23rd Australasian Transport Research Forum
Perth, Western Australia. 29 September – 1 October 1999

Systematic Errors in Mobility Surveys

Werner Brög and Erhard Erl
Socialdata GmbH, Germany

Abstract

Mobility surveys play an important role in transport planning because they often provide input for decisions resulting in an investment of millions of dollars.

Although such surveys are common practice since many decades, there is little concern about the quality of these surveys and consequently the validity of results achieved.

But even if data quality is in question, the discussions are often centered around random errors, significance and representiveness.

However, detailed research into quality of mobility surveys shows, that systematic errors, related to design and conduct of such surveys do have often much greater effects. Since systematic errors can be avoided, it is very important to provide better knowledge about effects of survey design elements which spoil the quality of mobility data.

The paper will present examples of such design elements which did have great effects on results and could have been easily avoided.

Contact Author

Werner Brög
Hans-Grässel-Weg 1
81375 Munchen
Germany

Phone: +49 89 7108-1 Fax: +49 89 716420 e-mail: socialdata@socialdata.de

1. Introduction

In transport research, statistical methods and modelling techniques are well developed and have reached a high level of expertise. The input for such analysis and models are very often empirical data, gathered in surveys on travel behaviour.

Correspondingly, there is much concern about random errors in the input data, e.g. significance or representativeness. On the other hand an empirical data base can contain systematic errors, which cannot be analysed nor described by statistical methods. Evermore the statisticians, transport planners and modellers are often not aware that this type of error occurs.

This paper concerns systematic errors and gives real-life examples of the impact of survey instruments and design factors on the validity of the results of studies on travel behaviour and the derived conclusions for planning decisions. The cases presented here are by far not exhaustive of the whole field of possible sources of error, but highlights the importance of such considerations.

Discussions about quality of mobility surveys are traditionally dominated by extensive concern with random errors, which suggest to the less experienced data user that statistically satisfactory random errors can be equated with the "reliability" of the results. This simplification is just as ill-considered as it is harmful. Particularly when measuring complex phenomena (like mobility behaviour), the systematic errors caused by the use of an inappropriate measurement method in complex survey designs and a questionable state of practice can exert a disproportionately larger influence on the results than that due to random errors. Systematic errors can in principle be anticipated and - up to a certain point – corrected. Failing to take these errors into account and the concentration on random errors can only be explained by the fact that as a rule, good statistical knowledge but little real experience with empirical methods are demanded in "traditional" transport research.

Problems found are relatively easily explained, if one is familiar with the respective survey designs, and understands the systematic errors arising from using them. However, a discussion of these effects - as important and useful as it would be for the further development of methodology - does not normally take place.

2. Systematic Errors in Survey Design

A Case study

For some years now, the **Dutch National Travel Survey (OVG)** which is conducted by the **Central Bureau of Statistics, Netherlands (CBS),** has had a significant decline in response. The OVG has a telephone-postal design. From 1978 until 1984 the data was collected in a face-to-face interview but due to budgetary reasons was collected by a telephone / mail interview from 1985 onwards. The survey population is defined as the resident population of the Netherlands The sampling unit is the household.

The decline of overall response rates stems partly from an increasing proportion of unlisted telephone numbers and therefore a declining accessibility of households. Furthermore, there is a tendency of a reduced willingness to participate These factors combined resulted in a drop of overall response rates from 51 % in 1985 to 35 % in 1998 (Moritz and Brög 1999).

The declining accessibility and rapidly dropping response rates gave rise to serious and increasing doubts with respect to representativeness of the sample and comparability of survey results. Parallel to these developments the demand from the side of policy makers for information with respect to transport and mobility increased These factors caused Statistics Netherlands, in co-operation with the Ministry of Transport and Public Works to look for an alternative design that would combine significantly improved response results with enhanced research flexibility.

Preliminary investigations for a new design of the Dutch National Travel Survey resulted in a choice for the **NEW KONTIV DESIGN (NKD),** developed by SOCIALDATA in Munich, Germany. (The NKD employs a mail-back technique.) Statistics Netherlands tested this design in a pilot study in September 1997.

