFP to stipulate the maximum proportion of survey responses that may be made by proxy reporting.

4.9.6 Q-6: VALIDATION STATISTICS

Validation is the process of verifying the accuracy of collected data by repeating the data collection process with a sample of households already surveyed. The validation process is not carried out by the same interviewer used in the original interview and the entire survey is not usually repeated. Validation serves to ensure that interviewers did a thorough job and that consistency in the data collection process was achieved among the interviewers. However, the results of the validation process can also be used to assess the quality of the data by observing the number of discrepancies obtained and the variance of this number among the sampled households.

What to include in a validation procedure and how to interpret the results would need to be established. The standard should address what should be asked in the re-interview, how the re-interview should be conducted, and how to interpret the results of the re-interview.

4.9.7 Q-7: DATA CLEANING STATISTICS

This measure of survey quality is simply a count of the number of data items that were found to be out of range, inconsistent with other values, or otherwise incorrect in some way. It is a statistic that would also be most useful in a pilot survey or pretest, where it would indicate problems with specific questions or the overall instrument design.

There are two necessary steps to defining a standard for this quality measure. First, it is necessary to specify a standard for the checks that should be performed to detect out-of-range and inconsistent values, which would specify which variables should be the target of such an investigation, and how inconsistencies would be detected. Second, the standard would stipulate how the edits are to be flagged and counted and to what they are to be compared to determine reasonableness.

4.9.8 Q-8: NUMBER OF MISSING VALUES

The number of missing items in the data can be used as a rough measure of the quality of the data. This can be expressed as a percentage of all data items in the data set to portray the proportion of all data items for which data is missing. It is important, obviously, that the missing items counted in this manner are truly missing (i.e., refused or no answer) and do not include item values for which a response is not possible (i.e., a 'not applicable' or skipped item). This standard would require that the standard on coding of missing data, discussed in subsection C-1 is established.

The number of missing items can be reduced by imputing values for at least some of the missing values. Thus, some of the effects of the missing data may be reduced by imputation. However, imputation adds no new information to the data and, therefore, the original measure of the proportion of data that is missing may be appropriate, whether imputation is conducted or not.

4.9.9 Q-9: ADHERENCE TO QUALITY GUIDELINES

The management of the travel survey process can be assessed by determining to what extent accepted quality guidelines were used in the execution of the survey. An initial checklist that could serve this purpose is the "Survey Design Checklist" listed in Appendix C of the book by Richardson, Ampt, and Meyburg (1995). However, it would be better to develop a set of questions specifically aimed at identifying the degree to which a survey has adhered to quality guidelines as required by, for example, the International Standards Organization (ISO) and translated into requirements for travel surveys (Richardson and Meyburg, 1996).

It is proposed that each travel survey should be assessed by determining to what extent it has complied with the minimum standards established by this conference.

4. CONTRIBUTIONS OF THE WORKSHOPS TO ELEMENTS OF THE FRAMEWORK

Table 2 identifies up to nine broad areas in which the fourteen workshops might make contributions to the framework, by offering guidelines or proposing standards for one or more of the specific elements within that area. It is expected that all workshops will contribute to section Q on Quality Performance Standards, to varying degrees. In addition to the areas and elements identified in Figure 1, most workshops will be considering the relevance of advances in their area in terms of the purposes for which the survey data is used, in relation to one or more of: understanding travel behaviour, policy development, and modelling.

The table distinguishes between areas where the workshop is likely to make a major contribution (dark shading) and other areas where some more limited input might be expected (lighter shading). However, workshops are invited to review all the areas and the 66 individual elements, and to contribute additionally to others where the members feel a wish to do so.

Table 2: Suggested Areas Where Workshops Might Contribute to the Framework

Topic Area	Workshop													
	A-1	A-2	A-3	A-4	A-5	A-6	A-7	B-1	B-2	B-3	B-4	B-5	B-6	B-7
P Pilots & Pre-tests														
I Instrument Design														
S Sample Design														
D Data Collection														
N New Surveys														
E Survey Execution														
C Data Coding														
A Analysis & Expansion														
Q Quality Performance														
Understanding														
Policy														
Modelling														

KEY:

- | | |
|--------------------------------------|--------------------------------------|
| A-1: Mixed Mode Surveys Surveys | B-1: Time-Use and Activity |
| A-2: Unit Non-Response | B-2: Item Non-Response |
| A-3: Multi-Cultural & Multi-Language | B-3: Impact of New Technologies |
| A-4: Instrument Design | B-4: Respondent Burden |
| A-5: Stated Response Surveys | B-5: Qualitative Methods |
| A-6: Non-Household Surveys | B-6: Freight and Commercial Vehicles |
| A-7: Multiday and Multiperiod Data | B-7: Data Interrogation & Management |

5. REFERENCES

Axhausen, K.W. (2000), "Presenting and Preserving Travel Data," *Transportation Research Circular No. E-C008*, Transportation Research Board, Washington, D.C., pp. II-F/1-19.

