RETROSPECTIVE DATA COLLECTION IN EUROPE

Mathis Schröder
Axel Börsch-Supan

172-2008
Retrospective Data Collection in Europe

Mathis Schröder
Axel Börsch-Supan

MEA and Mannheim University

December 2008

Abstract

Retrospective data play an important role in social science research. This paper provides an overview of problems arising during the process and also presents the solutions found in the field. It also shows details of the collection of retrospective data in the Survey of Health, Ageing and Retirement in Europe (SHARE) in the SHARELIFE project.

Correspondence:
Mathis Schröder, Mannheim University, L13,17, 68131 Mannheim, Germany
e-mail: schroeder@mea.uni-mannheim.de
1. Introduction

Collecting individual data for social research is one of the essential tasks of the social sciences. Suitable data allows doing policy research, deriving and evaluating welfare programs, and in turn helps to increase the general wellbeing of individuals. Longitudinal data is especially advantageous, when the focus is on changes over time – for example how people move from work into retirement when they become older. However, the collection of individual data is very resource intensive, and a longitudinal effort creates additional problems (e.g. inconsistencies, attrition) that can compromise the data quality.

A less fund-intensive alternative for collecting individual data is to do this in a one-time cross section, but to include a distinct longitudinal component by asking individuals about long stretches of their past. This builds on the fact that even data that is meant to be collected contemporaneously is to some extent retrospective when the person is interviewed (e.g. last year’s income), and so only the time horizon is altered, not the conceptual framework. Nevertheless, retrospective data come at the cost of an increased amount of recall errors, the further back in time interviews go. Researchers in cognitive psychology and survey methodology have developed several strategies and tools that have been shown to significantly reduce recall problems.

This paper motivates the use of a retrospective data collection for the Survey of Health, Ageing, and Retirement in Europe (SHARE)\(^1\). SHARE is a longitudinal study that started in 11 European countries in 2004, asking about 30,000 respondents aged 50 or older about their contemporaneous living situation. A second wave was fielded in 2006, which included two new countries as well as refresher samples to keep the initial sample size. SHARE covers multiple aspects, reaching from economic variables and demographics to health variables. SHARE is an interesting dataset for a retrospective data collection, since all its respondents have a history that spans at least 50 years. Especially to understand how people reach the

\(^1\) See also Börsch Supan et al, 2005, for a detailed discussion of SHARE.
current points in their lives, it is important to know about their pasts. SHARELIFE – the project name – is based on the most recent research on retrospective data collection and contributes to improvements in the field by implementing newly developed tools.

This paper first gives an overview of common findings about retrospective data and the associated quality. These findings are based on data where a comparison is possible, which of course is rarely the case (if the true data are available, a retrospective part is not necessary). Then follows a look at commonly used tools and methods to increase recall precision, as well as an overview of how retrospective data collection is done in other, comparable studies. In the final section, SHARELIFE is described as a tool used to collect retrospective data.

2. Retrospective data quality

In any survey the main issue of answering questions, retrospective or not, is accuracy. Most socio-economic surveys have a retrospective component to them, for example when asking about events that happened in the last year or when they want to know about frequencies of events. That there is room to improve accuracy in retrospective surveys has been shown in many studies across different subjects. In the following, several examples of known issues are illustrated and briefly discussed.

Peters (1988) uses the National Longitudinal Survey of Work Experience (NLS), a panel that started in 1966. As marital status – and thus changes in it – is recorded each year, she can construct a marriage history for her study group, which are young women. In 1978 and in 1983, the women are asked about their last three marriages, which gives the opportunity to compare the retrospectively collected data with the ongoing records of marital history. Taking the “at-the-time” recordings as the true data, Peters finds a concordance of between the panel and the retrospective data of 76%. The underlying recall period is at the maximum fifteen years, with a mean distance between interview and event of slightly more than five years.
When looking at the determinants of this recall error, she finds that the recall is reduced by 3-4% per year. In addition, education is reducing recall error.

