Household or individual – Advantages and disadvantages of different interview selection strategies

Stefan Hubrich*, Rico Wittwer

Technische Universität Dresden, Faculty of Transportation and Traffic Science, Chair of Transport and Infrastructure Planning, 01062 Dresden, Germany

Abstract

In light of the growing importance of transport analysis since the 1960s, interviewing households has become a proven means of gathering information concerning the travel behavior of the population. In spite of the exploitation of various classical methods and the partial use of new media, the acceptance of household interviews in Germany has deteriorated sharply over the last few years. This can be seen in the continuously declining response rates. A low response rate may have a negative impact on the quality of the data gathered and on the acceptance of the published results. The Technische Universität Dresden is currently analyzing the possibility of halting the declining response rate by combining different survey methods that differ in regard to the target group. One of the key questions in considering a target-group-specific approach is the decision of whether to choose the household or the individual as the primary unit of analysis. Deciding on a favorable sampling strategy requires careful reflection on the advantages and disadvantages in terms of statistics, content and the practicability of the survey.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/). Peer-review under responsibility of PANAM 2014.

Keywords: Household survey; travel survey; travel behavior; cross-sectional survey; reference day survey; unit of analysis; selection strategy; response rate; cluster sample; sampling method

* Stefan Hubrich. Tel.: +49-351-463-34126; fax: +49-351-463-37264.
E-mail address: stefan.hubrich@tu-dresden.de
1. Introduction

Since the 1970s, household questionnaires have been an established practice in Germany for surveying the travel behavior of residents. Due to the separation of Germany after WWII, two cross-sectional surveys have been developed within the last 40 years. After Reunification in 1990, these surveys were continued at regular intervals as Mobility in Germany (MiD), conducted by the Federal Ministry of Transport, and Mobility in Cities – SrV (“SrV” stands for, in German, “System of Representative Transportation Surveys”), conducted by cities, municipalities and transportation companies and associations. The two large German cross-sectional surveys are connected by the German Mobility Panel (MOP), particularly during the periods between MiD surveys. This longitudinal instrument also makes intrapersonal behavior analysis possible.

The necessity of conducting household surveys as an essential basis for strategic transportation planning was first recognized and discussed academically in Germany at the Technische Universität Dresden, in the former GDR, at the end of the 1960s. After various test surveys, the instrument was first implemented in 1972. Since then, continued technical guidance and organization of the SrV survey has been entrusted to the Technische Universität Dresden, which sees itself as the trustee of the cities and has the task of continually developing the method further and adapting the survey to current applicable conditions, as well as taking care of routine administration.

The SrV, which is arranged as a reference day survey, is repeated regularly, i.e. every five years, and was conducted for the tenth time in 2013. Its results primarily aid the planning and politics of municipal transportation development in overcoming one of the central tasks: the continual observation and evaluation of developments and effects. The data build a significant foundation for strategic and control-oriented plans, as well as for demand and effect calculations of measures, based on transportation models. The results serve politics and the public sphere as material that describes the transportation landscape and its development over time, as well as from city to city. Furthermore, the data are made available to research organizations for secondary analysis.

Measured by continually worsening response rates, the acceptance of household surveys has decreased dramatically in western countries in the last years and decades (see de Leeuw & de Heer, 2001, Singer, 2010, Massey & Tourangeau, 2013). This trend can also be identified for the latest rounds of the MiD and SrV surveys, despite extensive, comprehensive use of classical methodology and the at least partial use of new media (e.g. online surveys). Figure 1 shows the development of the SrV and KONTIV/MiD responses since 1972. The KONTIV surveys of 1976, 1982 and 1989 are the predecessors to Mobility in Germany (MiD) – “KONTIV” stands for, in German, “Continual survey of Transportation Behavior”.

![Fig. 1. Development of the SrV and KONTIV/MiD responses since 1972](KONTIV/MiD survey response rates: Scheiner, 2009, Follmer et al., 2003 and 2010)
A small number of responses has a general, negative effect on the quality of the data gathered and on the acceptance of published results, although response rate alone is an insufficient quality criterion (Shlomo, Skinner, & Schouten, 2012, Schneekloth & Leven, 2003, and Koch, 1998). Response rate only influences survey results negatively if the error is systematic, i.e. selectivity and original non-response effects occur (Groves, 2006). Thus, the Technische Universität Dresden is now investigating to what degree a target-group-specific combination of different survey methods can stop the ever-sinking response rate, and what consequences new survey methods can have on, among other factors, the representativeness and quality of the data, i.e. which problems arise when connecting different survey methods.

