Data Quality Problems in Travel Surveys
An International Overview

Wim F. de Heer and Ger Moritz
Statistic Netherlands

ABSTRACT

Nonresponse in household surveys is a matter of great concern. Data quality may be affected by nonresponse. Not only the attitude of the public, but also sample and survey design, fieldwork approach and characteristics of organization influence response results. Weighting techniques cannot reduce nonresponse bias totally. This will be explained in general and specific for Travel Surveys in 13 countries.

INTRODUCTION

This paper will focus on sampling, response and nonresponse and weighting and their relation to data quality. Data quality refers to the total survey error, which in general is unknown. Groves and others (1-4) distinguish between sampling and non-sampling error, where the non-sampling error can be subdivided into coverage or frame error, nonresponse error and measurement (process) error. As statisticians most of the time pay attention to sampling errors and estimates published by statistical agencies usually cover only sampling errors, the emphasis in this paper will be on the non-sampling error, as far as this is related to nonresponse and non-coverage. Sampling and weighting techniques receive attention in this paper as far as these techniques can be seen as instruments to reduce these types of non-sampling errors.

The goals set for this paper can not be too high, because research into non-sampling errors and especially non-sampling errors in travel surveys, is still in its infancy, although developing rapidly (5). The objective is to describe how serious the nonresponse problem can be, how it can be taken care of and to what extent survey organizations recognize nonresponse as a problem. This will be done in two ways:

- firstly, for the main aspects (‘sample and survey design’, ‘fieldwork strategy, organization and nonresponse’ and ‘weighting’) a general and practical overview will be presented,
- secondly, a description of present practices with respect to travel surveys will be given on the basis of a short mail survey in 13 countries.

This mail inventory was carried out with a short questionnaire which was sent to 15 agencies, mostly governmental agencies. There was no specific selection of agencies used other then the author’s acquaintance with contact persons in different countries. In Annex 1 the questionnaire is presented and in Annex 2 a summary of the information per agency.
SAMPLE AND SURVEY DESIGN OF TRAVEL SURVEYS

Sample Design in Travel Surveys

In all travel surveys mentioned in this paper sampling is used as a technique to reduce costs and to improve accuracy, precision, speed, and feasibility (6). All sample surveys suffer from sampling error, which is due to surveying only a subset of the population when heterogeneity exist among persons in the population (1). Beside that, the sample and survey design have consequences for the data collection methods and techniques to be chosen, which in turn affect the accuracy of the results because of errors of nonresponse and measurement.

Budgetary limits mostly restrict sample sizes and are usually set prior to determining necessary precision. The challenge is to optimize sample schemes to improve cost-efficiency and get the desired precision of the information wanted. In the area of travel surveys, analysis of sample size requirements to get the specified precision is very complicated, because of the very different use of the data (7). No general aim or single objective of analysis exists. Linked to this is the problem of quantification of the main variables of analysis. Optimizing the sample size requires information on the variables of interest of the population. Therefore optimizing sample schemes and sizes to get the necessary statistical reliability is problematic.

Definition of the sample unit needs special attention. The choice of the sample unit relates to the kind of information needed. In travel research sampling issues are complicated because the objective is to get reliable data on travel behaviour in space and time. Not only does the target population need to be defined as accurately as possible, but also the time period and the area that has to be covered requires special attention. The sample scheme should deal with the probabilities of selecting any person of any member of the target population within the given time period (8). Another complication in travel research is the fact that surveying on the personal level alone neglects inter-personal dependencies and cluster-effects on the household level. Besides that individual and household travel patterns have some relation with the location where persons or households live (9).

Defining the sampling frame and sample unit is a very important stage. Constructing or choosing the sample frames can imply the use of a specific method or approach and/or the choice of the measurement instrument. For travel surveys lists of addresses or person- or telephon-registers are usually used as sample frame.

Sampling schemes are dependent on the availability of sampling frames. Ideally the sample frame covers the full population of interest including information on ways to contact the elements of that population. Construction of a perfect sample frame is problematic. Often researchers have to deal with coverage errors, duplications and contact problems. Sometimes the sampled population is more restricted than the target population. Information on the differences between the sampled and the target population must be established in order to the influences on the conclusions drawn from the sampled population. If for example a telephone register is used, this implies telephone contacts. In this case the assumption is, that households owning a phone with a listed number are representative for the target population with respect to the survey’s objectives. But the question is to what extent does this assumption hold true? Most of the time there is no information on the population that is excluded from the frame. There might exist a
coverage error. Random digit dialing offers a solution for the unlisted numbers, but there remains a possible coverage error if a part of the population has no phone at all and it raises questions with respect to spatial representation, weighting schemes, privacy, duplications and over-coverage (9). If the sample frame does not cover the whole target population one should look for additional frames or one should find out whether the omitted part of the population differs from the covered part or not.

Sometimes ‘intercept sampling’ is used if the main objective is to sample trips within a given area and there is no reliable sample frame to sample households or persons. However, the greater the area to be covered, the more a household or persons sample is the appropriate way.

Most common sampling techniques or combinations of these techniques in travel surveys are:

1. **Stratified random sampling**: Geographic stratification is a common practice in travel survey sampling. Over-representation in particular areas (urban) is common to compensate for lower response rates in these areas.

2. **Systematic sampling**: Order of the sampling frame has to be unbiased regarding the target variables.

3. **Cluster sampling**: All elements of the selected clusters can be included in the survey, or, in a second stage, elements from the selected clusters may be drawn (two-stage sampling). In most travel surveys household travel information is gathered from all members of the household. Sometimes one member of the household is drawn, as a means of cluster sampling within the household.

4. **Double or two-phase sampling**: When advance information on auxiliary variables is lacking, selecting a large preliminary sample is profitable such that necessary population information can be estimated. In the second phase, a stratified random sample can be drawn from the first phase sample.

5. **Choice-based sampling**: Basis for the sampling frame is the choice behavior of a specific group. Low incidence of particular behavior requires a more efficient sampling technique. Getting enough observations for these groups would otherwise be problematic. Recruitment is linked to their choice-behavior, for example regular users of public transport systems are recruited while using the public transport system. This is not a real probability sample and it lacks a weighting frame. As an alternative, double sampling could be used.

6. **Sampling in time and space**: A special feature of travel surveys is that one has to take into consideration two variance components; variance in time and space. Sampling in time and space is specially developed to deal with these two components (8).

**Sampling Design Features in Travel Survey Practice**

Among the agencies surveyed, most travel surveys are *general population* surveys (9 out of 13). In 4 surveys the population in a region is the target population. In 5 countries a population or persons register is used as the *sample frame*, in 3 countries a telephone register is used, 3 countries use a list of addresses and in 2 countries another survey or the census database is used as a sample frame.

In only 4 countries are there *continuous surveys*, carried out every year, in 5 countries travel surveys are carried out regularly but with time intervals of several years
and in 4 countries travel surveys are carried irregularly. In 7 of the investigated travel surveys the household is defined as the sample unit. In some cases statistical estimates are established person-based. In 5 surveys only one person per household was selected. In one survey the target population consisted of household with cars and these households were also selected from the census data. Substitution of sample units was not used at all. In most surveys proxies were allowed, although mostly only for young children (6). In 4 surveys proxies were not allowed at all.