Auriat (2000) studies moving histories. The retrospective data source is a survey of couples in Belgium, who are asked separately and as a couple about their residential changes from when they were 14 years old. As they are between 41 and 57 years old at the time of the interview, there is a potentially large number of moves to report. The comparison data come from the Belgium National Population Register, where all moves have to be reported within 8 days of changing the address. Auriat sets the error margin to three months: if the reported move date is within three months of the official move date, there is no error. Considering at most the first three moves after marriage, she finds that at least 30% are off by more than 3 months. Quite interestingly, however, she finds that if the reports are wrong, respondents miss the mark by yearly amounts, i.e. the difference to the true date is by multiples of 12 months rather than by any other difference. This is attributed to memories of the season, in which the move happened. She finds little evidence that time since event matters, which might be an effect of the moves all being long ago.

Mathiowitz and Duncan (1988) use the 1983 PSID to assess the accuracy of retrospective unemployment information. The respondents are asked about their monthly unemployment history in the previous two calendar years. From this, an indicator of “any unemployment” is generated as well as the detailed unemployment history up to the interview. The employment records of a large firm are used to validate this information. A look at the indicator yields corresponding answers in 89% of the data – i.e., 89% of respondents indeed had an unemployment period if they reported that in the PSID and vice versa. However, the duration of unemployment is often misreported: specifically spells with a short duration of 12 weeks or less are omitted by the respondents in more than 50% of the time, and still 37% of spells that last 29 weeks or longer are not reported. Interestingly, time since the event is not related to the recall error, maybe an effect of the short time period considered here.
A study related to health by Means et al (1989) was done for the US department of Health and Human Services. They collect data on doctor visits for any health condition over the last 12 months, and compare the reported visits to those in the actual medical records. Only about 41% of all visits are recalled, and, interestingly, the seriousness of the condition does not matter in terms of recall. Recurring events, i.e. those where the respondent had to come back for the same condition, were recalled with less precision: visits for the same condition that led to three or more visits are each less accurately recalled than less frequent ones, with a margin of 21 percentage points, or 32 vs. 53%.

There has also been an effort to test the memory of public events, done by Gaskell et al (2000). In this study, people were asked to remember two public events: Margaret Thatcher’s resignation, which at the time of the interview had been 19 months ago, and the Hillsborough football disaster, 37 at the time of the interview months ago. Correct responses were relatively few, only 15% of respondents recalled the exact year and month of the resignation, whereas 10% were able to exactly date the football disaster. Given the previous examples, this hints at the possibility, that public events are less accurately remembered than events in respondents’ personal histories.

These examples show that there is evidence of recall bias in various topics, although there are differences among the topics in terms of how easy things are forgotten. Personal moves are much better recalled than visits to the doctor, for example, especially when considering the differences in time horizon in these studies. This leads to the very intuitive point that events that are more important in a respondent’s life are more easily remembered. This hypothesis of a saliency effect is tested in two studies by Akerlof and Yellen (1985) and Juerges (2007), who both look at the correct report of unemployment spells that had happened a year prior to the interview, dependent on the impact this had at the time on the respondent’s life. While Akerlof and Yellen measure this indirectly, Juerges uses the reported life satisfaction to explain recall bias. Both studies find that the more important the event was at
the time, the more accurate it is remembered later. However, this is of little practical purpose, as it is unknown to the researcher, what are the salient events in a person’s life. In the following section, measured to improve recall are discussed.

3. Measures to improve retrospective data

Before considering how to improve retrospective data practically, one must think about how memory functions. Research in cognitive psychology has proposed four types of autobiographical memory (see Conway, 1996): episodic memories, which are memories that for a particular memory; extended special events, for a well defined period of time; generic memories, for recurring events; and lifetime periods, the major periods in one’s life. In retrospective data collection we are not interested in the episodic memories, but mainly in extended special events that lie within lifetime periods.

