Retrospective Surveys: Some Experiences in the Context of Measuring Lifecycle Events

Marloes Verhoeven, Theo Arentze, Harry Timmermans and Peter van der Waerden
Urban Planning Group, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, phone +31 40 247 2934, fax +31 40 2438488, E-mail: m.verhoeven@bwk.tue.nl

Abstract. Data on lifecycle events are difficult and expensive to collect using (pseudo-) panels. Retrospective surveys which asked respondents to recall events, may constitute a valuable alternative. This paper reports on our experiences with the design and implementation of an Internet-based retrospective survey. Because the value of retrospective surveys will depend on the salience of the events, influencing the memory trace and retrievability of the event, a statistical analysis was conducted to examine the effect of age, cohort and time elapsed between event occurrence and event reporting. Results suggest that although retrospective surveys are not error-free, respondents’ ability to recall major lifecycle events and report some of their key attributes is satisfactory. These findings imply that retrospective surveys constitute a valuable and cheap alternative to collecting data on the sequence of lifetrjectory events.

1. INTRODUCTION

Traditionally, the role of transportation planning has been to plan and design new roads capacity. Models of transport demand were helpful in this regard in the sense that they predicted future demand as a function of demographic processes and planned changes in land use. Many countries and major metropolitan areas therefore collected travel survey data, typically focusing on one or two days, which were assumed to be representative of typical days and typical travel patterns. Knowledge and experiences with the design and administration of such travel surveys has accumulated world-wide over the years.

More recently, however, this traditional task of transportation planning has been augmented. Transport demand forecasting has gradually moved from trip-based via tour-based to activity-based models to deal with additional questions such as transport externalities. It means that the traditional travel survey, asking about the nature of trips, has been gradually replaced by activity and time use diaries, asking more detail about in-home and out-of-home activities. The focus however has remained on one or two-day diaries, rarely expanded into a full week. However, the emphasis of transportation planning especially in developed countries have further shifted from capacity expansion to the formulation of transportation policies aimed at effectively managing travel demand. The aim of such policies is to induce behavioral change: less car use, different departure times, etc. Management programs, information provision and specific incentives have been formulated to induce people to change their behavior. Sometimes, these policies are aimed at the public at large, sometimes they are focused on particular segments of the population. In evaluating these policies, or in assessing the potential impact of such policies, data are required on people’s motivations, tendencies to comply with social norms and information provision, learning, habits and triggers of behavioral change. It goes without saying that these data needs cannot be fulfilled and therefore some of the issues facing planners cannot be addressed by one-day or two day, cross-sectional data. Response rates, sample size, level of detail, etc. do not matter because by their very nature, the typical travel survey

1 Paper for the 87th Annual Meeting of the Transportation Research Board, Washington, DC to ABJ40-Committee on Travel Survey Methods.

Marloes Verhoeven, Theo Arentze, Harry Timmermans and Peter van der Waerden are respectively PhD student, associate professor, professor and research associate.
data do not contain any information about behavioral response and the dynamics underlying travel behavior.

Data needs and appropriate data collection will depend on the specific purpose and planning objectives. One instrument, important in the context of marketing programs is to understand what life trajectory changes trigger behavioral change. Van der Waerden et al. [1], [2] and Klöckner [3] argued that travelers may change their habits due to critical incidents and key lifecycle events. Critical incidents are certain events such as an accident that often cause a highly negative experience such that individuals reconsider their current behavior. In contrast, key lifecycle events are unavoidable (demographic) events, such as reaching the age to have a driver’s license or planned events that occur during a lifecycle (leaving home, getting married, first child, retirement, new job, new house, etc). Such events may lead to changes in available resources and choice options.

The potential relevance of these concepts has been evidenced in a number of largely qualitative studies. For example, Stanbridge et al. [4] (see also [5]) examined the effects of residential relocation and concluded that travel considerations are part of the prompt for the relocation itself, that travel entered the process of searching for a new property; and most importantly that relocation forced or prompted reappraisal of travel options once post-relocation journey experiences were encountered. Chen and Chen [6] applying hazard models to the Puget Sound panel data also found that a change in residential location affected the time allocation and travel patterns of individuals. Other similar evidence has been found by Krizek [7], Prillwitz and Lanzendorf [8], [9], Rocci [10] and Hannes et al. [11]. Lanzendorf [12] examined the impact of the birth of a child and found evidence of changes in transport mode choice decisions.