Although it was not possible to collect extensive information on the sample design, it appears that in most surveys a stratified sampling technique was used. According to the Travel Survey Manual (10) sample size of most household travel surveys were reduced using more efficient stratified sampling techniques. Most common stratification of sampling in the travel surveys is based on geographic (region, big cities) and demographic characteristics (age and sex). Over-sampling of regions or big cities exists in 4 countries and in 3 countries, some demographic subgroups are over-represented.

Main reason for over-sampling is an under-representation of these areas in the final sample, due to higher nonresponse rates. The aim mostly is to get higher precision. In one country multi-motorized households (having several cars) are over-represented also for reasons of higher precision. Census data is the basis for such over-sampling. No stratification at all is still common practice in more than one third of the questioned countries. There is no information on the quality of the sample frames used.

Conclusion

It seems that every country has its own thoughts of what might be an optimal or practical sample design. Whether travel behavior needs to be observed as on ongoing process or every 5 or 6 years, whether the household or the person is the observational unit, whether a particular subgroup or regions or cities need to be over-sampled or not, there seems to be no common standard. With respect to quality of the sample frame, there also seems to be no accepted standard.

RESPONSE, NONRESPONSE AND DATA QUALITY

Why is Nonresponse a Problem?

Nonresponse is a severe problem in social statistics and survey research. It makes survey estimates questionable because of the potential and unmeasurable bias. It is a largely unknown factor and it is not known to what extent postsurvey adjustments can deal with this bias. As the nonresponse error is a function of the nonresponse rate and the difference between average scores among the nonrespondents and the respondents, increasing response rates do not always reduce the nonresponse error (1). Improvement of response rates might not be enough. Reducing the nonresponse bias must be the ultimate goal.

It is generally accepted that nonresponse rates are increasing. People are said to become more reluctant to cooperate and be away from home more often. Although there is a lot of literature about the nonresponse problem and nonresponse causes, it seems difficult to establish a clear picture of the current situation.
One might ask: “Is it really important to know what is going on?” or “Why not just improve response rates?” Improving response rates or keeping response on an acceptable level might be very expensive and the question remains to be answered whether there is a nonresponse bias or not. On the other hand if the nonresponse rate is very high one might easily blame the public, stating that the public’s attitude towards survey participation has worsened and accept the situation as it is. With a clear view of what is going on in other regions or countries, one might be encouraged to try to improve response rates. Information on other surveys might help to benchmark or evaluate one’s own survey. It will give ideas on how to improve the design and the fieldwork in order to obtain good response results and lower the respondent burden. To illustrate this we will first describe what is done to get such an overview.

**Description and Explanation of the Nonresponse Problem**

**Description of the Situation**

Nonresponse is a phenomenon as old as survey-research itself. During the last twenty years quite a number of authors have devoted some time and energy to this problem. It seems however that there is a growing concern with respect to this problem during the last five years (11). Norman Bradburn used his first official speech as president of the American Association of Public Opinion Research to ask for more attention to this problem (12). In 1990, an International Workgroup on Household Survey Nonresponse was founded to exchange experiences, expertise and findings. One of the first conclusions of this group was that there was a great need of reliable and empirical overviews of response trends. The workgroup initiated an International Survey on Nonresponse as an attempt to establish these overviews (13, 14).

Steegh made one of the first descriptive overviews of response trend of academic surveys in the USA between 1952 and 1979. She found that ‘refusals’ were increasing (15). Goyder used meta-analyses to evaluate American and Canadian surveys with respect to nonresponse. His conclusions were that nonresponse was increasing for face-to-face surveys, but for mail-surveys nonresponse was stable (16). Groves concluded in his book “Survey Errors and Survey Costs” (1), that for academic surveys and surveys of market research bureaus, nonresponse was increasing, but for gubernemental surveys, response rates were stable. Schnell produced extensive research on the nonresponse problem in Germany (17). He concluded that nonresponse had only increased a few percentages points between 1970 and 1990. More importantly Schnell stated that great differences exists between survey organizations in Germany and these differences are caused for the greatest part by methodological and fieldwork inaccuracies or failures. He also concluded that there are no clear indications that people in Germany are less cooperative nowadays than they were in 1970. He experienced that it was very difficult to find comparable and reliable data.

Nevertheless the need for a reliable overview of response and nonresponse trends still exists, because most data are not comparable. The presented time series seem to reflect ‘apple-and-orange comparisons’ (18). Comparisons are distorted by changes in respondent-definition, mode of data-collection, content and design of the survey, fieldwork procedures and strategies as well as organizational changes of the survey agencies. The already mentioned International Survey was designed to collect data on response,
nonresponse and nonresponse reasons as well as information with respect to sample and survey design, fieldwork and survey organization.

This kind of information is essential to establishing an international standard with respect to the level and quality of response. Such a standard can give customers and survey or research institutes a tool to evaluate the quality of their survey and their data.

The International Survey collected data on response and nonresponse and also information with respect to factors or characteristics of sample and survey design, fieldwork approach and survey organization. This was done for general population surveys, which were carried out regularly by governmental agencies, on comparable topics. Labor Force Surveys and Budget Surveys are such surveys. These surveys are carried out in many countries on a regular basis and have more or less the same objectives.

In this paragraph data will be presented for Labor Force Surveys in order to show that there are relatively big differences in response results between countries and that this can be caused to great extent by design and organizational aspects.

Different response trends and response levels are to be seen in Table 1. In the UK, Ireland, Spain, Canada and the USA response trends seem to be quite stable with some deviations in a single year. In Finland, France and especially in Sweden a trend downwards can be observed. In Belgium the response rate seems to increase since 1993 and in The Netherlands a decreasing response rate from 1988 to 1994 seems to be bent into a stable one.

Besides these differences in trends there are also great differences in response levels. In 6 countries there is a more or less steady response level over 90%. Although the trend in Finland is downwards the level is still more then 90%. Spain shows a response level of about 90%. Two countries (Belgium and the UK) have a response level a few percent over 80 and although in Sweden there is a clear downward trend the level is nonetheless far more then 80%. The response level in The Netherlands is compared with other countries very low (about 60%). However compared with response levels from other research institutes in The Netherlands this level seems to be quite normal for the Dutch situation and the way surveys are carried out in The Netherlands, especially with respect to the fieldwork strategy. With specific fieldwork measures as recalls, re-assignment, use of mixed modes it is also possible to obtain response rates of 75% en even 85% in The Netherlands as is shown by Louwen. For the National Housing Survey, a face-to-face survey that is carried out since 1977 every four years in The Netherlands, it was possible to obtain response rates of about 75%, also in 1993/1994. It is important to know, that in some countries (Belgium, France, Germany) participation in the Labor Force Survey is mandatory. In other countries like Finland and Sweden the Labor Force Survey is a telephone survey and is less burdensome then in other countries where the Labor Force Survey is a face-to-face survey (Netherlands, United Kingdom, USA).

Table 1 in a way gives an example of the data that are available. It also shows that it is quite possible to establish a reliable overview of what is going on. What can be done for Labor Force and other general population surveys, can also be done for Travel Surveys, depending on the abilities of the agency.