There are several intuitive ways to improve the recall of an event itself, which have been shown to make a significant difference in recall. Several studies have shown that remembering events is more likely the longer people have time to think about the question. For example, Cannel et al (1977) report that adding meaningless parts to the question increases accuracy at least for educated respondents (with no significant effects for less educated people). Similarly, Loftus et al. (1990) reduce the number of misreports by asking the same question twice and varying the reference period. Explicitly asking for more effort from the respondent has been shown to increase recall as well (Cannel et al, 1981). However, these approaches are not necessarily suitable for a project like SHARELIFE – interviewers would feel awkward to ask the same question twice for example, and respondents would wonder why a certain question is said to be more important than another one. In addition it is hard to judge whether an event did not happen or the respondent does not remember it, and thus the countermeasures are hard to implement.
If the event in question is remembered, dating accuracy is most important. This is also important for SHARELIFE, as we are interested in precise data. Again, there are some intuitive ways, but also some more elaborate possibilities have been developed in the field. Immediately obvious is the restriction to time periods that are recent – stemming from the results that events that are further back in time are usually less accurately remembered than things that happened more recently. Findings on changing the recall order (i.e. backward, forward or “free recall”) are mixed and suggest that the effects of recall order depend on the subject. For example, Jobe et al. (1990), find that free recall provides more accurate results when asking about household health care visits, whereas Loftus et al. (1992) find no significant effect for recall order when asking about HMO visits. Another option is “bounded recall” – an artificial restriction of the reference period has been shown to increase accuracy (e.g. Neter and Waksberg, 1964; or Auriat, 1993). Bounded recall can be especially useful in a panel study, where the time between two interviews is naturally defined as the reference period. In the second interview it is possible to remind the respondent of an answer given last time, and then ask her to report any changes that happened since. But the reference period can also be restricted to other periods of the respondent’s life – for instance to lifetime periods or extended events in the terminology used above.

Another form of bounded recall is the use of “temporal landmarks”, which hinges on the idea that there are certain events in one’s life that are outstanding – personal events such as the birth of a child or marriage, but also some public events, such as the assassination of J. F. Kennedy or the win of one’s national team in a big sports competition. These events, once they are known, can be used to anchor the respondent’s memory and place other personal events relatively to the landmark: the respondent might not know when exactly something happened, but if she knows it was in the year when people landed on the moon, this information can help.
Several studies have tested the use of public landmarks. For example, Loftus and Marburger (1983) use the eruption of Mt. St. Helens in Washington State. Six months after this event, they ask about crime victimization, using either “in the last six months” or “since the eruption of St. Helens” as an entry to the question. They find that the inclusion of landmarks increases the accuracy of the reports.

Besides changing the type or the content of questions, graphical devices have been shown to improve data quality as well. The simplest way here is the use of timelines, where the respondent is asked to record life events on an axis, and then place other events accordingly around it. This one-dimensional version has been extended to a life grid or calendar with the use of Event History Calendars (EHC), as described by Freedman et al. (1988) or Belli (1998). The idea is similar to that of timelines, just on multiple dimensions. When going through a questionnaire, life events are recorded into a large grid, where a set of topics such as children, partners, or work are combined with the time dimension, which is usually on the horizontal. The calendar then allows the respondent to see important events of different areas of her life in parallel. Belli (1998) argues that the EHC enhances the respondent’s ability to recall, as the standard recall mechanisms are triggered by this approach. These mechanisms or associations are threefold, all of which are supported by the EHC:

1. Top-down retrieval, meaning that a higher order structure indexing allows moving into lower order structures, or similar, the memory moves from general structures to specific ones.

2. Within a theme, events are ordered and therefore recalled sequentially.

3. Across themes, recall happens in parallel, meaning that one event can trigger the memory of a different theme because it happened in the same episode.