In addition to these studies showing proof of concept, other studies showed how such information can be incorporated into formal models to predict or assess the impact of life trajectory events on travel resources or travel behavior. To derive an operational model that can be embedded in multi-agent simulations, Verhoeven et al. [13], [14] used a Bayesian network to represent the interdependencies between lifecycle events, resources and activity-travel patterns. Beige and Axhausen [15] used hazard/competing risk models to simulate the length of time intervals between events.

To conduct such analyses, data is required about the key events experienced by individuals and their chronological order. Panel data or pseudo-panel data are not relevant to collect such data, because they typically are not administered to the same respondents for a large number of years. In addition to the costs of administering a panel, attritions rates of 40% are quite standard. Retrospective surveys offer an alternative. Behrens and Del Mistro [16] define retrospective surveys as once-off surveys of individuals which ask respondents to recall past behavioral changes and the events and circumstances surrounding these changes. Survey questions can be both quantitative and qualitative in nature. Examples of small scale applications of such surveys focus on non-standardized interviews, e.g., [16], [17]. The method does not appear to have been applied extensively in practice. For model building purposes, however, one needs larger samples. The question then becomes whether retrospective surveys can be successfully designed and administrated through Internet. Self-completion mail surveys seem less adequate as the protocol for completing the survey is relatively complex. Internet-based surveys have the potential advantage that some help and consistency checking is possible.

The paper reports experiences with the design and administration with such a retrospective survey. It should be articulated from the outset that this was not a methodological study, systematically comparing between decisions in the design and administration of such surveys. The data were collected in the context of the conceptual framework suggested in Verhoeven et al. [13] in an attempt to collect the data needed for the Bayesian network representation of the effect of lifecycle events on travel resources and transport mode choice. Because research on lifecycle events is rapidly increasing, we hope that our experiences may be valuable for others, planning to collect similar data.
2. RETROSPECTIVE SURVEYS

Retrospective surveys are surveys which ask respondents to provide information about events, activities or other phenomena that happened to them in the past. Hence, they are invited to recall such phenomena in retrospect. Behrens and Del Mistro [16] describe the advantages and disadvantages of (quasi-) longitudinal personal travel data collection methods, such as panel surveys, repeated cross sectional surveys, cohort pseudo-panel surveys and retrospective surveys. The potential advantages of the retrospective surveys method are that they (1) are not dependent on the existence of existing datasets, and (2) do not present great administrative complexity and time delay in data collection of panel surveys. On the other hand, because respondents are invited to recall phenomena in the past, it goes without saying that the reliability of retrospective surveys depends fundamentally on the nature of the phenomena about which questions are asked. In general, one may argue that individuals build up memory traces about their experiences. Such memory traces will be stronger for those experiences that are more important to them, for experiences that are more unique, dramatic, etc. Vice versa, memory traces of insignificant experiences will be weak. Moreover, potentially incomplete or inaccurate responses likely occur if the event recalled and the time of recollection are far apart. Assuming that data quality, that is memory recall, is monotonically related to the strength of the memory trace, the quality of retrospective surveys may be sufficient if the retrospective questions are concerned with special, memorable, events or phenomena, especially when the time elapsed between the occurring of the event and the time of the survey is not too far apart.

This logic was supported by Behrens and Del Mistro [16]. They concluded based on their retrospective travel survey experiment that even when considerable time had elapsed since making a behavioral change, respondents did not report uncertainty in their recollection of the number of years that had passed since the change. Especially the follow-up telephonic qualitative interview to explore the reliability of the answers from the surveys and to establish how confident respondents were with their answers indicated this. The explanation for this is that all recalled travel behavior changes were associated with a form of ‘life shock’ or trauma, which are memorable events. On the other hand, Baddeley [18] argued that collecting reliable information about experienced events is difficult; respondents often cannot recall the events accurately, especially all details of the event. In an earlier study, Baddeley [18] concluded that forgetting is not uniform. The information about the most recent experience of an event is likely to be more accurate and reliable than information about earlier occasions. Especially details of a given event are difficult to recall accurately. Some memorable personal events are easier to remember.