With this kind of data it seems to be quite feasible to establish standards. The more differences can be explained by differences in content, design, fieldwork and organization the more it seems to be possible to use international comparable data for bench-marking, improvement of surveys and setting standards. Moreover what Table 1 shows is that it seems to be quite possible to obtain relatively high response results.
<table>
<thead>
<tr>
<th>Year</th>
<th>Austr</th>
<th>Bel</th>
<th>Can</th>
<th>Dan</th>
<th>Finl</th>
<th>Fr</th>
<th>Ger1</th>
<th>Ger2</th>
<th>Hun</th>
<th>Irl</th>
<th>NL</th>
<th>Pol</th>
<th>Slov</th>
<th>Sp</th>
<th>Swe</th>
<th>UK</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>82</td>
<td></td>
<td></td>
<td>96</td>
<td></td>
<td></td>
<td>94</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>96</td>
</tr>
<tr>
<td>1984</td>
<td>83</td>
<td>89</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95</td>
</tr>
<tr>
<td>1985</td>
<td>83</td>
<td>88</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td>96</td>
<td>77</td>
<td>91</td>
<td>93</td>
<td>82</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1986</td>
<td>81</td>
<td>82</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>90</td>
<td>82</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td>78</td>
<td>82</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>90</td>
<td>81</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td>82</td>
<td>81</td>
<td>93</td>
<td>93</td>
<td></td>
<td></td>
<td>96</td>
<td>61</td>
<td>89</td>
<td>83</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td>83</td>
<td>82</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td>96</td>
<td>61</td>
<td>88</td>
<td>84</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>83</td>
<td>82</td>
<td>92</td>
<td>97</td>
<td></td>
<td></td>
<td>96</td>
<td>60</td>
<td>88</td>
<td>90</td>
<td>88</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>96</td>
<td>82</td>
<td>95</td>
<td>84</td>
<td>92</td>
<td>96</td>
<td>96</td>
<td>95</td>
<td>91</td>
<td>89</td>
<td>86</td>
<td>84</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td>96</td>
<td>81</td>
<td>96</td>
<td>82</td>
<td>93</td>
<td>93</td>
<td>97</td>
<td>96</td>
<td>83</td>
<td>96</td>
<td>95</td>
<td>88</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>97</td>
<td>85</td>
<td>-</td>
<td>82</td>
<td>92</td>
<td>92</td>
<td>97</td>
<td>98</td>
<td>83</td>
<td>96</td>
<td>58</td>
<td>94</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>96</td>
<td>86</td>
<td>95</td>
<td>76</td>
<td>93</td>
<td>92</td>
<td>97</td>
<td>98</td>
<td>85</td>
<td>91</td>
<td>59</td>
<td>91</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>96</td>
<td>84</td>
<td>94</td>
<td>74</td>
<td>92</td>
<td>91</td>
<td>97</td>
<td>99</td>
<td>83</td>
<td>-</td>
<td>60</td>
<td>90</td>
<td>87</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>96</td>
<td>-</td>
<td>95</td>
<td>75</td>
<td>91</td>
<td>91</td>
<td>-</td>
<td>81</td>
<td>-</td>
<td>-</td>
<td>59</td>
<td>90</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Austr=Australia; Bel=Belgium; Can=Canada; Dan=Danmark; Finl=Finland; Fr=France; Ger1=former western Germany; Ger2=former eastern Germany; Hun=Hungary; Irl=Ireland; NL=Netherlands; Pol=Poland; Slo=Slovenia; Sp=Spain; Swe=Sweden; UK=United Kingdom; USA=United States of America.
Differences between countries can be explained to a great extent by differences in design, fieldwork approach and organizations (19). There might be also a cultural or ‘climate’ effect, i.e. the situation in one country cannot be compared with the situation in another country, but this only a part of the explanation.

**Conclusion**

Nonresponse is not a disaster caused by nature. It seems to be quite possible to obtain high response rates, because response results are for a great part affected by design, fieldwork and organizational aspects.

**Description and Explanation of Aspects**

To explain what is stated in the conclusion above the most important aspects of design, strategy and organization will be described in short and it will be explained how these aspects might affect response results. Although response rates for Labor Force Surveys may not typify those in Travel Surveys, it will be clear that the same aspects and characteristics may play a role and that it might be possible to improve response results for Travel Surveys also, if Travel Surveys will be designed and organized taking into account the role these aspects and characteristics play with respect to respondent burden and respondent friendliness.

The subject or topic of the survey might play a role. In Ireland for instance ‘employment’ is a very important topic and this seem to motivate people to participate in the Labor Force Survey, which has a very high response rate, although the fieldwork-strategy is not particularly designed to obtain the highest response rates possible. Beside that there are not many surveys in Ireland, so the survey burden is very low, if compared with other countries. ‘Car-owners’ might like surveys on cars more than they like labor force surveys. ‘Travel’ might be a more motivating topic then ‘Employment’. The type of sample frame is important as was already explained earlier. Sometimes a persons-register is used with the advantage that the approach strategy (letter) can be addressed direct and personally to the sampled person. Where a frame of addresses is used one needs to select an observational unit on the spot, which makes it more difficult to find the selected person at home. If the household is to be questioned one needs information on all household members and, if proxies are not allowed, this enlarges the respondent burden and affects response rates negatively. Also the survey method can affect respondent burden and response rates. Panels seem to be more burdensome than cross-sectional surveys as the same respondents have to participate in the panel survey more then once. The mode of data collection and the data collection instruments can play an important role. Mail surveys and telephone surveys seem to be less ‘intrusive’ than face-to-face’ surveys. Especially in mail surveys, complex questionnaires might result in greater reluctance to respond then with simple and open diaries. In Sweden and Finland the Labor Force Surveys are telephone surveys, while in all other countries the face-to-face mode of data-collection is used. This might explain why in Sweden and Finland response rates show a steady trend downwards. It seems to become more difficult to contact people by telephone. Substitution of refusals or ‘not at homes’ is some times used. For the Labor Force Surveys presented in Table 1, substitution is only used in Spain. It makes it difficult to compare response figures. In general this is judged to be an ‘inferior’ survey method as you cannot control the fieldwork...
and the nonresponse. The same counts however for surveys without substitution and a (large) unknown nonresponse bias. Use of proxy respondents is more common for Labor Force Surveys. Only in Australia proxies were not allowed and since 1994 proxies were not allowed in Denmark and as a consequence of this the response rates went down in Denmark.

Although most survey agencies use advance letters, the fieldwork strategy differs very much between countries. Sometimes call scheduling is used, i.e. calls and visits are scheduled to maximize the contact-probability and as a consequence the response probability. For Labor Force Surveys this is the case in the UK, Ireland, USA and Canada. Sometimes agencies allow their interviewers to go or to call as often as they want (high number of contact attempts). Refusal conversion is used by most agencies, because it is known that many refusals are caused by specific circumstances (no time) that are changed when interviewers come back some other time. In the Netherlands ‘refusal conversion’ is not used for the Labor Force Survey because of what might be called the very ‘sensitive’ survey climate. Incentives for respondents are generally only used if the respondent burden is very high as is the case with most expenditure surveys. Interviewer incentives in general means ‘differential payment’ related to the performance of interviewer with respect to response results. Where these kind of incentives are used it is mostly done in combination with performance monitoring system. Sometimes paying per response is seen as a way of motivating interviewers to perform as well as they can.