A basic investigative focus in dividing such target groups is the question of whether the travel patterns of every person in a household, or only those of one, randomly chosen member, should be surveyed for a compilation of mobility behavior. This is of particular interest from a statistical perspective.

In all large household surveys on mobility behavior in Germany, the household (with all of its members) has been the primary analysis unit for the past 40 years. Households were chosen randomly from the registry of residents by their members with probabilities proportional to size (PPS), and were contacted as survey units, at which time information was gathered regarding both mobility on specific reference days and mobility conditions for each household member. The reason for PPS is the characteristic of the German registry of residents, in which there is no logical connection between households and the people living in them. This process is denoted here as Strategy 1, and is predominantly founded on content-oriented aspects, particularly the complexity of the interaction of mobility decisions within the household, and on practical surveying advantages, primarily the long time required to conduct face-to-face surveys.

Another strategy is to take the person chosen from the registry of residents as the survey unit with simple random sampling (SRS). This would mean that reference day mobility would only be determined and evaluated for that one person. Furthermore, general information about the structure and features of the household, as well as about the other household members (socio-demographic information, mobility-relevant personal details) would be requested. However, for evaluations at the personal level, only the characteristics of the selected person are used. This process is denoted here as Strategy 2. Figure 2 compares and contrasts the two strategies.

Recent investigations (Bonnel & Armoogum, 2005) show that, in the worldwide survey landscape of household surveys on transportation behavior, differences definitely exist regarding the sampling strategy, and therefore, the analysis unit (see Table 1). While in Germany, as already mentioned, Great Britain and the USA, all members of a household are interviewed about their mobility behavior on the reference day, in other countries, this data is only gathered for one, randomly chosen member of the household, particularly in Scandinavia. Furthermore, solutions that strike a compromise between the two strategies exist in countries such as Switzerland or France, where, depending on the size of the household, one or two people are interviewed. In recent years, the present strategy has been increasingly called into question in Germany, as the statistical perspective, in particular, seems to speak for selecting the individual as the unit of analysis.

![Fig. 2. Gathered information when using survey Strategies 1 and 2, comparatively](image-url)
Table 1. People interviewed in selected household surveys (according to Bonnel & Armoogum, 2005)

<table>
<thead>
<tr>
<th>Number of surveys</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Only selected individual</td>
<td>5 (39%)</td>
</tr>
<tr>
<td>One or two individuals</td>
<td>2 (15%)</td>
</tr>
<tr>
<td>All household members</td>
<td>6 (46%)</td>
</tr>
</tbody>
</table>

2. Objectives

The goal of this paper is to compare the advantages and disadvantages of the two different selection/analysis strategies. The focus is particularly on the requirements and conditions of the cross-sectional survey Mobility in Cities – SrV. The fundamental character of this survey is, however, transferable to other cross-sectional surveys of travel behavior, especially those that use the registry of residents as the basis for selection. The background of these considerations is also, as explored above, the declining number of responses in the German MiD and SrV surveys, as well as for transportation surveys in general. This development increases the amount of effort required to achieve the planned sample, has a negative effect on data quality, as mentioned above, and endangers the acceptance of the data. Since hardly any discussion of the question of sampling strategy can be found in the literature, this paper is an attempt to initiate an in-depth discussion of this question at the international level.

3. Methodology

To answer the question of which survey strategy is most advantageous, one must methodically investigate three evaluation areas: statistics, i.e. sampling theory, content, i.e. completeness in the proposed realm of influence for mobility decisions, and feasibility/practicality, i.e. the practical consequences of the survey, such as cost, response rate, etc.

In order to not only analyze the statistical effects theoretically, but also to demonstrate them practically, comprehensive survey simulations were conducted, which showed, among other observations, the different variations/spread of the results, depending on the sampling strategy used.

The basis for this is a “pseudo-population”, which consisted of the 38,965 people belonging to 18,372 households, who were interviewed in Berlin for Mobility in Cities – SrV 2008 and who built the selection basis for 200 samples. For 100 of those samples, 1,000 people each were chosen and became the analysis unit for their respective samples (Strategy 2). For the other 100 samples, 370 people per sample were randomly chosen, and they, along with the members of their households, likewise amounted to around 1,000 persons. In these samples, the complete household was counted as the analysis unit for each sample (Strategy 1). The inclusion probabilities of Strategy 1 result in using PPS as sampling theory.