Selective memory can play a role when retrospective questions are used. It may happen that respondents in fact ‘do-not-know’ the answer. The respondent can be stimulated to give an adequate response in several ways [19]: (1) encourage the respondent to use personal records and (2) stimulate a more thorough question-answer sequence by using a longer introduction. The researcher should be careful to ensure that the respondent understands the introduction and the question. Special techniques, like ‘time-line follow-back methodology’ and ‘domain-dependent encoding’ of memory, are used to probe the memory of respondents and improve their recall. The latter technique uses extra introductory questions, to bring the respondent back to the situation in which the researcher is interested.

In sum, this brief summary of the literature suggests that more research is needed to better understand the pros and cons of retrospective surveys. This paper is meant to contribute to that evolving literature and addresses the specific question to what extent retrospective surveys are useful and adequate to collect information about lifecycle events.
3. INTERNET-BASED SURVEYS

In principle, retrospective data can be collected through self-administered surveys (mail or internet), face-to-face interviews or telephone interviews. Interviews are more time consuming than self-administered surveys and have certain disadvantages for collecting retrospective data. Recently, Internet-based surveys have become quite popular, compared to surveys distributed by ordinary mail. Besides the low costs compared to mail, allowing faster processing of data and distribution by e-mail an Internet-based survey has some technical advantages which are helpful with respect to collecting retrospective data.

As mentioned by De Leeuw [19], it is important for the designer of the survey to fully understand what happens in a question-answer process [20], [21], [22] to stimulate the respondent to give an adequate response. First, the respondent has to understand the question, that is determine the intended meaning. After that, the respondent has to recall relevant information from his/her memory. Particularly, with retrospective surveys this is a difficult cognitive task. After retrieval from memory a judgement is ‘computed’. After a private judgement is formed in the mind of the respondent, he/she has to communicate the answer to the researcher. In the last step the respondent may want to edit the response before it is finally given, especially with sensitive topics.

Examples of technical advantages of Internet-based surveys include: (1) extra information can be presented in pop-ups to support better understanding of a question, (2) dynamic routing, i.e., a dynamic sequence of questions depending on previous answers, skip irrelevent questions and different question phrasing, is possible, (3) drop down lists can be used to effectively handle pre-coded answers, to decrease process time for respondents and ensure that answers are in a desired format, (4) checking possibilities can be used to make sure that answers are in a required range or format (numeric, number of characters) or to reduce item nonresponse if the question is mandatory. A disadvantage of these checking possibilities is that warning/error messages, to return to the question and correct their “error” before being allowed to proceed to the next question, have been shown to increase respondent frustration and number of respondents prematurely abandoning or terminating of the survey [23].

Furthermore, as Christian notes “The design and visual presentation of survey information, through the manipulation of verbal, numeric, symbolic, and graphical languages, can facilitate respondents’ answering of survey questions and help them “get it right the first time”, [24]. The likelihood of receiving error messages, that web surveyors may use to decrease item nonresponse and to verify that responses are in an acceptable format, can be reduced with helping web respondents. Visual design techniques, to reduce respondent frustration, increase response efficiency and improve the overall survey experience for respondents, can be applied by websurvey designers. Item-nonresponse can be reduced with appropriate data collection methods and a well-designed questionnaire [19]. The role of the interviewer is taken over by a self-administered Internet-based survey and handles the questionnaire logic and flow which makes it easier for the respondent to answer, while the respondent remains the ‘locus of control’ and determines the pace of the interview. The quality of the answer will be improved, because the respondent has more time to understand the question and retrieve and compose an answer. Although these advantages of computer-based questionnaires hold more in general, they arguably are particularly relevant for retrospective surveys given the memory demands these surveys impose.

4. DESIGN AND APPLICATION OF A RETROSPECTIVE INTERNET-BASED SURVEY

In this section, we consider an application of a web-based survey that the authors conducted in 2004 to collect retrospective data about lifecycle events. The aim of the retrospective data is to develop an activity-based model that will include the impact of lifecycle or life trajectory events on transport resources and transport mode choice
decisions. We used a self-administered Internet-based survey to collect retrospective data considering the potential advantages of administering the survey through the internet discussed above.