The more the fieldwork strategy is designed to get high cooperation rates the better the response results will be. An example is the Labor Force Survey in Belgium. The fieldwork strategy was changed drastically in 1993 and from that year on response results higher.

A largely unknown factor is what might be called the survey climate. In The Netherlands for instance the survey burden seems to be very high. Many surveys are carried out and also the growth of call-centers and telemarketing seems to have a negative effect on the survey climate. Also public debates on privacy might deteriorate the survey climate as was the case in The Netherlands (around 1981) and in Sweden (Metropolitan debate 1985). In Ireland on the other hand it is said that the survey climate is very positive. The ‘survey burden’ is low, only a few important are carried out, which leads to the situation where with relatively little effort one can obtain very high response results.

There are many aspects that must be taken into account if one wants to design the survey in such a way that a high response rate can be obtained. The question however is how the different possibilities with respect to design, fieldwork and organization must be used to get the best results. For this it seems to be necessary to understand how and why people react to the survey request.

Theory of Survey Participation

To understand how and why people react as they do to the survey demand the ‘theory of reasoned action’ and the (preliminary) ‘theory of survey participation’ can be of help (20, 21). In short, the theory of reasoned action hypothesizes that the intention to perform a specific behavior is an accurate predictor of behavior if the intention and behavior correspond closely, if the intention does not change in the interval between intention and action, and if the behavior is under the individual’s personal control.
Groves and Cialdini suppose that the most important moment with respect to the decision to participate in a survey or not is the moment of interaction between the intended respondent and the interviewer (or the survey/interview organization in case of mail or self-administered surveys) (21). Mostly people take heuristic decisions when there is a demand for survey participation (21). “They are likely to base their decisions on one or two highly prominent and normally diagnostic considerations (e.g., the length of the survey or the authoritativeness of the interviewer)”, as Groves and Cialdini put it (21). This also means, that the decision is highly dependent of the situation at the moment of interaction and the performance of the interviewer. If people are very busy, for instance furnishing their home, preparing meals or feeding the baby, this is not a good time to ask for survey-participation, because these are ‘rational’ reasons to refuse. Interviewers who understand these ‘signals’ and only ask to make an appointment are in general more successful, they change the situation in the sense, that they make the rational argument of the potential respondent less relevant. This might be the explanation why ‘refusal conversion’ in general is an effective measure to improve response rates.

This seems to be in line with the findings of Braver and Cialdini, that “the tendency to comply with a survey request is not only a durable disposition reflecting the individual’s attitudes, personality traits and values, as compared to a momentary inclination” (22). Also Jansen’s conclusion, that ‘total refusers’and ‘selective refusers’ showed no difference with respect to survey participation, supports the idea that circumstances can affect the decision taking process a lot and that it is not very useful to distinguish between ‘soft’ and ‘hard refusers’ (23). Of course there are certain groups in the population with low response rates (low education, low income, elderly) due to their group characteristic, but even for these groups situational or momentary factors seem to play a role as the experiment of Jansen showed (23).

The two theories seem to come together at the point in the decision taking process of the intended respondent, where reluctant respondents try to find ‘reasonable’ arguments not to participate and the interviewer tries to disarm the ‘reasonability’ of the argument. The very fact, that ‘maintaining interaction’ and ‘refusal conversion’ work can be seen as indications that people in general do not flatly refuse, but always seek to underpin their reluctant behavior with arguments of reason.

Now the question is “what can be learned from these theoretical assumptions?” It seems that the more rational arguments intended respondents have, the more effort is needed from interviewers and the interview organization to convince people to comply to the survey request. ‘Rational’ arguments can be found in the situation or circumstances - but situations can change and refusal conversion at different times is therefore very useful - and in the ‘respondent burden’ and ‘survey burden’ or ‘survey climate’. The respondent burden is the effort that is needed from the respondent to participate in a particular survey. The respondent burden is highly determined by the sample and survey design. The survey burden or survey climate can be seen as the total burden caused by survey organizations and others in a particular society. If people are often asked to participate in surveys or to answer ‘direct marketing questions’ they might get the idea that survey organizations do not care about them and the burden they put on the population. This, and of course the burden of a particular survey request itself, might give people more or less rational arguments to say ‘no’.

On the other hand, if the survey request hardly causes any burden (e.g. a very short telephone survey) or is not intrusive (mail survey) people tend to say ‘yes’, because in
general it seems that people do not like to say ‘no’ to an other person [“helping tendencies”, (21)]. To say ‘no’ one needs arguments to ‘rationalize’ this decision, because in general it is not common or social behaviour to refuse a (survey) request without any reason. In the case of a mail survey, the survey request must be made personal to make full use of these theoretical assumptions. The burden of a particular survey and the (perceived) survey burden in general might provide these arguments to intended respondents. The burden is an important factor, as already is shown by Bradburn and Lyberg & Dean (24, 25). So for the survey design stage the strategy must be: minimize the respondent burden to make it easy for respondents to say ‘yes’. If this is not possible, the strategy for the fieldwork must be: make it difficult for potential respondents to say ‘no’.

Nonresponse in Travel Surveys

In order to be able to evaluate Travel Surveys in relation to these theoretical assumptions and interpretations and also the experiences with respect to Labor Force Surveys, we first will describe how the travel surveys in the 13 countries are designed and organized as far as this can be related to the nonresponse problem.

Survey and Sample Characteristics and Respondent Burden

In the section on sampling design features in travel survey practice, the sample characteristics of the travel survey on which information is gathered are described. It appears that in 50% of the cases data are collected on the household level and that proxies are allowed for young children only. If the household is the objective of study and proxies are not allowed, the respondent burden will be substantial; because this will require that someone in the household must ensure also the completion of questionnaires and diaries of all other household members. If one person per household is selected or directly receives the survey request - population registers offer this opportunity - the survey burden is less, because this person only has to bother about his own questionnaires and diaries. So in the case of households being the observational unit, this might lead to more drop outs and more item or person-nonresponse.

All travel surveys investigated are voluntary surveys which means that respondents can freely decide to participate or not. All travel surveys are also cross-sectional surveys which are less burdensome if compared with panel-surveys. For panel-surveys, not only does panel-attrition seem to be a problem, but also the initial response. It seems to be much more difficult to convince people to participate in panel-surveys, as they are asked to participate for a longer period in time with repeated interviews.

Other important aspects of the survey design that have an effect on the respondent burden are the mode and technique of data collection and the observational period. Most travel surveys (8 out of 13) use diaries in combination with questionnaires. A diary has the advantage that the respondent can complete it on his own, without the presence of an interviewer looking over his shoulder, but it can be quite burdensome as the respondent has to think of reporting his behaviour for the entire day or period. Questionnaires asking for travel data retrospectively have an advantage, that the respondent only has to react, but if the observational period is too long respondents will find it difficult to dig in their memory, which can be unpleasant. We will not go into the possibilities of measurement errors associated with different measurement instruments. In relation to the nonresponse
problem the respondent burden seems to be the most important aspect of the survey design. And for this also the mode of data-collection is important.