Brög, W., Fallast, K., Katteler, H., Sammer, G. Schwertner, B., (1983) "Selected Results of a Standardised Survey Instrument for Large-Scale Travel Surveys in Several European Countries." *New Survey Methods in Transport*, VNU Science Press BV, International Conference "New Survey Methods in Transport", Hungerford Hill, N.S.W., Australia

Brög, W. (2000), "The New KONTIV Design: A Total Survey Design for Surveys on Mobility Behaviour." International Conference on Establishment Surveys – II, Buffalo, New York.

Cambridge Systematics, Inc. (1996), "Travel Survey Manual." U.S. Department of Transportation and U.S. Environmental Protection Agency 564 pp.

Council of American Survey Research Organization (CASRO) (1982), "On the Definition of Response Rates", Special Report of the CASRO Task Force on Completion Rates, Lester R. Frankel (Chmn.), 6 pages.

Council of American Survey Research Organizations (CASRO) (1997), "Code of Standards and Ethics for Survey Research." New York (<http://www.casro.org/casro.htm>) 8 pages.
<http://www.casro.org/casro.htm>.

Council of American Survey Research Organizations (CASRO) (1998), "Survey Research Quality Guidelines.", New York (<http://www.casro.org/guidelines.htm>) 14 pages.
<http://www.casro.org/guidelines.htm>.

Dillman, D., *Mail and Telephone Surveys: The Total Design Method*, John Wiley, New York, 1978.

Dudala, T. and Stopher, P.R. (2001), "Survey Data Repair Using Hot-Deck Imputation," paper presented to the 8th TRB Conference on Transportation Planning Methods Applications, Corpus Christi, April.

European Society for Opinion and Marketing Research (ESOMAR) (1999), "Rules," Amsterdam, The Netherlands, 4 pp. http://www.esomar.nl/codes_3.html .

Hensher, David A. and Peter R. Stopher, "Are More Profiles Better Than Less? Searching for Parsimony and Relevance in Stated Choice Experiments," *Transportation Research Record No. 1719*, pp. 165-174.

International Standard Organization (ISO) (1997), "Quality Management Principles and Guidelines on their Application." Report ISO/TC 176/SC 2/N 376, 19 pp.

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

Marketing Research Association (MRA) (2000a), "MRA Code of Data Collection Standards with Notes," Connecticut, 6 pp. http://www.mra-net.org/industry/code_dcs.cfm.

Marketing Research Association (MRA) (2000b), "Respondent Bill of Rights," Connecticut, 2 pp. http://www.mra-net.org/industry/bill_of_rights.cfm

McGuckin, N., Liss, S. and Keyes, M.A. (2001), "Hang-Ups – Why People Refuse to Participate in Telephone Surveys," paper to be presented at the International Conference on Survey Quality and Innovation, Kruger Park, South Africa, August.

Richardson, A. J., E. S. Ampt, and A. H. Meyburg (1995), *Survey Methods for Transport Planning*. Eucalyptus Press, 492 pp.

Richardson, A. J., and A. Pisarski(1997), "Guidelines for Quality Assurance in Travel & Activity Surveys." International Conference on Transport Survey Quality and Innovation, *Transport Surveys: Raising the Standard*, Grainau, Germany, 30 pp.

Richardson, A. J. (2000), "Behavioral Mechanisms of Non-Response in Mailback Travel Surveys", paper number P00-2551, 79th Annual Meeting of the Transportation Research Board, Washington D.C.

Statistics Canada (1998), "Quality Guidelines." Third edition, Catalogue number 12-539-XIE, Ottawa, Canada,. <http://www.statcan.ca:80/english/freepub/12-539-XIE>

Stopher, P.R. and Metcalf, H. (1996), *Methods for Household Travel Surveys. NCHRP Synthesis of Highway Practice 236*, National Academy Press, Washington, D.C.

Stopher, P.R. (1998), "A Review of Separate and Joint Strategies for the Use of Data on Revealed and Stated Choices," *Transportation*, Vol. 25, No. 2, pp. 187-205.

Tooley, M. (1996), "Incentives and Rates of Return for Travel Surveys," *Transportation Research Record No. 1551*, Washington, D.C., pp.67-73.

Zimkowski, M., R. , R. Ghadialy, an S. Pedlow (1997), "Non-response in Household Travel Surveys." National Opinion Research Center (NORC), prepared for the Federal Highway Administration, October.

<http://www.bts.gov/tmip/papers/surveys/nonresponse/nonresp.htm>