**Lifecylce events**
A structural lifecycle event is defined as a major event in a person’s life, such as a marriage or move. Based on previous work by van der Waerden et al. [1], [2] a list of seven structural lifecycle events was compiled. These seven events are defined as changes in residential location, in household composition, in work location, in study location, in car possession and availability, in possession of public transport pass, and in household income. Changes in household composition for example are marked by births, marriage, divorce, etc. We believe that these processes are true markers of an unfolding life trajectory. Be it positive or negative, these events are unique and generally are the hallmark of individual’s most important lifecycle decisions. Similarly, even though perhaps less intense, changes in residential location, a new car, change of job, etc. also represent unique events in one’s life. Therefore, we assumed that if respondents can be sufficiently motivated to actively participate in the survey and can handle the technical challenges induced by Internet-surveys, there should be no reason to expect any less reliable results for a retrospective survey than for any other type of survey. We will report our experiences in this regard and test this assumption below.

**Design of the survey**
Respondents were prompted about seven structural lifecycle events in their life. They were asked to indicate whether they experienced each of these events, and, if so, to provide additional information in a matrix about the timing of the event (month and year), the cause of the change (i.e., the specific type of event) that took place and the before and after situation for every change to a maximum of ten changes. Different types of changes involved in an event were defined (Figure 1) so that the respondents will understand what they had to recall. We assumed that the description helped the respondent to recall those specific changes of an event. First the respondents indicated if they experienced a change or not. If they experienced at least one change, the respondent automatically saw the matrix question about that specific event. As an example, Figure 2 represents the matrix question of *change in residential location*. All matrices of the seven lifecycle events were identically structured except that this matrix of *change in residential location* had seven columns, whereas the other matrices had only five. The five columns in the matrices referred to: the timing of the event: (1) month and (2) year, (3) the before and (4) after situation and (5) the cause of the change (i.e., the specific type of event). The two extra columns in this matrix (Figure 2) were included to collect more specific information about their housing: (6) housing type and (7) rent / owned residence.

The respondents could indicate changes to a maximum of ten, from the most recent change to previous changes in the past. We assumed that the information about more than ten changes in the past would be less accurate and reliable. Besides that, changes far back in the past probably don’t influence a person’s (current) travel behavior. Trying to remember is hard work, for that reason the maximum number of changes was set to ten. To stimulate the recall process we started with the most recent change. A mix of open ended and pre-coded questions was used in the retrospective survey. The question about the occurrence of an event was a pre-coded question (Figure 1), which was mandatory. Some questions/columns in the matrix question (Figure 2) were open ended questions. Like before and after situation, the other answers were pre-coded (pull down menu). Only the matrices for the structural lifecycle events housing, work and study had open ended answer space in the columns before and after situation. The matrices for the other structural lifecycle events had only predefined answer categories, thus also for the before and after situation.

The matrix question was not mandatory. If this question matrix was made mandatory the respondents should indicate ten changes (rows) and should answer all questions in each column. This would probably irritate the respondent and result in a lower
response rate of completed surveys. An option would be to add a predefined answer ‘no change’ in each row, but this will require a lot of extra time of the respondent to complete each matrix question, which could result in prematurely termination of the survey.

**Procedure, response rate and sample**

The following four step procedure was adopted: (1) invitation e-mail with research aim including a link to the (2) online application form, (3) e-mail with link to the (4) Internet-based survey. To maximize response rates and motivate respondents to participate, a lottery for gift vouchers among completed surveys was organized.

To collect the required data a convenience sample was drawn. E-mail addresses were collected from a set of colleagues and universities in the Netherlands. Approximately 2500 e-mails were sent with a request to participate and to forward the mail to potential respondents. In total, 939 persons agreed to participate and were emailed the address of the web-based surveys. From this group of persons, 807 started and 710 finished the survey. Since 939 respondents registered for the Internet-based survey and 807 respondents started with the survey, a first-level nonresponse rate was 14%. Second-level nonresponse rate indicates the percentage of respondents that terminated the survey before it was completed. This rate is 12%, thus 710 respondents completed the whole survey.

As for sample composition, 59 percent were males, while 41 percent were females. This is in line with previous findings that Internet-based samples tend to be biased in the sense that males are overrepresented in the group of internet users or more inclined to respond than females. Younger respondents are overrepresented in the sample and the elderly are underrepresented. The use of an Internet-based survey could result in this distribution, since not all elderly people probably possesses a personal computer and have access to an internet connection in contrast to younger people. Respondents with a higher education are overrepresented and respondents with a lower education are underrepresented in the sample. The sampling method, i.e. the way e-mails were distributed, probably caused this over- and underrepresentation.