Half of the investigated travel surveys are telephone surveys. The diaries are then mostly sent by mail. Only two of the surveys are mail surveys and four use face-to-face or home-interviews. One might say that home-interviews or -visits cause more respondent burden, because people have to invite another strange person into their home, which might be considered as ‘intrusive’. For some people however a home-visit might be a welcome change in their lonesome existence. Mail surveys seem to cause the least respondent burden as there is no personal contact and the respondent can choose when the questionnaires are being completed and sent back. Mail surveys without any contact between respondents and survey agency however, have very low response rates and it is necessary to have at least one call or reminder to improve response. If the mail survey includes one or more telephone calls to remind or motivate respondents the difference between mail and telephone survey becomes more and more unclear, certainly when telephone surveys also rely on mail. However there might be a quite a substantial difference. In a telephone interview respondents are asked to respond to questions, if they like that or not, they have to provide answers and they might not be in the mood. If a mail survey with telephone calls to remind or motivate respondents is well designed with regard to survey burden and how respondents might perceive this, it might have important advantages compared with telephone and face-to-face surveys.

A special important feature is the observational period, i.e. the length of the time on which respondents have to report their behaviour. In half of the investigated travel surveys the observational period is relatively short, about one day. The other half of the surveys seem to pose quite a burden on the respondents. One of these surveys has a observational period of two days, which seems reasonable, but others have observational periods varying from one week to one month for ‘fuel consumption’ or reporting of ‘long distance trips’ and 3 months for long-distance trips.

Fieldwork Strategy

The more the survey poses a burden on the respondent the more effort is needed from the fieldwork-organization to convince reluctant respondents. Organizations that do not use special instruments or strategies might face higher nonresponse rates than is acceptable. The most important characteristics of fieldwork strategies that aim at the highest possible response results are:

- special approach strategies to contact the respondent and to explain that participation in the survey is more or less a logical step; elements of these strategies are: advance letters & leaflets, call scheduling and call backs,
- strategy to convince reluctant respondents and convert refusers,
- use of respondent incentives,
- monitoring and controlling the fieldwork very closely,
- interviewer performance rewards, sanctions on bad performances and differential payment

Most of the travel surveys on which information is gathered use one or more of these instruments. Eight surveys carry out a special approach strategy using call scheduling and
telephone calls to motivate respondents. Although for every survey advance letters are used, for four surveys there is no special approach strategy. Only 3 surveys use respondent incentives of which two have quite a large respondent burden, i.e. an observational period of 7 or 13 days. For most of the investigated surveys monitoring and controlling the fieldwork is part of the strategy. For only 2 surveys does there appear to be no clear policy on this point.

One might say that paying per response or interview is as such an incentive for interviewers to perform as best as they can. However it might also lead to a particular behavior of interviewers to select ‘good addresses’ and pay less attention to reluctant respondents. To convince these respondents or to convert refusals interviewers need more time. For this a hourly payment seems to be the best solution. Most of the survey organizations from which data are obtained use some form of rewarding interviewer performances. This counts for 10 out of 13 organizations. Two of them use negative instruments, i.e. sanctions for bad performance. Only three organizations use no rewards at all. Hourly payment or paying per interview is not adopted when bad or good interviewer performances are rewarded. Six of the 10 organizations that make use, rewarding performances pay by the hour and 4 pay per interview.

Survey Climate

The survey climate can be an important factor. As explanation for survey climate has already been provided in the section on description and explanation of aspects. A survey organization can make use of publicity campaigns to try to influence the survey climate. For none of the investigated travel surveys were publicity campaigns used. This seems to be quite logical, because 9 out of 13 organization report that the survey climate in their country is positive or even very positive.

Response and Nonresponse Data of Travel Surveys

In Table 2, response data of the 13 investigated travel surveys are presented. As is already pointed out in the section on description and explanation of the nonresponse problem, it is not easy to gather reliable and comparable data on response and nonresponse. The table shows quite a divers picture, not only with respect to the obtained response results, but also with respect to the quality of the information provided. It seems that not all countries were able to make a differentiation with respect to the most common nonresponse reasons. Not all response and nonresponse data add up to 100%. The categories in the table were meant to be mutually exclusive. It is however not certain that all countries treated ‘partial response’ in that way. If there was more time and means to carry out an extensive investigation the picture might have been better as far as the completeness of the information is concerned. But what about the differences with respect to the obtained response rates?

If we take “full response” as the intended goal with respect to data quality then for 7 travel surveys the response rate is below 70%, 6 even have response rate just over or below 50%. If full response and partial response are taken together the picture seems to be a little bit better: 8 surveys have a response rate of more than 70% and ‘only’ 5 have a response rate under 70%, of which 3 have a response rate below 60%. Still there are big differences. It appears to be very difficult to draw any conclusion on what response rate can be seen as a standard quality indicator.
TABLE 2  Response Data for Travel Surveys (% of fieldwork sample)

<table>
<thead>
<tr>
<th>Definition</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1= full response</td>
<td>80</td>
<td>27</td>
<td>41</td>
<td>81</td>
<td>68</td>
<td>75</td>
<td>44</td>
<td>51</td>
<td>13</td>
<td>91</td>
<td>39</td>
<td>78</td>
<td>74</td>
</tr>
<tr>
<td>2= partial response</td>
<td>-</td>
<td>38</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>14</td>
<td>20</td>
<td>61</td>
<td>-</td>
<td>0</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>3= refusals</td>
<td>-</td>
<td>13</td>
<td>-</td>
<td>18</td>
<td>22</td>
<td>-</td>
<td>14</td>
<td>19</td>
<td>2</td>
<td>3</td>
<td>34</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>4= non-contacts</td>
<td>-</td>
<td>22</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>3</td>
<td>15</td>
<td>4</td>
<td>34</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>5= other nonresponse</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>22</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>-</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>6= drop outs</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>16</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>20</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Response Rates, Fieldwork Strategies, and Survey Characteristics

In Table 3 an overall evaluation is presented of the 13 investigated travel surveys as far as respondent burden, fieldwork strategy, survey climate and obtained response rates is concerned. The respondent burden is evaluated positive, if the burden is low (maximum 1 day reporting period, telephone or mail survey). The fieldwork strategy is evaluated as positive, i.e. designed to obtain the highest response results as possible, if call scheduling is used or calls to remind or motivate respondents, the fieldwork is monitored and controlled and interviewer performances are rewarded or if at least two of these three groups of instruments are being used. The survey climate in a particular country is judged to be (very) positive according to the answers given to the very question. Response rates have a “+” sign if the response rate (full response and partial response is 70% at least).