As Schwarz and Oyserman (2001) suggest, EHCs improve recall by making use of several of the other approaches mentioned above. One clearly is the use of landmarks, as during the
process of filling the calendar, the interviewer can always prompt by using previously entered events, for example the birth of children. But also public landmarks can be used very easily, if the interviewer has a possibility to refer to them. The calendar is in principle also open in terms of the order in which it is filled – as long as the questions are asked flexible enough, the topics do not need to be followed in a specific order. An additional feature is that the calendar allows both the respondent and the interviewer to easily cross-check events and correct errors that otherwise would have remained undetected.

4. Earlier Implementations of Event History Calendars

The first EHC has been implemented by Freedman et al. (1988) in a pen and paper version. They used this technique on a sample of nine hundred 23 year olds, who were asked in detail about life events that had happened since their 15th birthday. Events were entered on a monthly basis, so overall, 120 entries were possible for each of the possible categories. Since this was done on actual paper, even though a lot of categories were included, the scope of the questionnaire accompanying the calendar was rather limited.

Belli et al. (2005) experimented with a life history for the Panel Study of Income Dynamics (PSID). The PSID meant to test the life history approach to see if recall could be improved with this technique compared to a regular q-list interview. The implementation was a computerized instrument done by a telephone interview, so the respondent did not have access to the calendar, but the calendar rather served as an input device for the interviewer. However, the interview itself was very flexible: Instead of asking direct questions, the interviewers were asked to simply give broad questions that would lead to the topics the researchers were interested in. This would guarantee that the respondent truly remembered events. However, this approach is very limited if the researcher is interested in things the respondent cannot be “hinted” at.
The English Longitudinal Study of Ageing (ELSA) implemented a face-to-face interview with an EHC in their study in 2007. As ELSA is very closely related to SHARE, the life history approach serves as the main role model for SHARELIFE. The main reason for ELSA to use a life history approach is the sample: since all of the respondents are at least 50 years old, there are long stretches of time the respondents were not observed in. Especially for health and socio-economic status, however, researchers would be very interested in the respondents past to relate it to the present. In this regard, the life history approach was meant to collect very important unobserved variables.

ELSA covered several areas of retrospective data: health, economics and social from early childhood, which were followed then by demographic, economic and health experiences through adulthood. A lot of development went into the implementation of the instrument, which started out as a very flexible type of interview, similar to the PSID experiment done by Belli et al (2001). However, throughout the course of the pre-tests and pilot studies, the study moved from a flexible pen-and-paper combined interview to a more standard way of questioning with a life grid as a supplement to the interview. It is this computerized version of the life history interview that serves as the basis for the SHARELIFE interview described in detail below.

5. Implementation of SHARELIFE

The decision of how to implement a life history interview within the framework of the SHARE project was based on the specific requirements and the research ideas behind them. SHARELIFE is meant to analyze the European welfare state by comparing individual decisions across time and countries and connecting these decisions with the institutional surroundings that people faced at the time. This implies a somewhat standardized interview which guarantees the comparability not only within a country but also across the European
boundaries. In addition, the face-to-face interviews from the first two waves of SHARE were meant to be continued. These considerations led to the decision to not use a fully flexible approach as taken by Belli et al. (2005) but rather base the life history interview on an approach similar to ELSA. This has the additional benefit that comparability with this very similar study is given.

There are several different modules in the current SHARELIFE interview, which are ordered according to what is most important and thus remembered most accurately in the respondent’s life. Although this order is based on the pilot studies done by ELSA, a flexible approach is allowed in the sense that the order can easily be changed if the respondent demands this. The most important part for the interview is to keep a balance between length and content, as a lot of things are scientifically interesting but each of them adds a little bit to the respondent burden.