5. **ANALYSES AND RESULTS**

The performance of our retrospective survey instrument and administration procedure will be assessed in terms of item non-response, error checking and data cleaning, and a statistical analysis of the effect of memory on reporting of events. More specifically, a binary logit model was estimated to predict the probability of reporting an event at a given year in the past.

**Item non-response**

When interpreting item-non response in the context of retrospective surveys, one should realize that respondents may skip questions for a number of reasons: (1) by mistake, (2) refuse to answer or (3) unable to provide a correct answer. This may be caused by a problem in the question-answer process (e.g., by not understanding the question or being unable to retrieve the necessary information), the lack of motivation of the respondent, by the topic of the question (e.g., sensitive issues), or by badly designed surveys. Note that missing data that is caused by the inability to recall the relevant information is of concern here. The matrix questions were not mandatory (see Figure 1), so there exists item nonresponse in these data.

To analyze item non-response we compared the item-nonresponse rate, within each reported change related to an event, of the questions about month, year, before situation, after situation and type of change (Figure 3). The different graphs shown in Figure 3 relate to different events. The categories displayed on the x-axis relate to different questions that were posed for each reported event. Graphs in Figure 3 indicate that in general the item-nonresponse is low. However, the first graph (housing) suggests some substantial variation in item-non response, depending on level of detail. For example, the nonresponse rate for the questions about housing type and bought/rent is higher than the
nonresponse rate for the other questions month, year, before and after situation, and type of change. It suggests that the ease of recalling the various aspects of events differ within and between different types of events and displays a tendency to increase over the number of events reported from most recent to most ancient. It seems that respondents had somewhat more problems recalling the specific information about some events than one would assume indeed are less memorable. Apart from ease of recalling, perceived burden of providing the answers could have played a role. Thus, item-nonresponse exist in our retrospective data, but overall, the percentage is relatively low, and more importantly, item non-response tends to decrease with increased salience of the event. Based on this performance criteria, one can conclude that indeed one does not need to be overconcerned about respondents’ ability and willingness to recall information about the major lifecycle events.

**Error checking and cleaning**

A second aspect concerns the quality of the response. Because errors and inconsistencies exist in all types of data, the key question here is whether the survey evidenced any abnormal amount of inconsistency. To address that question, several error checking possibilities were considered. In the matrix question, errors possibly may appear as: (1) inconsistencies between the before situation of the most recent change (first row) and the after situation of the previous change (second row), and (2) changes were not indicated in the right order, from recent to previous changes. Possible causes for these errors are given and a way to fix these errors is described.

Findings related to item non-reponse have already been reported above. This further analysis is aimed at checking whether the missing information could be corrected based on other information available in the data. Of course, such consistency checks not only applies for respondents with missing information, but for all respondents. Table 1 indicates the number of incomplete cases for each event. The number of incomplete cases that could be corrected are reported in the second row. Ultimately, the maximum number of non-usuable cases for analysis was only 12, which is a small percentage of the total number of cases, 710. It suggests that overall the quality of the surveys is seemed at least satisfactory.

**Memory and recording of events**

As we have argued, a critical factor influencing the usefulness and reliability of retrospective surveys concerns the question how well respondents are able to retrieve from memory the events they experienced in the past. We hypothesized that this will depend on the nature of the event and on time elapsed between moment of recollection (i.e., the moment of completing the survey) and the moment a particular event happened. More specifically, we expect that the probability of a reported event will decrease with increasing time elapsed between the queried period and the current moment (of the survey).

To test this hypothesis, we estimated a binary logit model to predict the probability of reporting an event at a certain year in the past (the queried year). For each respondent and each event category, each year in the past constitutes an observation of whether or not the respondent reported an event of that category for that year. This means, for example, if a respondent could look 15 years back in time regarding a particular event, we have 15 observations for that respondent and event. A univariate analysis where elapsed time is the only explanatory variable in the model is, however, not adequate for testing our hypothesis. With varying elapsed time (history) also the age of the respondent at the queried year varies simultaneously and, obviously, age may have an effect on occurrence probability of an event as well. Therefore, to correct for an age effect we also include the age of the respondent at the moment of the queried period (age-event) as an independent variable in the model. Finally, we include education, gender to correct for demographic attributes.