TABLE 3  Evaluation of Travel Surveys with Respect to Burden, Fieldwork, Climate, and Obtained Response Rate

<table>
<thead>
<tr>
<th>Aspects</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burden (low = +)</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Field. Strategy</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Climate</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Response rate</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Of course, the information presented in Table 3 is only descriptive and roughly combined. The most encouraging pattern of associations for researchers would be between all positive indicators, i.e. a low burden, an active field strategy, a positive climate and then a high response rate (countries 1, 6, 10 and 12). This does not mean that every active fieldwork strategy is good. If the respondent burden is low, the survey climate is positive, it seems to be better not to use a too aggressive fieldwork strategy, because this might affect the survey climate negatively in the long run. In country number 8 with a bad survey climate, strategies seem to be used as needed and consequently a positive response result is obtained. In countries 4, 9 and 13 the respondent burden appears to be high, but with an active fieldwork strategy in a positive survey climate these countries are able to get high (> 70%) response rates.
In country 2 with more or less the same situation (also a high respondent burden, the fieldwork strategy is active and the survey climate is positive) the response rate is relatively low. This might have been caused by the fact that the sampled households already participated in another survey. In countries number 3 and 7, agents seemed to think that survey participation is something that you cannot influence or were not aware of the possibilities to improve response rates. In both countries, the survey climate is judged to be not very good and still their fieldwork strategy is not designed to get the highest possible response results. In country number 5, the respondent burden is low, they use an active fieldwork strategy, the survey climate is positive, but still the response rate is relatively low, although 68% response rate might be a good result in that particular country. In country number 11, one seems not to be able to perform as is needed given the bad survey climate. Maybe the fieldwork strategy is not as active as it should be, although is has two aspects of what seem to be important.

The aim of this paragraph was not so much to show that differences in response rates can be explained by aspects as respondent burden, fieldwork strategy and survey climate, but more that it is necessary to think of these aspects as an integrated whole at the design stage of the survey. Most of the time it is the budget that puts limits on the design of the survey and not the survey researcher. It is also very important to think of the respondent burden. In general respondents seem to be quite willing to cooperate if the respondents task is not to difficult, does not costs too much time and if respondents are not pushed. Respondents which are pushed to cooperate might give wrong answers (measurement error). So, besides budget pressure, the researchers need for data, also the respondent burden should be a main design factor as this might affect the appearance of measurement errors and the survey climate in the long term. The most important aspect with respect to response and nonresponse rate is the possibility of nonresponse bias. If one uses a very active fieldwork strategy to get the highest possible response results and does not have the slightest idea if there might be (some) nonresponse bias, it is not sure whether the invested money pays off.

To some extent the picture as presented in Table 3 looks better than it is. If only ‘full response’ was taken as indication for response results, instead of 5 countries 7 countries would have bad response results. And of course if the discussion is about standards and raising the standards the definition of response should be ‘full response’, i.e. all response records that are available and useable for analyses.

**Conclusion**

It seems that there is a lot of work to be done with respect to travel surveys to develop an optimal combination of design and fieldwork strategy to lower respondent burden and to obtain good response results. There is not yet an accepted standard with respect to a minimum response rate nor is there a accepted definition of response. Standards for response results should be combined with indicators for nonresponse bias.

**WEIGHTING TO ADJUST FOR SAMPLING AND NON-SAMPLING ERRORS**

**Use of Weighting Procedures in General**

Weighting as a method for post-survey adjustment can be used for different reasons. If for example a stratified sample design is used, which means for example that some groups are
over- or underrepresented, the survey outcomes need to be corrected before they can serve as reliable population estimates. Distributions in the sample must be adjusted to distributions in the population.

Weighting also can be used to correct for unequal inclusion probabilities. If, for instance a person-based sample is used and the survey also must provide estimates on households, the weighting procedure must correct for the fact that one-person households had lower inclusion probabilities than more-person households.

If, in the case of travel, time-budget or expenditure surveys, the sampling procedure emphasizes weekdays or weekends, the weighting scheme must correct for this, especially if behavior is different during weekdays or weekends.

Weighting is not only used to improve precision, in other words to reduce the sampling error, it is also used as an instrument to reduce non-sampling errors as bias due to non-coverage and nonresponse.

If not all the population variables are known, more complex weighting techniques are used (linear or multiplicative weighting). Research on weighting techniques suggests, that there is not much difference with respect to the estimates (26).

If the sample frame does not cover the whole target population, weighting can be used to adjust the survey outcomes for this bias. The same holds for nonresponse. Non-coverage and nonresponse are to some extent comparable problems. Non-coverage means that a particular group of the population is not included in the sample frame, so it was not possible to collect data on this group. (Unit)-nonresponse means this non-responding group is also not represented in the sample outcomes.

Non-coverage and nonresponse are severe problems if these groups have different characteristics with respect to the dependent or target variables in the survey. If there is no difference there is no problem. There is, however, mostly no information on these groups with respect to the target variables. Weighting to reduce bias can be used if the auxiliary variables in the weighting scheme are strongly correlated to the target variables. If for instance nonresponse is high for car users, commuting every day from suburb to the city, correction with auxiliary variables as ‘car ownership’, ‘having work’, ‘urbanicity’, ‘age’ and ‘marital status’ might reduce such nonresponse bias. But one never knows to what extent. The assumption in the weighting scheme is that the nonresponse group has (more or less) the same ‘car-use behavior’ as the respondents with the same characteristics. The problem however is that there is no information at all most of the time. So, in the example described, it is not known that the group of car users, commuting every day from suburb to the city, have higher nonresponse rates. Besides that, if there is information on nonrespondents, it is not known if the correlation between auxiliary and dependent variables for the particular nonresponse and response groups are strong enough. It is also important where the auxiliary information on variables and their distributions came from, e.g. the same survey or from other, more reliable sources.

Armoogum et al. reports several examples of empirical evidence in travel surveys with respect to nonresponse bias. Nonresponse correlates with (27):

- low income
- low education
- non-car & high car ownership
- household size (one-person & four or more person households)
- unemployment
• living area in metropolitan areas
• low and high mobility.

There is also some empirical evidence in some countries that there is more fieldwork effort needed to contact highly mobile persons, which can also be seen as an indication for nonresponse bias.

If there is reliable information with respect to these variables and their distribution in the population, it is quite possible to reweigh sample outcomes and consequently bias will be reduced. Most of the time, however, there is no reliable information from other sources with respect to the dependent variables, which is logical, because this is the very aim of the travel survey itself.

So the question remains what can be done to improve weighting schemes and reduce bias.

**Improvement of Weighting Schemes and Reduction of Bias**

The following methods of improvement can be mentioned:

• a modular and integrated survey design,
• use of process information provided by reluctant or hard-to-get respondents as proxies for final refusals or non-contacts,
• use of special and in the field registered information that is related to the target variables provided by reluctant respondents and refusals,
• special research under nonrespondents with respect to the target variables,
• collection of information that highly correlates with the dependent variables during the very interaction with the nonrespondent or at the nonrespondent’s home,

In a modular and integrated survey design - omnibus-surveys are to some extent comparable - there is a basic questionnaire with questions on background variables and on key elements of all target variables (health, employment, housing, crime, travel). If it is a really modular and integrated survey design every respondent has to provide information on the basic questionnaire, but only parts of the sample or selected subgroups have to complete the modules on specific subjects. The advantages are:

• the nonresponse on the basic questionnaire and the specific modules is not related to the subject, as it is for substantive surveys on only one specific subject, and might therefore be less selective,
• the nonresponse for the basic questionnaire can be lower than the nonresponse for the substantive surveys on only one specific subject,
• the nonresponse on the modules can be corrected using information with respect to the target variables from the basic questionnaire.

