

Workshop N°9: Validating Shifts in the Total Design of Travel Surveys

Workshop Chair: Tony Richardson

Description:

The process of trying to achieve an optimum balance in survey design decisions to achieve the best total quality is known as the “total survey design” approach. In this approach, major efforts are taken to better understand, and therefore, to control both sampling and non-sampling errors throughout the design, capture, processing, and analysis of survey data. New approaches available to design travel surveys promise the capability of collecting better quality data while accommodating increasing budget restrictions and expectations. But the implications of these shifts in total survey design have not been well researched or documented. This workshop focuses on important issues in understanding the implications of implementing changes in survey design, such as improvements in telephone instruments, using GPS devices, or developing online survey systems. What are the implications in terms of the validity and reliability of the resulting information and for its utility for transportation planning and policy-making?

Associated Papers:

1. Using accelerometer equipped GPS devices in place of paper travel diaries to reduce respondent burden in a national travel survey

Abby Sneade (Department for Transport, UK)

Best practice and professional protocols dictate that survey managers continually seek to reduce respondent burden. The current financial climate also requires cost savings to be identified wherever possible. Previous research undertaken by the Department for Transport (DfT) concluded that GPS data collection has real promise for the GB National Travel Survey (NTS) and offers the most suitable option for delivering affordable and practical improvements in the quality and reliability of the NTS diary data with no fundamental barriers of feasibility or public acceptability. The use of GPS devices in place of travel diaries would also record more accurate data, including precise time and location start and end points and improved estimates of the distance travelled.

The respondent burden of the NTS travel diary is high. The quality of data is dependent on respondents’ interpretation of the instructions and accurate completion. Some respondents with visual impairments, low literacy levels or limited English language may struggle to complete the diary. The comparative burden of GPS is low: simply carrying the device and remembering to charge it for a few hours each evening, much like a mobile phone.

Potential drawbacks associated with switching to GPS data collection include the loss of some diary items, not all of which can alternatively be collected using interview questions, although in time other emerging sources may substitute some of this data. For practical and ethical reasons it may be necessary to exclude children, or at least those under the age of 12, from a GPS sample.

In September 2010, DfT began preparation for a pilot survey that will define how the NTS might be run using GPS devices in place of the travel diary. This pilot will employ a representative achieved sample of 600-1,000 GB residents aged 12+ who will each be interviewed and asked to carry a GPS device during the seven day travel week. The pilot methodology will remain unchanged from the current NTS methodology wherever possible. The core of the pilot questionnaire will also remain unchanged, but will exclude non-essential questions to create space to test new questions to assist in the processing of GPS trace data, weight results, or replace deleted diary items. The pilot will be conducted by the current NTS contractor, the National Centre for Social Research (NatCen). Data collection will take place in February and March 2011.

The GPS data and elements of the interview data will be processed by a sub-contractor. We anticipate that a mixture of rule based and probability based algorithms will be used to process the GPS data using personal GIS data collected during the interviews; standard GIS data on points of interest and the transport network; public transport service data; time; average speed travelled and the accelerometer traces.

DfT and NatCen are currently securing a processing supplier and therefore can not presently name the supplier or provide a detailed methodology; however, these would be included in the final submitted paper. The processing will identify trip and trip stage start and end points, predict the purpose and mode of these and link the processed GPS data to interview data. Output files will also contain metadata on selected locations and the probabilities assigned to the inferred mode and purpose.

A DfT statistician will analyse the processed data and prepare a summary report on the findings by Q3 2011. A technical report outlining the methodology and any remarkable experiences during fieldwork will be prepared by NatCen. Details of these findings would be included in the final paper.

It is anticipated that inferred data will be validated using a visual tool displaying mapped GPS traces and the inferred purpose/ mode for each trip/trip stage. This would be completed by a researcher rather than the respondent as it is not practicable to use online methods as nine million GB residents have never used the internet and 27 per cent of households do not have internet access, and the complexity and costs of additional mixed-mode follow-up surveys are not feasible. Although the use of 'proxy' validation by researchers will not entirely eliminate uncertainty about inferences, it will provide a valuable sense check and allow improvements to be made to the processing system algorithms.

We can only speculate on the outcome of the pilot at this point in time, however, commentary on this would be included in the final paper. We anticipate that the results will determine how well the GPS NTS methodology works; that it will identify any unforeseen problems and whether this method provides similar results to traditional methods. It is expected to reveal whether the algorithms used to predict mode and purpose are accurate and well functioning and provide some indication as to whether the certainty of these predictions are within acceptable parameters. Special consideration of the response rates for children aged 12-15 pilot will also determine the feasibility of including this group in a GPS NTS sample.