In the pilot study, a response rate of 74 % was obtained from households that were approached using the NKD procedure (table 1). This result was significantly better than the response rate in the control group using the OVG procedure, achieving a response rate of 44 % of the households. Unlike the procedure in the OVG-group, the households in the NKD-group, for which no telephone number could be obtained, were approached by mail. For these households a response rate of 45 % was obtained. The higher response rate in the NKD-group, however, can only be partially explained by this difference. The response rate for the households for which telephone numbers were known was considerably higher in the NKD-group: 81 % versus 55 % in the OVG group.

Table 1: Response Rates

	OVG*	NKD		
		total	tel. (+)	tel. (-)
Net sample (households)	1014	960	765	195
Responding households	446	708	620	88
Response rates	44 %	74%	81 %	45 %

* OVG-control-group: same period, region and sample frame.

The NKD pilot survey resulted in 1.72 activities and 3.03 trips per person per day (table 2).

The OVG control group showed 2.32 activities and 3.88 trips per person per day. The difference in these mobility figures between NKD and OVG led to the suspicion, that the NKD would achieve higher response rates, but the OVG more reliable results, because the latter measured a mobility level (trips per person), which was 28 % higher.

Table 2: Mobility figures: comparison NKD - OVG

	NKD	OVG
Activities per person / day	1.72	2.32
Trips per person / day	3.03	3.88

If the trip rate of the NKD survey (3.03) is set equal to 100, the trip rate of the OVG survey (3.88) accounts to 128. When compared on this basis there are big differences in the distribution of modes:

In the NKD survey 17 % of all trips were walking trips and 31 % cycling trips. Every third trip was undertaken with a car as driver and every seventh with a car as passenger. Public transport was used for 3 % of all trips. In the OVG survey the share of all modes is higher, as the base – trips per person / day – is higher (128 compared to the original NKD-value = 100). Especially the mode shares for the car (as driver and passenger) and cycling are remarkably higher.

Table 3: **Mode choice: comparison NKD-OVG**

	NKD %	OVG %
Walking	17	19
Bicycle	31	41
Motorbike	1	1
Car as driver	34	42
Car as passenger	14	19
Public Transport	3	4
Other	0	2
TOTAL	100	128
(Trips per person/per day)	(Base:3.03)	(Base:3.88)

Through in depth analysis into systematic errors, three major influences explaining the differences were identified:

- coding conventions
- response rate
- self-selection by respondents

In order to identity the effect of **coding**, the NKD questionnaires were coded independently by two separate coding teams, using the same coding conventions. However, it can be shown, that even with identical questionnaires and coding conventions, there are still differences in the results:

Table 4: Coding-Effects: Impacts on trip number; based on real number of trips / person / day

	NKD	
	Team A % (trips)	Team B % (trips)
Coding errors	+0.5	+3.0
Unlinked trips	-4.9	-
Split of roundtrips	+0.4	-
Questionnaire unclear	-0.6	-+0.4
Trip on private ground	-0.5	-
Sample day replaced	-0.9	-
Commercial traffic	-0.4	+0.4
Missing questionnaire	-	+0.3
Playing of children	-0.5	-
Non-reported-trips / Over-reported-trips (Exploration)	+0.1	0.0
TOTAL	-6.8	+4.1

The most important reason for a higher trip rate in the Team A coding was the acceptance of unlinked trips. (Although the used trip definition was for linked trips only.) Since it is well known that some respondents do report unlinked instead of linked trips, it is important that the coders correct such respondent-related errors. (And this is also a good example highlighting the role of coding as a correctional procedure instead of a "blind punching in of data").

On the other hand, coding is often not seen as an important part of data validation and thus not given enough time and attendance. The result of that can be seen in the underreporting of trips by Team B. (In this specific case, the reason was that the coding had to be done by Team B coders under extreme time pressure.)

So in total the coding produces an over reporting of trips by Team A in the range of 7 % and an underreporting by Team B in the range of 4 %.