But still there will be nonresponse on which there’s no information at all.

A common approach in the literature to exploring the potential effect of nonresponse on survey variables is to use reluctant or hard to get respondents as proxies for final refusals or non-contacts (28).

A way of reducing nonresponse bias that goes to the very heart of the bias is what Couper did with his experiment with respect to the National Election Study in the USA. He
asked the interviewers to register exactly what arguments respondents and nonrespondents mentioned (e.g. “not interested in politics”) as a reaction to the survey request (28). This information then was used in the weighting scheme. Couper showed that it is not sufficient to show that reluctant respondents share similar characteristics as others. It is important to also attempt to measure correlates of the measures to convert reluctant respondents with the depending variables. So not only variables that can be derived from the sample frame or other auxiliary data (region, urbanicity, gender) or process variables (temporary refusals, number of calls) can be used in weighting models, but also information on the reluctancy seems to be important especially when potential respondents attempt to disqualify themselves from participation on the basis of the survey topic (e.g. ‘lack of interest in politics’, ‘lack of strong opinions in attitudes surveys’, ‘age or ill-health for surveys of the elderly’, ‘good or bad health for health surveys’, ‘retired or unemployed for labor force surveys’, ‘not enough money for expenditure surveys’, no car for travel surveys’, etc).

Special research of nonrespondents can be carried out, but is in general very expensive. One needs special approach strategies and most of the time special interviewers. The aim should not be to convert as many nonrespondents as possible, because the remaining group of nonrespondents might be the most selective one. The aim must be to get reliable information on all nonrespondents or groups of nonrespondents. The problem is that it is difficult to distinguish relevant groups among the nonrespondents.

A way to do this is to collect auxiliary information at the nonrespondents address (type of dwelling, presence of car, information provided spontaneously by neighbors ‘people living here are at work’) and trying to get some key-information with respect to the target variables (29). Lynn (30) made an attempt of this kind of research of nonrespondents with respect to the British Crime Survey. His findings are not published up to now, but seem to be very interesting.

**How is Weighting Used in Travel Surveys?**

In the questionnaire we sent out to a number of agencies we also requested information on the use of adjustment procedures and the reduction of bias. Only rough information could be gathered, but nevertheless it might give an impression on how the problem with respect to bias is handled.

Almost all of the agencies indicate that there is some evidence for the existence of nonresponse bias. The main variables or characteristics that are under-represented in the survey are:

- very high and very low income groups,
- high and low mileage drivers,
- young single males and females,
- zero vehicle use,
- older and younger persons,
- persons living in metropolitan areas.

In all the investigated surveys weighting techniques are used to reduce nonresponse bias. Sometimes explicitly as an objective of weighting, sometimes as side-activity assuming that the weighting also will correct for nonresponse bias. On the household level, auxiliary
information is used on regional characteristics and household size. On the person level, age and sex are the variables that are mostly used. Only in two of the surveys is information used that is highly correlated with travel behavior (availability of car). Weighting variables used are mostly general background variables, of which one does not know the correlation with the dependent variables for the nonrespondents.

Most survey organizations do not really know to what extent nonresponse bias is reduced by post-survey adjustments. The only thing they seem to rely on are the assumptions with respect to the correlation between the auxiliary variables and the depending variables. Only one agency reports that they carry out specific nonresponse research.

**Conclusion**

Although there seems to be indication for the existence of nonresponse bias and it is not really known to what extent nonresponse-bias is corrected by weighting techniques, there is not much activity with respect to nonresponse research.

**WHAT NEEDS TO BE DONE?**

When discussing the quality standards of travel surveys, it seems quite logical to pay a lot of attention to the nonresponse problem and especially to nonresponse bias. Only improving response might not pay off, because it does not necessarily reduce nonresponse bias. Weighting with general population characteristics might not be enough, because the correlation between auxiliary variables and dependent variables might be weak. There seems to be evidence that nonresponse does not necessarily causes bias, which means that there is not a general attitude in favor of, or against, survey participation.

Jansen showed in an experiment that so called ‘hard-refusals’ that refused to a prior survey request responded as well as the ‘temporary refusals’ or even better to a subsequent survey request (23). These were unexpected findings. We all know that ‘refusal conversion’ does pay off, which can also be interpreted as evidence in this respect. Also the findings of Foster seem to point in this direction. She corrected survey outcomes with census data from respondents and non-respondents (31). She showed that nonresponse bias did not increase strongly in line with increasing nonresponse rates. Her findings also suggest that characteristics of the survey (e.g. topic or design features) have more impact on bias than the overall nonresponse rate. For some subgroups she found relatively high correction factors especially in the expenditure survey and food survey. These surveys are carried out with diaries and have a high respondent burden. Higher correction factors were needed for groups for which one might suppose that participation in that type of survey is very difficult (elderly, ethnic minorities, etc.). So, there seems to be some evidence, that in general for “average surveys” the nonresponse bias is small. Nonresponse bias will increase if the survey has some specific characteristics that correlate with characteristics of groups of the population. Expenditure surveys and other surveys that make use of diaries for a longer period are very difficult for people that are not used to the kind of administrative behavior that is expected. In a telephone travel survey it is very difficult to contact people that are very mobile. This means that the survey design itself can cause nonresponse bias if one does not pay enough attention to the possibility, that some groups will be ‘excluded’ by the way the survey is designed or carried out.
CONCLUSION

So the general conclusion is that it seems not to be very useful to try to get the highest possible response rates by all available means. It seems to be better to try to avoid design or fieldwork failures that might result in decreasing participation rates of specific subgroups and to detect particular groups among the nonrespondents and to collect from them information related to the target variables to find out whether there is nonresponse bias or not.

REFERENCES


The workshop focused its attention on household surveys, in view of the fact that time limitations would not allow for a productive discussion of all survey types. The workshop members agreed it would be best to discuss the three issues in the workshop title in a different order:

- sampling
- nonresponse
- weighting

**SAMPLING**

**Sampling Frames: Quality Indicator and Guidelines**

It was appreciated that availability and quality of frames would vary from country to country and were, of course, dictated by how the target population was defined. This is true even within the scope of household surveys, when the target population is car owners, cyclists, regular bus users, etc.

Despite these differences, it was agreed that, in any case, availability and quality of sampling frames should be documented. When a sampling frame was known to cover or suspected of covering less than 90 percent of the target population, this should be brought to the attention of the client, along with the recommendation that more budget should be made available in order to have the means for covering at least 90 percent. A great number of the workshop members voted for a coverage of 95 percent and more.

**Sample Size: Quality Indicators and Guidelines**

Whilst the percentage coverage of the target population (or an estimate thereof) should be documented, this should not be used as a quality indicator. Assuming we are able to draw a non-biased sample (nonresponse, etc., are discussed later), the sample site should be calculated from

- the known (or estimated) variance of key variables on the questionnaire
- the degree of statistical, spatial and temporal resolution needed (by the user)

Issues relating to the likely effect of the sample size on confidence intervals and allowable random errors should be clearly explained to the client, and documented in reporting. Instead of complaining to the client that the budget did not allow for a large enough sample, it would be better to talk through sampling issues in an attempt to arrive at the best methodology possible within budgetary constraints, and to mutually agree whether it is still worthwhile to move forward with the proposal survey, or if it is necessary to recommend a different approach [e.g., a more qualitative study/(further) desk research into secondary data].
Stratification: Quality Indicators and Guidelines

Designing a stratification scheme that was successful in optimising the sampling procedure with respect to many variables at once would be very difficult and mostly impossible. Often, secondary data on the variables being measured by the survey instrument were not available (particularly with respect to attitudinal questions).