For each sample, the mobility activity on the reference day, the average number of trips per day, the share of motorized private transport in the modal split, as well as the average distance and duration of a trip were calculated, and the data from samples obtained with Strategy 1 and Strategy 2 were put together for comparison. To estimate the accompanying confidence interval (CI, with an error rate of 5 %), the procedure for unlimited random sampling could be used for the samples obtained with Strategy 2, while those samples using Strategy 1 required a complex variance estimation process. This complex process is based on a Taylor linearization approach (documented in Wittwer, 2008). The variance estimation was conducted with the Complex Samples supplementary module, which is available in SPSS 18. This way, the most important influences on sample design, particularly the stratification effect and the cluster effect on the household level, can be considered in the variance estimation.

Since no error process was modelled, the parameters of the samples selected with Strategy 2 were determined without weighting. The use of Strategy 1 leads to a biased sample resulting from the PPS design. In addition, due to the questioning of all household members, the socio-demographic structure of the sample does not correspond to the “pseudo-population”. For these reasons, the evaluation was carried out by weighting the Strategy 1 sample according to household size (4 groups), age (5 groups), and gender (2 groups). Figure 3 summarizes the methodology of the conducted simulations.
The consequences for the content are still being discussed with a number of colleagues in the field, and are thus only mentioned briefly in this paper.

In order to assess the practical aspects for surveying, the CATI (Computer Assisted Telephone Interview) protocols originating from the survey run of 2008 were evaluated. Thus, the total length, among other aspects, of all (telephone) interviews from Mobility in Cities – SrV 2008 (questioning of all household members) could be put up against the (fictitious) case of a sample equal in size in which only individuals are surveyed. For the MiD 2008 as well, other additional variables are available with the Scientific Use File.

4. Results

4.1. Statistical aspects

From a statistical perspective, the sampling strategy used currently (Strategy 1) has both advantages and disadvantages. The household represents a meaningful social unit. Therefore, the fact that the dependence of investigative characteristics within a household can be statistically examined more deeply is an advantage. In modelling, such situations can, by now, be adequately treated through the use of multi-level models (for basic information about multi-level models, see e.g. de Leeuw & Meijer, 2007). If only the mobility characteristics of the person chosen from the residents’ registry are compiled (Strategy 2), such analysis possibilities are very limited.

A critical disadvantage of the currently used strategy is, however, that the sampling process leads to clustering in the sample at the household level, and thus to a loss of efficiency and, finally, inaccuracy in the determination of parameters. Another disadvantage is the dependent relationships within measured data (particularly at the trip level, e.g. in choosing a mode of transportation). The independence of residuals from within the sample is an essential prerequisite for the use of inferential statistics procedures. This disadvantage does not arise with Strategy 2. Instead, the sampling process is significantly simplified, and the cluster effect is avoided in the evaluation on the household and individual levels, leading to an efficient determination of parameters.
The simulations described above, conducted at the Technische Universität Dresden, impressively show the relationships expected by sample theory, namely, higher precision of estimation (a clearly smaller dispersion of the point estimations among each other) for Strategy 2, with the same expected value estimated, compared with Strategy 1.

The averages calculated from each of 100 samples, which were compared with respect to the values investigated, provide the data basis for the boxplots in Figure 4. The left data sets belong to Strategy 1 (all members of the household), while the right set is for Strategy 2 (individuals).

The “box” in the boxplots contains the middle two quartiles, i.e. the middle 50% of all calculated average values. The diamond represents the mean value, while the horizontal line in the box denotes the median of the calculated averages. The antennae (whiskers) denote the corresponding minimum and maximum of the averages. Their length is limited to 1.5 times the inter-quartile distance (height of the box) or, possibly the smallest and largest average values (if these are within 1.5 times the inter-quartile distance). The averages’ minimum and maximum are represented by a cross. The red line shows the characteristic of the individual parameters with respect to the whole population.

It is clear that the calculated averages for both sampling strategies, for all six parameters, are near the characteristic values for the whole population. Equally recognisable is that, for the same sample size (1,000 people), the range of variation of the results is significantly smaller for surveys of individuals (Strategy 2), compared to surveys of entire households (Strategy 1).

If one compares the 100 samples in pairs, an overlap in the confidence intervals can be observed in almost every case. Figure 5 depicts average trip distance as an example. If the confidence intervals on an error level of alpha = 5% do not overlap, one can assume a significant difference in the averages when the confidence intervals are calculated exactly. This process has approximately the same importance as a two-sided parametric significance test (Sedlmeyer, 1996). It can thus be determined that the type of sample selection (Strategy 1 or 2) does not have a significant influence on the characteristic of the parameters themselves.