It can be easily assumed, that under "tougher circumstances" (e.g. different questionnaires, less supervision of coders, more coders involved, slightly different conventions) the effect of coding only can result in differences in the resulting trip rate of up to 20-30 % (and there are specific examples where this has happened!).

Another example is that of the replacement of sample days. This is typical for respondents, who are generally very active, but were for some unusual reason (e.g. illness), immobile on the sample day. They tend to replace the "odd sample day" (in their view) with a "normal sample day", thus reporting a higher mobility.

A second important systematic error in mobility surveys is the effect of different **response rates.** It is well documented in the respective literature, that often the survey design motivates people with a higher mobility to participate in the survey than those with low or no mobility (e.g. Brög and Meyburg 1980). As shown above, the response rate in the OVG was significantly lower than in the NKD. However, in the NKD, where a strict mail back design was used, results can be analysed by the speed of response and following this well known procedure the results of the NKD could be calculated for the same response rate. This analysis showed, that the trip rate in the "44 % - sample" of the NKD was 11 % higher than in the total sample which gives reason to assume, that the OVG is affected by over reporting of trips by at least that amount.

Finally, **"self-selection"** of respondents is another source of error, which is often ignored in survey analysis. In the OVG design respondents were first approached by telephone and mainly sociodemographic information was collected. Then the respondents were asked, whether they were willing to fill in diary and only when they agreed to do this, was a diary mailed to them. Thus there is a clear element of self-selection. Experiments show, that self selection in mobility surveys result in higher trip rates (ranging between 10 and 30 %, Brög, Erl, 1999, paper for the European Conference of Ministers of Transport (CEMT).

If only these three categories of systematic error are corrected for, the trip rate in the OVG drops from a level of 128 (in comparison to 100 in the NKD) to a level of 97. This indicates that the presumably higher trip rate is in fact a lower trip rate which is affected by underreporting of trips (which is due to a less clear questionnaire in the OVG than the NKD):

Table 5: Mode choice-correction of the OVG survey

	OVG corrected for		
	Original %	Coding / Self selection / response rate %	NKD %
Walking	19	16	17
Bicycle	41	32	31
Motorbike	1	1	1
Car as driver	42	33	34
Car as passenger	19	12	14
Public Transport	4	3	3
Other	2	0	0
TOTAL	128	97	100

3. Systematic Errors in Diary Design – Selected Examples –

The following examples demonstrate how presumably "unimportant" aspects of questionnaire design can spoil the results of mobility surveys. For understandable reasons, these examples have to be presented anonymously; however, they are all "real life" examples

In 1986 two travel surveys were conducted in parallel in the same city: the city travel survey and the national travel survey. The surveys were based on a large number of respondents (more than 2,500 households each) using the same design and the same administration but with a small difference in the questionnaire: unlike Version B, Version A didn't offer the category "car as passenger". This resulted in respondents confusing "Car as passenger" with "Bus as passenger" whereas in Version A they could enter the right answer to the open-ended question in their own words and the categorisation was done by the coder