In cases where it was not possible to confidently demonstrate to the client that an increase in sampling efficiency was likely to result from stratification based on data relating to dependent variables such as age and sex, it would be good practice to argue the case for some pretesting of key variable variances. A well-documented, robust case for stratification was a good quality indicator, since it demonstrated considered use of budget and sampling effort.

Unit Selection

Unit selection rules should be specified in detail and strictly be followed. When non-random selection procedures are used—as in supplementing a sample for rare values—these exceptions should be explicitly reported.

In stratified sample frames, usually a proportionate selection of sample units is recommended, in view of the fact that several target variables have to be investigated in household surveys.

NONRESPONSE

Terminology

Different definitions were used in the literature with respect to response rates. A workable definition of response rates was felt to be very useful by the group. The following process schematic was agreed on (see Figure 1).

- Corrections for under-/overcoverage include such cases where it is known/found that two or more “households” live at the same, single address, as it appears in the sampling frame, and where students are double-counted, for example, at university/college halls of residence and at their parents’ house. (This issue is covered later in sample weighting.)
- Administrative errors would include instances where one or more members of a household included in the target population have died/moved/been taken into terminal care, etc.
- Instances where questionnaires returned have clearly been intentionally abused (e.g., the respondent writes his/her name as “Mickey Mouse,” with an income of $10 billion) are effectively counted as a refusal (although this must be decided even after a questionnaire is returned).
- For mail-back surveys, some nonresponse will be caused by “apathy”, i.e., respondents intended to complete the questionnaire, but did not get around to it (in time).
- There may be a point at which an unacceptable item “nonresponse” (as determined by the client/researcher) effectively counts as a unit “nonresponse”, since the entire record must be discounted.
Quality Indicators and Guidelines

In addition to a detailed documentation of the full methodology to help the client understand the context of the nonresponse (e.g., reporting the number of household callbacks, including a copy of the mail-back cover letter), it was agreed that the four types of nonresponse identified should be quantified for the client. Comparing the net (achieved) sample with the fieldwork sample characteristics offers some limited help in understanding nonresponse. Nonresponse is a problem particularly if the nonrespondents have different characteristics with respect to the dependent or target variables, as compared with respondents—i.e., nonrespondents do not represent a random sample of respondents.

Unfortunately, there is usually no information on nonrespondents with respect to many target variables, such that a priori weighting can be used only if the auxiliary variables in the weighting scheme are strongly correlated to the target variables. For example, some nonresponse will be due to “lazy” people. Correcting of nonresponse in “young males”, for example, would be of limited (but not zero) value in the absence of a correlation between “young males” and “laziness” (the cause of the nonresponse). It was agreed that some budget should always be set aside for recontacting nonrespondents in an attempt to make at least qualitative statements about any nonresponse bias.
Guidelines for Minimizing the Effect of Nonresponse

- Ensure that the respondent burden is kept to a minimum, for example, through the use of cognitive laboratories to help in the thorough pretesting of the survey instrument.
- Incentives may be used, but should be treated with caution in case respondent bias is increased.
- Fully document all suspected reasons of nonresponse, share this experience with both the client and the research community, and equally document corrective procedures.
- Attempt to convert refusals through follow-up/repeat contacts/reminders. Many refusals are “circumstantial” (not “core”), and thus may be converted.
- Consider scope for carrying out specific research to understand the reasons for refusal, as opposed to an objective of correcting for it, in order to help future research (e.g., by weighting the sample to increase the selection probability of anticipated nonresponse.)
- The response rate should not be used alone as a quality indicator. It may be far better to weight up a “good quality”—50 percent—sample rather than push the response rate up to 70 percent with “poor quality” data, and thus make the overall (weighted) quality even worse. Respondent pushing can also harm the long-term survey climate.
- Assure the respondent that the data will be used!
- If target variables continue to change as a function of the number of recontacts (or with time/reminders for a mail-back survey), there will be all the more reason for trying to understand nonresponse and continuing to pursue this until the effect ceases, or is understood well enough to make an attempt at sensible controls/corrections.
- For panel surveys, in particular, a full description to the respondent of exactly what is involved (and over what time scale) is likely to increase initial nonresponse (refusal), but to have lower attrition rates, and thus give the researcher a better foundation for planning sampling requirements.

WEIGHTING

Schematic Description of Weighting

The weighting procedure usually should be

- First stage: sample weighting - Sample weighting corrects for unequal probabilities of selection. That could be correct ign frame errors and/or requiring disproportionate samples. This stage should be compulsory for disproportionate samples.
- Middle stage: unit/item weighting (nonresponse weighting) - This stage may weight respondents up/down due to unit nonresponse; it may correct for item nonresponse by weighting or by the imputation of missing values.
- Final stage: grossing up

After the corrections above, overall sample characteristics may (still) not reflect the population proportions on key variables, for example, because the random sample did not give a representative sample. This stage is often done primarily for political/cosmetic reasons.
The order of the second two stages may be inverted. Whether the weighting takes place by households first, then by people, or vice versa, depends on objectives, but will rarely give the same result. We can overcome this by

1. iterative weighting  
2. weighting to cells, if population cross-tests are known

**WEIGHTING: GUIDELINES**

- Document the weighting procedure fully. Add weighting variables to the data sent in a manner that allows them to be easily identified. Keep them there for other users. Identify the effects the weighting has for the variance of dependent variables, and how this has been taken into account (e.g., with significance testing).
- Keep the weighting as straightforward as possible, for example, avoid the adoption of different weighting procedures in different sections of a report—this makes it very difficult for the client to follow.
- Weighting non-contacts (e.g., despite callbacks) is probably better used for correcting (based on data for late/difficult contacts) rather than refusals. There is little literature evidence to suggest that late/difficult contacts are similar to refusals.

**Weighting: Quality Guidelines**

The difference between the weighted and non-weighted data sets is not a good quality indicator, because:

- the sampling could have been disproportional to allow for specific weighting of particular subgroups/cells;  
- this may encourage political/cosmetic weighting; one should be free to weight to the optimal sample characteristics, as far as this can be statistically justified.

**Underreporting: Quality Control**

- The use of secondary counts/volumetric data just to check diary data is reasonable. Often inadvisable is to rely on secondary data for corrections/weighting to diary data.
- Undertake validation surveys to double-check the quality of the collected data and to understand (at least qualitatively) any shortcomings in the survey instrument in eliciting full trip reporting (e.g., questionnaire design errors).
- Surveys were felt to be a good mechanism for assessing the quality of diary data.

**ACKNOWLEDGMENTS**

The initial draft of this workshop report was made by Andrew Cook. The workshop chairman got very helpful comments from Chester Wilmot and Carlos Acre. The chairman feels very obliged to all workshop members for the engaged and constructive discussion.