The pilot will identify the scale of any cost savings to be made in replacing paper data collection materials and the subsequent manual coding and data entry with GPS devices and automated data processing.

If the pilot proves successful, GPS methodology could be as the sole data collection method for the NTS from 2013. A switch to 100 per cent GPS collection in such a well respected and established national travel survey could herald a significant milestone for the global travel survey community, and may provide others wishing to move towards GPS methodology with the evidence and leverage to do so.

2. The role of web interviews as part of a national travel survey.

Linda Christensen (DTU Transport), Carsten Jensen (DTU Transport) and Hjalmar Christiansen (DTU Transport)

Background

Over the years National Travel Surveys (NTS) have been the most important source for both construction of national transport models and policy making according to the travellers' behaviour. However, data collection with interviews is a very expensive way of getting information about travel activities. National and local authorities might therefore often look for possibilities to collect these data in a cheaper way. A web-based survey is an obvious solution to save money.

The most normal method to collect data by web-based interviews is to contact a survey firm which has a big panel of costumers willing to answer these surveys. This will save most money in relation to the actual methods. However, the resulting travel behaviour of respondents in a NTS is very sensitive to the data collection method (Christensen, 2005). Therefore, it is necessary to make sure that the methodology will not change the collected data.

First of all survey data has to be representative for the population according to both socio-economic parameters and to travel behaviour. Age s an example of a simple factor of

importance for car ownership, mode choice, and for number and distances of daily trips. Education level and/or income are important for travel distance.

However, access to computers, response rate on web-based surveys, and willingness to be registered in a survey panel are dependent on age, educational level etc., and the available time to complete the surveys. And time for participating is among other things dependent on travel time during the day. A decision to choose a web-based survey will therefore influence the results of the survey.

On the other hand personal and telephone based interview surveys are sensitive to methodology too. Some of the problems with the web-based surveys are seen in traditional interview surveys too, for instance that the response rate depends on age, education, income level etc. too. The question is whether the socio-economic influence is stronger in case of a web-based survey than in a traditional survey methodology.

Furthermore the possibility to get contact to the respondents is dependent on when and how much time the respondent is at home, which again is correlated with used travel time and the time it takes place. A web-based survey might reduce the problems as it is possible to get contact to some of those who are not available in the most frequent interviewer contact periods.

However, the most important problem with an interviewer based survey is the interviewer effect. A special interviewer effect discussed in the literature and observed in the earlier Danish NTS is soft refusal. This is the case when the respondent is not willing to participate but does not say it directly. He will for instance agree on an appointment for a later call. But after numerous calls and appointments he is accepting to participate to avoid more calls. The soft refusal strategy could for instance be to lie and say that he had not left home during the day.

In a web survey the results are independent of the interviewer effect. But of course it is still dependent on how the questions are formulated and even more when no interviewer can explain what the intention is. And the respondent is also trying to get the survey finished after an acceptable period.

Purpose

The purpose of the paper is to document the advantages and disadvantages of using a web survey in a National Travel Survey data collection and indicate which problems to be most aware of. The questions to be asked are first of all whether one of the two methods is the best, or whether it would be better to let them complement each other.

The paper will address the following questions:

- Differences in response rate dependent on data collection methods. Which method is most useful related to different groups. Will the web survey method make it possible to get contact to respondents who are not easily accessible in a telephone based survey
- The effect of data collection methodology on travel behaviour analysed by no travel rate, number of trips and length of trips
- Quality of data in general
- The possibility to save money by using a web-based survey

Methods

For the explained reasons it was decided not to use a survey panel for the Danish NTS. The web survey is only a part of the survey where the representatively sampled respondents choose if they want to participate on the web. If they are not answering on the web they are contacted by telephone for a CATI.

The analyses are primarily based on multivariate regression analyses on the datasets by SAS. At the end of 2009 91.000 respondents had been sampled since May 2006 when the survey started. 13 % of the sample is responding on web which means that 22 % of the overall resulting interviews are done by web.

Conclusion

The analyses show that the web survey is contributing to a higher response rate which is not compensated by later lower response rate in the CATI. This is clear for the adult groups, especially for middle-aged people, couples with no or only one child. Furthermore the web is a really good media to get contact to children of both genders.

The web interviews reduces to some extent the problem with missing telephone numbers. But as the response rate is also lower on the web it does not add much information to the travel behaviour.

But the web is not a solution to every problem. The young, those outside the labour market, those with language problems etc., are not met on the web. If the response rate for these should be increased other solutions have to be found. For the young oversampling might be considered. For those with language problems some special interviewers with knowledge of urdu etc. would help.