Figure 1: Questionnaire design of two surveys

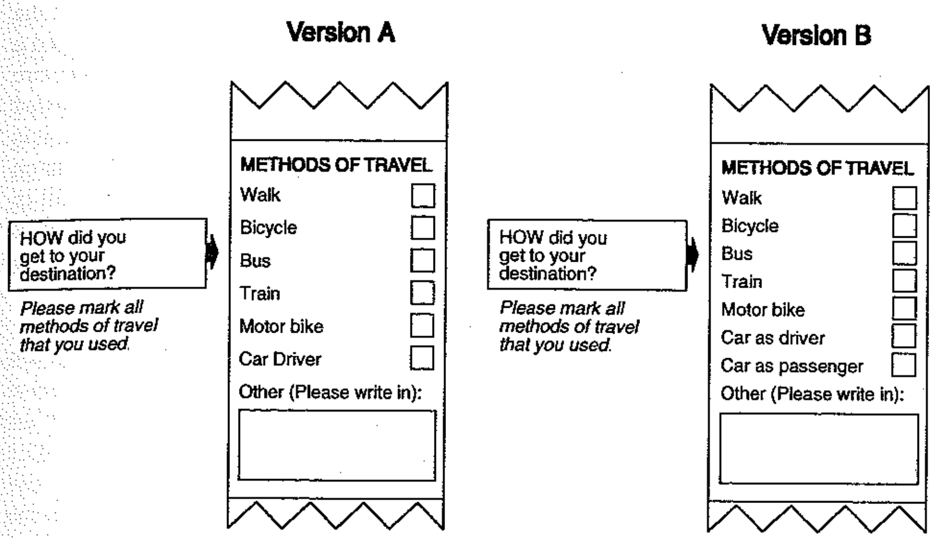


Table 6: Mode choice, different questionnaire design

	Version A %	Version B %
Walking	12	12
Bicycle	6	6
Motorbike	1	1
Cas as driver	58	58
Car as passenger	16	18
Public Transport	6	4
Other	1	1
TOTAL	100	100

All modes had equal shares in both surveys, except the share for car as passenger which was 16 % in version A and 18 % in version B and the public transport share which was 6 % and 4 % resp. Consequently, public transport "lost" about one third of its patronage in Version B due to the fact, that the category "car as passenger" was misunderstood and bus users ticked the wrong box!

An important principle in questionnaire design is to make it "as simple and clear as possible" for the respondent. If this principle is not sufficiently observed, distortions in the results must be expected. The questionnaire in a German Survey of the Elderly provided a specimen trip as an example for the respondents. One of the problems of the questionnaire design, however, was the misleading trip definition in the specimen. The example showed no return trip. Since activity patterns which consist of only one trip are very infrequent, it is not wise that they be used as examples. Furthermore there were three trips to be filled in on the front-page of the questionnaire and no indication for turning to the back page to fill in more trips. It could therefore be expected that a rather large proportion of return trips were not reported and that there was a large number of questionnaires with no more than three trips reported.

The results of the Survey of the Elderly could be compared with "Elderly" from the National Travel Survey, (the so-called German "KONTIV").

Table 8: Comparison of methodological effects in two different surveys

		Survey of Elderly	KONTIV
Basis (Person Travel Days)		9,279 %	24,569 %
Person Travel Days with	1 trip	17	1
	2 trips	47	52
	3 trips	17	8
	4 trips +	19	39
		100	100

It has already been pointed out that, based on the example given to respondents, return trips were not recorded as separate activities. This influence can be clearly seen in the results. The high proportion of person travel days for which only one trip was recorded cannot be explained in any other way. Activity patterns with only one trip are very rare, as shown both by the KONTIV survey and other studies. They can occur when someone returns home on the travel day from a destination where they had spent the previous day or when someone goes to a destination and returns at a later date. It can be reasonably argued that such activity patterns appear rather less frequently for persons aged 60 and over than for other age groups.

And the fact, that there was no indication to turn over the page after three trips in the Survey of the Elderly produced twice as many person travel days in the national survey. As a result of this, the trip rate in the KONTIV was nearly one quarter higher than in the elderly survey.

The importance of every detail in the design of a diary becomes even clearer where the diary collects trips in a "line format" (although literature shows the disadvantage of this approach, see Brög, Däsler, et al 1983) using a specific order of categories and boxes:

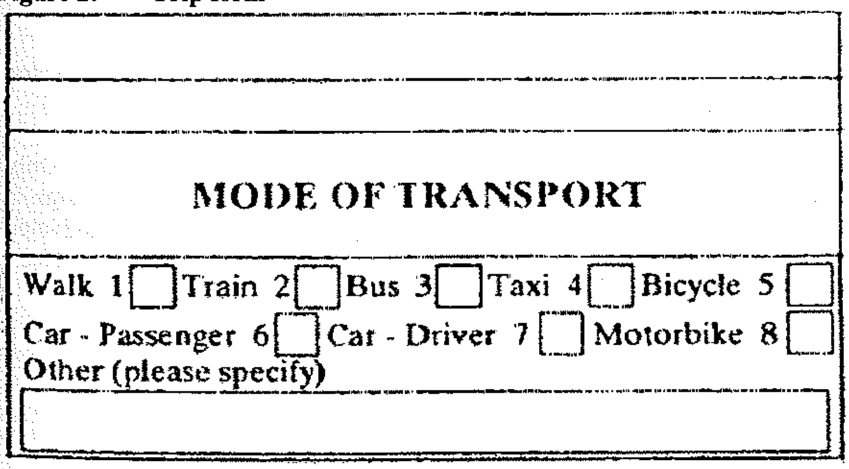
The following example shows an extract of a trip diary using this line format. Each line asks the respondents to give details about each trip, beginning with start time, starting address, ending address, arrival time, trip purpose, mode of transport, number of people in car, where the car was parked and the parking cost.
Up to nine trips could be recorded on a single A4 sheet, with trip pupose and mode of transport response boxes pre-coded.

The following extract of the "Mode of Transport" column shows where this has caused respondents to tick wrong boxes, in particular when they wanted to tick "Car-Driver". They were invited to tick the box between the 6 and Car-Driver (Figure 2). But this was the box for "Car-Passenger"! As a result, in a sample of 7,000 households, there were more than 1,000 trips with "Car-Passenger" but "number of people in car" (which was also collected) was reported as one. There were another 15,000 trips with "Car-Driver" in doubt, with correction impossible because the car occupancy was not reported or was higher than one.

Similar problems can occur with other modes (e.g. ticking train instead of bus, taxi instead of bicycle, etc.) and with trip purposes (e.g. ticking shop instead of education, lunch instead of recreation, etc), where this line format is used.

As a rule, systematic errors cannot be calculated, but can often be corrected (as distinct from random errors, which can be calculated, but not corrected). This represents one of the rare examples, where a systematic error could not even be corrected and the survey results would have only very limited value, if any.

Figure 2: Trip form

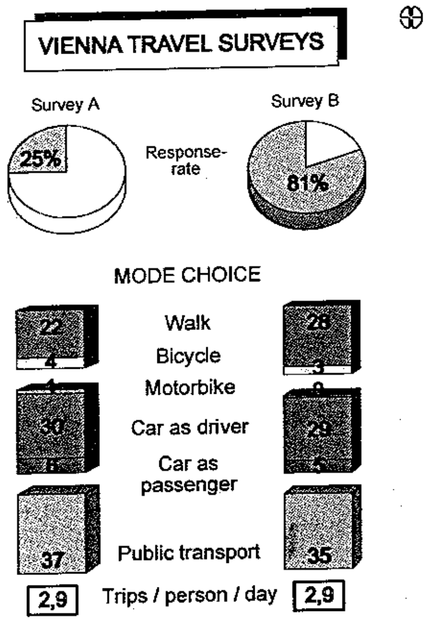

MODE OF TRANSPORT

Walk 1 ☐ Train 2 ☐ Bus 3 ☐ Taxi 4 ☐ Bicycle 5 ☐
Car - Passenger 6 ☐ Car - Driver 7 ☐ Motorbike 8 ☐
Other (please specify)

4. Correction of Systematic Error

In Vienna, Austria, two mobility surveys were conducted (A+B) using a mail-back design. Survey A had a low and survey B a high response rate (25 % as to 81 %). The trips per person per day were equal in both cases (2.9) and this was seen as an indication, that even a survey with a low response rate can provide reliable data on mobility. However, an inspection of mode choice has already indicated, that there might be hidden problems in survey A:

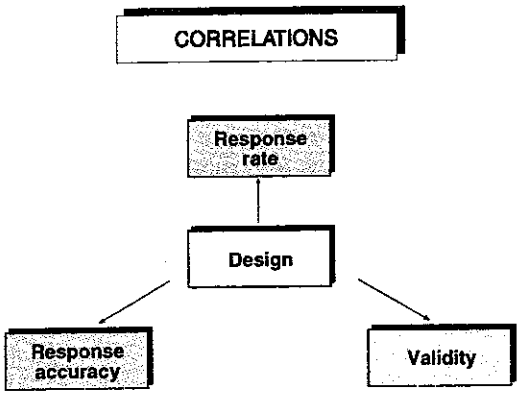
Figure 3: Characteristics of two travel surveys in Vienna