Although in this way we can correct for an age-at-the-queried-period effect, we should note that there is another possible effect of age. That is, if we vary elapsed time,
while keeping age at the queried period constant, we vary simultaneously the current age of the respondent and with that the cohort group to which the respondent belongs. For example, if older people of today experienced an event less often when they were young compared to younger people when they were young then the elapsed-time variable would have a negative effect on the probability of the event being reported. Clearly, this would not be a memory effect but a cohort group effect. If it were possible to include current age as an explanatory variable in the model as well, then we could correct for this cohort effect. Obviously, however, this would give estimation problems, since a linear relationship exists between elapsed time, age at queried period and current age. In sum, the situation is that we can correct for age at the queried period and confound a memory and cohort effect or we can correct for current age and confound a memory effect with an age-at-queried-period effect in the estimated coefficient for elapsed time. Thus, we should keep in mind that the analysis does not allow us to test our hypothesis in a rigorous sense.

Table 2 represents the results of the binary logit model. If the parameter for history is negative, the probability of a reported event decreases with increasing time elapsed between the queried period and the current moment (of the survey). This could indicate a memory effect (and/or cohort effect). A positive parameter for history means that the probability of a reported event increases with increasing time elapsed between the recall period and the current moment (of the survey). This could indicate that there is no memory effect (and/or cohort effect). As it appears, the history parameter for the events: household composition (household), changes in car possession and availability (car), possession of public transport pass (PT), and household income (income) are all negative and significantly different from zero. For the events Housing (i.e., change in residential location), Work and Study the history parameter is not significantly different from zero. These results indicate that memory may have an effect in reporting of events in case of the household composition, income and transport-mode related events, whereas it does not seem to play a role in case of Housing, Work and Study related events. The negative effect on report probability seems to be largest for the public transport pass events (parameter = -0.090). In case of household, car and income the effect of history is substantially lower than the effect of age-event. Thus, this analysis suggests that changes in Public Transport pass are not that memorable. An alternative cohort explanation would be that, over the years, availability and use of public transport passes have increased (in The Netherlands) and, with that, probability of events related to these passes increased. Keeping in mind the possible confounding of cohort and memory effect, we note that nevertheless these findings are consistent with our assumption that many lifecycle events are relatively easy to retrieve from memory, supporting the potential value of retrospective surveys.

6. CONCLUSIONS AND DISCUSSION

If one wishes to include lifecycle or trajectory events in any analysis or model of activity-travel patterns, data on such events should be collected. Conventional data collection approaches such as (quasi-) longitudinal personal travel data collection methods, including panel surveys, repeated cross sectional surveys, and cohort pseudo-panel surveys are typically not collected as part of national surveys and moreover require substantial financial resources to administer. Potentially, therefore, retrospective surveys, especially administered through the Internet, are a good alternative, but relatively little is known about their performance. One would expect that the quality of data coming from a retrospective survey depends on the nature of the event about which information is collected and on the time elapsed between the occurring of the events and the time of the retrospective survey.

To contribute to that evolving literature, and address the specific question to what extent retrospective surveys are useful and adequate to collect information about life trajectory events, this paper has reported some experiences with the design and administration of an Internet-based retrospective survey. The results support the claim that the data on lifecycle events can be collected with sufficient reliability. This conclusion is
based on three indicators. Item non-response in general was relatively low, especially for those lifecycle events that serve as markers unfolding one’s life. Moreover, inconsistencies in the data were not very different from experiences with traditional travel surveys and most problems could be relatively fixed using other pieces of information in the survey. Finally, a statistical analysis, albeit not capable of avoiding confounding cohort and memory effects, indicated that memory / cohort effects were not found especially for the more salient lifecycle events.

In conclusion, the results of this study support the claim that retrospective surveys are relatively easy, cheap and reliable means of collecting data on lifecycle events. The use of Internet has the additional advantage that technical tools can be used to enhance the memory recall process, prompt respondents for incomplete or unlikely information and avoid coding errors. On the other hand, reliance on Internet has the usual problems of sample bias.
REFERENCES


List of figures and tables

FIGURE 1 Introduction question Occurrence change in Residential Housing.
FIGURE 2 Matrix Question change in Residential Location.
FIGURE 3 Item-nonresponse seven Events.