The analyses show that there is little difference in travel activity between those who are answering by web and those answering on CATI. The only group for which such a difference is observed is the elderly over 65 for whom answering on the web results in both the travel rate and the average travel distance per trip being higher.

The two survey media results in the same travel behaviour according to the share of respondents with no trips. But the web interviews reveal a too low number of trips in the CATI interview. It is the shorter trips which are left out – at least on average - resulting in only 2 % less travel distance. But still it is a problem which should be followed in the future.

The conclusion is that web interviews are saving money, in our case 74 % per web interview. But our survey shows that the web interviews need much more post processing than the CATI where the interviewers are trained. So it is important not to base a conclusion of the savings on the price for collecting web interviews instead of CATI. Taking the post processing and the low response rate into account the overall cost reduction is only 15 % when it is decided to use the procedure in the Danish survey.

As it is proved that the web interviews adds some quality to the survey by getting into contact with people who are not participating else it should be considered to do an extra effort to contact those who were impossible to contact to by telephone calls by sending out a new letter. If 8 % are answering on the web this time the price per extra interview will be 110 DKK which is a bit less than the price for a telephone interview.

3. The Complexities of Multi-Modal Trips: Recommendations for Practice.

Kelly Clifton (Portland State University), T Keith Lawton (Private consultant) and Neba Noyan (Portland State University)

The ability to capture and represent the various multi-modal trip stages is important for a variety of reasons. There is increasing need to understand more about the role of transit access/egress modes to support the “last mile” of these trips. Planners desire this information to understand potential markets for walking/cycling and vehicle sharing services (automobile and bicycling). Non-motorized trips are historically undercounted, particularly when considered as part of a multi-modal trip and information about these trips are important for evaluating non-motorized infrastructure investments and health benefits. Short automobile trips should be represented in vehicle-miles traveled estimates and greenhouse gas emissions calculations. Trip generation for mixed use or infill developments presents challenges with how to account for internal capture from trips made within a site or attracted from adjacent locations. Accurate travel duration and trip lengths depend upon capturing this information

Additionally, the definition of what constitutes a multi-modal trip needs to be addressed, but depends largely on the needs of the end data user. For example, pedestrian planners, urban designers, and health researchers often care about aggregate levels of walking, even very short trips constituting a few blocks, such as walking from a car to the final destination. But these short trips can fall outside the needs of many demand modelers working at a larger scale and collecting this information via survey can significantly increase respondent burden. The use of GPS and other passive monitoring device can aid

in the collection of these short trips, but impose another kind of burden on the respondent and require significant post-processing. Distinguishing between trip tours (multiple activities/trip purposes linked over the course of a day) and multi-modal trips (single purpose activities/trip purposes with multiple modes) also requires some clear rules about how best to represent in the data files. The pressures to make travel surveys useful for purposes beyond travel demand modeling mount the need for standards that consider these issues is important.

This paper will work to develop standards in how to handle these multi-modal trips in travel survey data collection, data structure, augmentation, and analysis. First, based upon experience with various travel surveys, the authors will make recommended approach to practice, with an emphasis on various aspects of data collection. These recommendations will consider what information to best collect directly from the respondent (versus other approaches), how to structure the survey instrument to capture multiple modes, and what information could be included with post-processing data augmentation.

These recommendations will be field tested as a part of a data collection effort in Portland, OR as part of a study of the urban contextual influences of trip generation. This effort specifically aims to understand several travel phenomena associated with multi-modal trips, including: a) mode choices, b) trip length distribution of each trip segment, c) placement of the segment within trip tours, and d) trip purpose.

Then, the paper will provide guidance on how these data can best represented in the data structure of the resulting database that allows for the various needs of end users while maintaining flexibility. Travel surveys tend to represent multi-modal trips in several ways. The primary trip purpose and mode are identified and captured as a single record. The “auxiliary modes” are either a) incorporated into this primary trip record, with less information about their trip length, mode, travel time and relationship to the primary trip purpose, b) included as separate trip records with “change mode” as the trip purpose, but with some ambiguity about the relationship to the primary trip, d) some combination of the a) and b), or c) omitted entirely. There are advantages and disadvantages for each approach for survey costs, respondent burden, and the needs of the end user. The paper will make recommendations for representation that distinguishes between multi-modal trips and legs of a trip tour, clearly represents trip purpose, preserves detailed trip stage travel information, and has flexibility for the end user.