However, the response rate is one quality indicator of surveys, which is highly correlated with other quality indicators:

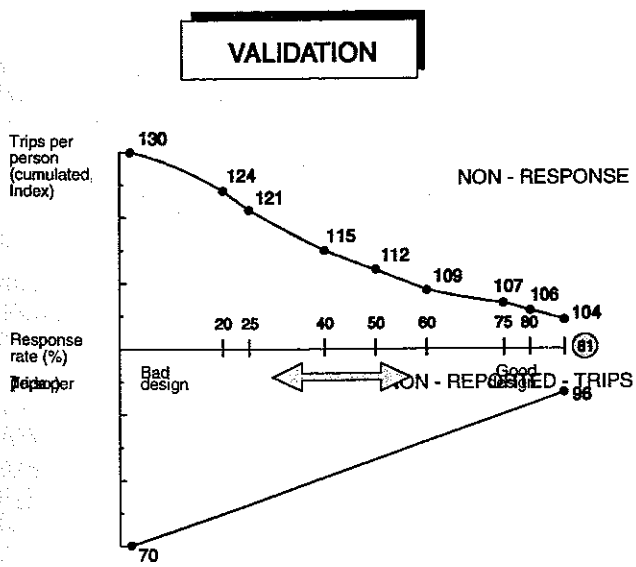
This correlation is easy to understand and can be traced back to the survey design. Only when using a good design, are respondents encouraged to respond. But a good design is also a necessary pre-requisite for accurate responses. And only if respondents are motivated to respond and have little trouble with the survey, will they report complete and valid data.

Figure 4: Correlations



This simple correlation can now be used to estimate the systematic errors hidden in the results of survey A. Since both surveys used a mail-back design it is possible to analyse survey B by speed of response. This analysis shows, as discussed earlier, that the trip rate falls with an increasing response rate. A low response rate usually has its reason in a bad design (this as also the case in survey A, where a confusing, hardly readable diary was employed). Therefore, when a low response rate is associated with a bad design, one of the effects is that the non-reported trip rate will be high with a low response rate and low non-reported trip rate with a high response rate. This correlation is shown in the following figure:

Figure 5: Validation of trip rate



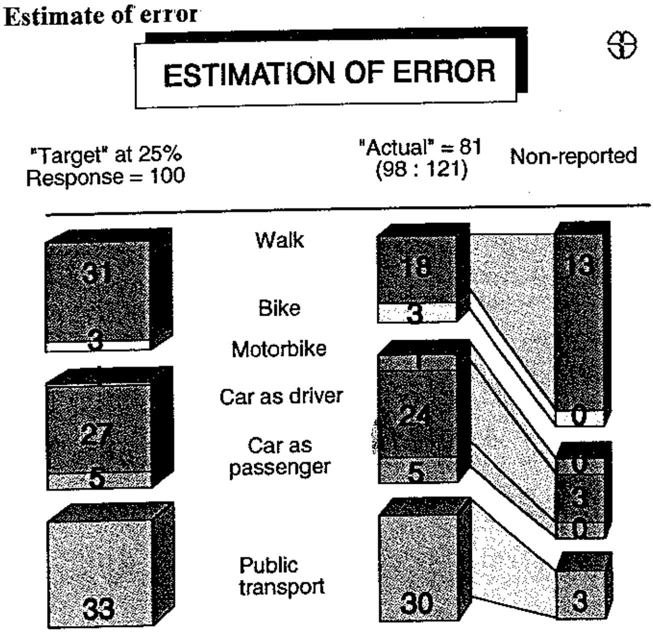
It can been seen, that the response effect and the non-reported trip effect are of the same size, which tend to cancel each other out

We can now use this correlation for a calculation of error:

The expected trip rate of survey A is 21 % higher than the trip rate of survey B. However the real trip rate was 2 % lower. This means, that survey A measured only 81 % of the trips it should have collected (98 out of 121).