Fig. 4. Boxplot representation for the parameters examined

Fig. 5. Comparison of the sample results of Strategy 1 and Strategy 2
Table 2. Number of cases with overlapping confidence intervals

<table>
<thead>
<tr>
<th>Parameter</th>
<th>At place of residence</th>
<th>Mobility participation</th>
<th>Trips per day</th>
<th>Private motorized transport</th>
<th>Distance</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence intervals overlapping</td>
<td>98 cases</td>
<td>99 cases</td>
<td>100 cases</td>
<td>98 cases</td>
<td>97 cases</td>
<td>96 cases</td>
</tr>
</tbody>
</table>

Table 3. Comparison of sampling strategies 1 and 2

<table>
<thead>
<tr>
<th>Entire household (Strategy 1)</th>
<th>Individual (Strategy 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sampling process</strong></td>
<td>One-level cluster sample:</td>
</tr>
<tr>
<td>(1) Random sampling of one person from the registry of residents</td>
<td>(1) Random sampling of one person from the registry of residents</td>
</tr>
<tr>
<td>(2) Gathering information regarding mobility on reference day of all household members</td>
<td>(2) Gathering information regarding mobility on reference day of the chosen person</td>
</tr>
<tr>
<td>(3) Conversion to a household-based sample leads to a PPS sample of households</td>
<td></td>
</tr>
<tr>
<td><strong>Sampling influences</strong></td>
<td>Large households are over-represented, thus, weighting is necessary even without non-response</td>
</tr>
<tr>
<td><strong>Expected effects on the results</strong></td>
<td>• Clustering leads to variance increase</td>
</tr>
<tr>
<td></td>
<td>• Complex methodology necessary for variance estimation</td>
</tr>
</tbody>
</table>

In a few cases (no more than 4 cases for average trip duration, see Table 2), the confidence intervals do not overlap, and so a significant difference in the average values must be assumed. Since the confidence interval is calculated based on an error probability of 5 %, it can be established that the intervals do not overlap in up to 5 of the 100 compared pairs. Therefore, the results are in the expected range.

The statistical aspects can be summarized thus: the absence of the cluster effect influences the accuracy of the survey positively. Sample design is greatly simplified. The survey characteristics are independent at the individual level. Statistical disadvantages of this sampling strategy (Strategy 2) could not be uncovered. Table 3 summarizes and compares the essential statistical aspects of both sampling strategies.

4.2. Content oriented aspects

With regard to content, the current strategy (Strategy 1) has advantages over Strategy 2 in that, by collecting the entirety of the mobility characteristics of a household, dependent relationships within the household are also surveyed and can be investigated scientifically through secondary analyses, so that the special place of the household as an organizational form for mobility is taken into account. This advantage is, however, not relevant to the primary goal of the Mobility in Cities – SrV survey, providing parameters for transportation planning and politics, and is not used by the commissioners of the study (cities and municipalities, transportation companies and associations). Scientific analyses in the German language based on this research that concretely investigate mobility decisions within a household at the trip level are unknown to the authors (with the exception of Gliebe & Koppelman, 2002, and Engelbrecht et al., 2005).

Such analyses are no longer possible if mobility information is gathered according to Strategy 2. However, the characteristics of the household and of the other household members, which are to be acquired with Strategy 2 as well, can function as explanatory values in statistical models. The implications of different procedures with respect to content in modelling processes present another question. First, the type of modelling process must be distinguished. For macroscopic transportation planning models, which usually follow the classic Four Step Model...
(FSM or 4SM): trip generation, distribution, mode choice, and route assignment, and mostly build on groups of people with the most similar behavior, a change in the sampling strategy has no consequences for content. The necessary initial and calibration values can still be made available (Wittwer, 2008).

The question is different for microscopic implementations. Such models are, in some ways, very different in design. Agent-based simulations are a representative form. A logical simulation concept is to use the information of entire households from the underlying behavior surveys. These households can then be fully simulated, i.e. with all members of the household and with consideration for the (temporal) availability of the household’s motor vehicles and other limiting conditions (such as the presence of other members). Therefore, this approach is based on information that is only available in surveys conducted with sampling Strategy 1. Until now, however, these restrictions within a household were not considered in models. In the German language area, this field of study has, however, been intensively researched at various institutions, and so a corresponding extension of the models is nevertheless to be expected. An important project in this field is the open source software MATSim, providing a framework to implement large-scale agent-based transport simulations (see http://www.matsim.org/). It must be noted, however, that in the view of the paper’s authors, at this time, the number of practical applications is currently relatively limited.