The paper will also touch on two additional aspects related to multi-modal trips: data augmentation and analysis. Data augmentation is often necessary to enrich the travel survey data with archived or new information from outside sources. Here, we will limit our discussion to only those aspects of augmentation that are related to multi-modal trips, such as the inclusion of travel speeds, distances, built environment data, etc. Finally, the paper will provide examples of ways to analyze the information provided in this data set to preserve key information important to the end user. Examples of this include how to account for multi-modal trips in trip rate calculations, accounting for the contributions of

various modes used in trip stages in the calculation of VMT, and how to represent non-motorized activity in household travel.

4. Origin–Destination Household Surveys in Argentina cities: methodological and practical challenges

Leda Pereyra (Proyecto de Transporte Urbano de Buenos Aires (PTUBA)) and Mariela Nerome (Proyecto de Transporte Urbano de Buenos Aires (PTUBA)).

This work is based in the experience that it is being done since the year 2006 in the development of Origin-Destination household surveys in metropolitan areas of Argentinean cities. The objective is to document the methodological and practical challenges that arose from the implemented projects, to show the learned lessons and action alternatives that have been made to solve some problems and to offer ways and guidelines according to the past experience for future implementations of this kind of surveys.

Since the year 2006, under the Urban Transport Project of Buenos Aires (PTUBA) from the National Secretary of Transport, with finance from the World Bank, have initiated actions to developed these kind of surveys in the Buenos Aires metropolitan Area and in other intermediate cities of Argentina, including Córdoba, Mendoza, Posadas, Rosario and Tucumán. Nowadays, PTUBA is also working on implementing these surveys in others cities in the year 2011.

From the point of view of urban and transport planning, these kind of surveys play an important role allowing transport planners and decision makers to know how the population travels and to analyze the accessibility, mobility and transport needs. Furthermore, these surveys are used to develop models that give the possibility to analyze transport policies and projects.

So far, PTUBA has finished the fieldwork and it is working in processing databases, performing analysis and developing reports of the surveys in the cities of Rosario, Buenos Aires, Córdoba and Mendoza. PTUBA is also performing fieldwork in the cities of Posadas and San Miguel de Tucumán.

PTUBA technical team has been doing the follow up and collecting several experiences, confronting methodological challenges due to the complexity of implementing travel household surveys. These surveys involve working on three different units of analysis (households, persons and trips). Such complexity leads to methodological and practical challenges in the design, execution and analysis of the surveys.

There are mainly three kinds of challenges i) the ones related to the sampling process and expansion, ii) the ones related to the development of the questionnaire and iii) the ones related to the field work.

In relation to challenges in the sampling process and expansion, PTUBA has worked in different cities with different methodologies, taking into account the available information. The main cases are summarized below:

- Cases of samples with replacement, using different methodologies, and others cases of samples without replacement where samples are augmented taking into account the estimated non-response rates. The cases of sample with replacement were more flexible allowing more cases with complete surveys, but excessive replacements of cases included sample bias that needed to be corrected later.

- In growing and changing cities, the 2001 National Census data did not accurately revealed the real number of blocks and housing in certain census radios. For this reason, it was required special procedures to list and count houses. It turned out to be an important issue given the sample design required a more updated list of housing.

- In one of the performed surveys, the sampling frame was based on the cadastre, which included the whole lots, including private homes and lots with other uses. This decision was taken because it was considered that the basis of the cadastre was more updated than the National Census. However, this decision introduced several complications and challenges during the fieldwork.

Related with the development of the questionnaires, the length of the forms tends to produce tiredness in respondents. It is important to remind that in this kind of Origin-Destination household surveys, information is collected from households (type of home, income and costs), about persons (educational level, occupation, perception about their city transport and their travels in the previous day (working day) to the one that the survey was done) and trips (origin destination, travel time, etc). For the application of these questionnaires it is necessary direct informants (all members of the household). The repetition of information required about the origin and destination addresses of every travel and its stages, lengthened the process to complete the questionnaire.

The fieldwork also produced important challenges:

- High rotation of the pollsters is generated because it is difficult in many cases to close a full survey within the 3 visits planned. This situation generated interviewers' frustration lengthening the fieldwork as permanent trainings must be done.

- There are a high number of incomplete surveys because some members are hard to contact and because of the resistance from the respondents to give specific information about some chapters of the questionnaire.

Taking into account the experience and the challenges found after the implementation of these projects, this work closes with a methodological proposal that considers the reality of existing data in Argentina, with a sampling methodology that can be replicable in other cities, with a more solid and simplified questionnaire and with appropriate expansion processes. Finally, it considers that the analysis of the information can not only serve the purposes of a transport model but also could serve as an essential tool of analysis of transport planning and design of public policies for our cities, allowing to improve the accessibility conditions for low-income sectors and the rest of the population.