If we now compare the mode choice for survey B at a response rate of 25 % (= 100) with mode choice of survey A on a basis of 81 %, we can estimate the number of non-reported trips for each mode:

Figure 6: Estimate of error



It becomes very clear, that the trip rate of 2.9 in survey A was misleading. In reality, survey A was biased by more mobile respondents who did not report a significant part of their trips (with mainly walking, public transport and car as driver missing).

This example also shows that systematic errors can be estimated – and corrected to a certain extent – if self-administered travel diaries are used, methodological information is available and coding conventions are known. But if errors are directly linked to the questionnaire design no further considerations concerning the extent of the errors is possible

5. Conclusion

This paper presented selected examples for systematic errors in the area of empirical survey methods, with the aim to make a contribution to an increased level of awareness and knowledge using data from travel behaviour research.

There are various influencing factors on the validity of survey data such as sampling methods, questionnaire design, survey administration, coding, conventions and definitions, in- and exclusions, which strongly influence survey results. It happens that even the researcher who is in charge of a travel behavioural study, is often not aware of errors which can occur in collecting the data. And even more so the users of the data, analysts, planners, modellers and political decision makers who, depending on and believe in, data which can be strongly influenced by systematic errors.

But the classical thinking in the field of empirical research is still dominated by numbers and significance-levels, but statistical figures are irrelevant if the systematic error is larger than any confidence interval.

Surveys are the output of a complex communication process. One has to know the inputs and the participants of the process to use and assess the results of this process correctly.

References

Ampt, L and Brög, W (1984) *Errors in Travel Surveys: The Significance of Methodology*; Paper for the ARRB, Hobart

Brög, W and Neumann, K-H (1977) *The Interviewee as a Human Being - A Critical Contribution to the Discussion of Methods in Empirical Social Research* Paper presented to the ESOMAR Seminar on "Ways and New Ways of Data Collection", Jouy-en-Josas, France

Brög, W (1979) *Passenger Transport: Mobility and Lifestyle - Sociological Aspects* Report to the "Eighth International ECMT Symposium on Theory and Practice in Transport Economics", Istanbul, Turkey

Brög, W and Meyburg, A H (1980) *Nonresponse Problem in Travel Surveys: An Empirical Investigation* Paper presented to the 59th Annual Meeting of the Transportation Research Board (TRB), Transportation Research Record No. 775, Washington, D.C.

Brög W and Daesler Ch and Förg O G (1983) *An Empirical Test of Various Survey Instruments for the Recording of Travel Behaviour* Paper presented to the 2nd International Conferenc "New Survey Methods in Transport", Hungerford Hill, N.S.W., Australia, in: Conference Proceedings "Workshop Papers Questionnaire Design and Piloting", p. 1-26

Brög, W and Meyburg, A H (1981) *Consideration of Non-Response Effects in Large-Scale Mobility Surveys* Paper presented to the 60th Annual Meeting of the Transportation Research Board (TRB), Transportation Research Record No. 807, Washington, D.C.

Brög, W (1997) *"Raising the Standard!" Keynote paper zur internationalen Konferenz - "Transport Survey Quality and Innovation"*, Grainau

Meyburg, A H and Brög, W (1981) *Validity Problems in Empirical Analyses of Non-Home Activity Patterns* Paper presented to the 60th Annual Meeting of the Transportation Research Board (TRB), Transportation Research Record No. 807, Washington, D.C.

Magistrat der Stadt Wien (1993) *Mobilität in Wien* Beiträge zur Stadtforschung, Stadtentwicklung und Stadtgestaltung, Bd 45, Wien

Moritz, G and Brög, W (1999) *Redesign of the Dutch Travel Survey Response improvement* Paper presented at Transportation Research Board (TRB) on Personal Travel: The Long and Short of It, Washington D.C. (Restricted)