4.3. Practical aspects for surveying

Approximately 77% of all households that took part in the Mobility in Cities – SrV 2008 were interviewed by telephone (Ahrens et al., 2009). The evaluation of the corresponding CATI protocols confirms the conjecture that a linear relationship exists between the total length of the interview and the size of a household (see Figure 6). With a coefficient of determination of $R^2 = 0.996$, a linear regression equation can be determined, which shows, in addition to a base time frame of about 13 minutes for single-person household, an interview time of about 6 minutes for each additional member of the household.

The total interview time for all people surveyed by telephone was 748,608 minutes, and the average household size for those interviewed by phone was about 2.23 people. If the method of surveying only one person per household were to be adopted for such a target sample, 2.23 times as many households would have to be contacted. Theoretically, regarding the CATI effort (total interview time), at least a 150% increase is to be expected.

The assumption that the work and cost of a survey are less when surveying all members of each household (Strategy 1) than when interviewing only 1 member (Strategy 2) is without any doubt reasonable for earlier survey runs (MiD until 1989, SrV until 1998), in which the households were personally sought by the interviewers (face-to-face). For current survey designs (telephone/written-by mail/written-online) or other combinations of methods, no practical experience with this assumption is available. For example, a higher response rate, based on the shorter necessary interview length, could decrease both effort and cost. Therefore, in the context of this discussion of methods, researchers are striving to collect practical experience in this area through field testing.

![Fig. 6. Length of telephone interview as a function of household size](image-url)
After the end of the field time for a run of Mobility in Cities – SrV, the raw data is checked at the Technische Universität Dresden for consistency and plausibility, and possible contradictions are indicated and rectified as appropriate. An approach for identifying problematic data is the reconciliation of trip information within a household. If the information about trips taken on the reference day is only available from one person, other approaches must be implemented more intensely in order to recognize possible implausibility.

On one hand, some of the missing information can be acquired by slightly extending the catalogue of questions. Thus, the current question about the number of accompanying household members on a trip should be supplemented by a concrete determination of which members made the trip being reported. This additional information is also helpful for more in-depth analysis of interactions within the household. At least for the collection of trip information by telephone, better information for the interviewer would be an approach to improve data quality during the interview by both target-group-specific control questions (e.g. “How did your child get to school on the reference day?” as a question to an adult in a household with at least one school-age child), and visual aids during data acquisition in the CATI environment. These could be graphical representations of the day’s schedule for better recollection of “forgotten” trips (similar to the software tool Gradiv, see Zumkeller et al., 2011), or maps of routes travelled in route panning software (implemented in SrV 2013 based on Google Maps API).

5. Discussion

In Germany, the questions presented in this paper are the subject of intense controversy among scholars in the field, and this article is meant to illustrate the current state of discourse on this subject in Germany. This is to build a foundation for a conflation of experiences from different countries in order to form a basis for evaluation that is as comprehensive as possible and to achieve an assessment of all advantages and disadvantages in view of future SrV surveys.

The high accuracy of the individual person sampling strategy increases the efficiency of the sample significantly. Thus, this strategy is clearly more advantageous, and the cost-benefit ratio is clearly better for the same per person cost. However, it is possible that a higher expenditure per person is to be expected when surveying individuals, compared to the current strategy, especially as a result of the need to contact a greater number of households to reach the same net sample size, and because the total interview time per successfully interviewed person increases. Other practical disadvantages of individual surveying include a lack of opportunities for control and for plausibility determination. It is indisputable that the household as a social unit plays an essential role in the determination of the members’ travel behavior. Investigations that analyze the relationships involved with the role of the household are, however, rare, and mostly deal with only general characteristics or mobility characteristics of the household, and not with the real mobility of all household members.

6. Conclusions

Both inquiry strategies exhibit various advantages and disadvantages. While practical surveying considerations in particular speak for the household as the unit of analysis, surveying individuals is recommended from a statistical perspective. With regard to content, information about the mobility of other household members on the reference day is absent in surveys of individuals, which severely limits analyses of interdependencies in mobility decisions within a household. Indeed, such analysis is not the goal of the cities that commission Mobility in Cities – SrV; rather, it is “merely” an academic motivation.
References


<http://mobilitaetspanel.ifv.uni-karlsruhe.de/index.php?eID=tx_mm_bccmsbase_zip&id=3637161252e782e08ded5>.