TABLE 1 Incomplete data year and type of change
TABLE 2 Results Binary Logit Model
CHANGE IN RESIDENTIAL LOCATION

How many changes occurred since you left the parental house?

Changes in residential location are defined as follows:
- moving to the first student room (house in the city where you study, which is not independent)
- move to another studentroom (not within the same house)
- independent living
- living together (move into your partner's house or move together to a different location)
- rent a (different) house
- buy a (different) house
- moving in with parents

Please indicate here whether you experienced this never or at least once.
○ never
○ at least once

powered by NetQuestionnaires

FIGURE 1 Introduction question Occurrence change in Residential Housing.
CHANGE IN RESIDENTIAL LOCATION

- moving to the final student room
- move to another student room (not within the same house)
- independent living
- living together
- rent a (different) house
- buy a (different) house
- moving in with parents

**Month** indicate the month of the recalled occurrence. 'no idea' is also an option
**Year** indicate here the year of the recalled occurrence.
**Before** indicate the street and city of the old residential location.
**After** indicate the street and city of the new residential location.
**Type of change** choose a pre-coded answer.
**Housing type** indicate the type of your new house.
**Bought/Rent** indicate if the new house is rent or bought.

Please answer the questions below for each change in residential location.

<table>
<thead>
<tr>
<th>month</th>
<th>year</th>
<th>before</th>
<th>after</th>
<th>type of change</th>
<th>housing type</th>
<th>bought/rent</th>
</tr>
</thead>
<tbody>
<tr>
<td>most recent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>previous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 2** Matrix Question change in Residential Location.
FIGURE 3 Item-nonresponse seven Events.
TABLE 1 Incomplete data (year and type of change)

<table>
<thead>
<tr>
<th>Incomplete cases</th>
<th>Housing</th>
<th>Household</th>
<th>Work</th>
<th>Study</th>
<th>Car</th>
<th>PT</th>
<th>Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases corrected</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>16</td>
<td>28</td>
<td>16</td>
</tr>
<tr>
<td>Usable cases</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unusable cases</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>history</td>
<td>age-event</td>
<td>education</td>
<td>gender</td>
<td>constant</td>
<td>Chi square</td>
<td>2LogL</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-----------</td>
<td>-----------</td>
<td>--------</td>
<td>----------</td>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Housing</td>
<td>Beta</td>
<td>-0.002</td>
<td>-0.064</td>
<td>0.049</td>
<td>0.103</td>
<td>0.457</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sign</td>
<td>0.343</td>
<td>0.000</td>
<td>0.058</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Household</td>
<td>Beta</td>
<td>-0.014</td>
<td>-0.022</td>
<td>-0.017</td>
<td>0.003</td>
<td>-0.949</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sign</td>
<td>0.000</td>
<td>0.000</td>
<td>0.583</td>
<td>0.933</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Work</td>
<td>Beta</td>
<td>0.003</td>
<td>-0.046</td>
<td>0.027</td>
<td>0.203</td>
<td>-0.124</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sign</td>
<td>0.231</td>
<td>0.000</td>
<td>0.361</td>
<td>0.000</td>
<td>0.266</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Beta</td>
<td>0.011</td>
<td>-0.124</td>
<td>0.068</td>
<td>0.120</td>
<td>0.757</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sign</td>
<td>0.368</td>
<td>0.000</td>
<td>0.122</td>
<td>0.013</td>
<td>0.030</td>
<td></td>
</tr>
<tr>
<td>Car</td>
<td>Beta</td>
<td>-0.014</td>
<td>-0.039</td>
<td>0.001</td>
<td>0.136</td>
<td>-0.637</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sign</td>
<td>0.001</td>
<td>0.000</td>
<td>0.985</td>
<td>0.001</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td>Beta</td>
<td>-0.090</td>
<td>-0.041</td>
<td>-0.057</td>
<td>0.184</td>
<td>-0.482</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sign</td>
<td>0.000</td>
<td>0.000</td>
<td>0.099</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>Beta</td>
<td>-0.010</td>
<td>-0.040</td>
<td>0.104</td>
<td>0.188</td>
<td>-0.429</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sign</td>
<td>0.009</td>
<td>0.000</td>
<td>0.005</td>
<td>0.000</td>
<td>0.001</td>
<td></td>
</tr>
</tbody>
</table>