The interview starts in default order with questions about the children, like birth year of the child, his or her name, gender, etc. Immediately, this information appears in the calendar for both the respondent and the interviewer to see, so that the interviewer has an easy way of linking questions to personal events (in this case, children). After the child section follows the module about partners, which asks about marriages, cohabitating partners as well as about other important relationships. Again, the main information like the start and end of a relationship is displayed on the screen. The places of living are recorded in the following section, where in the pilot studies so far, the previous life events proved very helpful, as the interviewers can prompt with the information they already have, e.g. “Did you live there after your second child was born?” or “Were you still with X when you moved?”. This anchoring gives tremendous help to the respondent.

This is followed by a section about the respondent’s living situation when he or she was ten years old. This detailed look at one point in the respondent’s childhood provides useful information about where our respondents come from, as some of these variables prove to be
good predictors of socio-economic status later in life. None of the parts of this section appear in the calendar, since the information is too detailed and only concerns this one point in time. This is different for the work section, which follows. This is one of the very detailed sections, where questions are asked about the respondent’s job and retirement history. Not only jobs are covered in detail, but also any periods of not working, be it due to unemployment, maternity or retirement. The work module is followed by a brief section about the use of financial assets during the respondent’s life, where mainly the entry points are of interest here.

The next two sections cover health, where the first one is about health status as a child and as an adult, and the second is about health care with a strong focus on the use of preventive medicine in a respondent’s life. These sections allow identifying differences in the actual health and health care use throughout Europe, which are important determinants of the welfare state. The final section of the interview is then covering general life events, where the respondents are asked to identify specific periods of their live, e.g. when they were happier or when they had to endure financial hardship. While the other sections have in some form been implemented in previous surveys, this section is new to such an instrument. The research interest here is on whether one can link the events already asked about to the directly reported well-being.

Besides the content of the instrument through the described modules, other aspects of the program have to be taken into account as well. The interface that the interviewer uses to go through the questionnaire is important for the success of the whole project. Figure 1 shows a screenshot of the instrument, which has been filled in already.
The screen is divided into six parts. On top is the question part, where the current text the interviewer is supposed to read out is placed in black font while interviewer instructions are set in bold and blue font. Below are the answer categories that are possible for this question – this of course is empty if there is text to be entered. The answers can either be clicked on with the mouse, or they can be typed into the answer line below the answer categories, where also open questions are answered. Below the answer line is the main new feature of the instrument, the calendar. As can be seen in the screenshot, several personal events already have already been entered for this respondent, for example 4 children in the first line, or several different places of living in the third. Below the calendar are two fields – one covering personal events on the right hand side, the other labelled “find events” on the left hand side. The personal events field is filled while going through the questionnaire, and depending on the year that is
Currently looked at in the calendar, the events listed in there change. The events list corresponds directly to the filled squares in the calendar and serves as a tool to get more details about the event in the calendar.

The field in the lower left of the screen is for the interviewer to use the idea that world history events may help the respondent to remember the timing of own personal events. This search tool is a completely new development for SHARELIFE, and may provide a vast improvement over previous usages of general events. Before this implementation, the respondents had a list of events for each year, but here the problem was that either the interviewer or the respondent had to have the knowledge of when an event had happened for these lists to be useful. For example, if the respondent remembered that her move was a month after J. F. Kennedy was shot, then this does not help if neither she nor the interviewer knew when that happened. So this led to searches for events, which cost time and was at times frustrating for both the interviewer and the respondent. The new search tool provides a solution for this, by allowing the interviewer to search for a specific event, as shown in Figure 2:

![Figure 2: Search Tool Screenshot](image)

Figure 2: Search Tool Screenshot
Here the interviewer entered “kenned” and a list of events that contain this string appears in the box. The current example shows two events, the inauguration and the assassination of J. F. Kennedy. As this part is newly developed, there are no data on how this tool performs in the field as of now.

6. Summary

This brief overview showed the problems arising when collecting retrospective data. We also review how these issues are addressed in the literature in general, and then describe the approach SHARELIFE, a collection of retrospective life histories in Europe, has taken. Further research on the current methods as well as their thorough evaluation will be an essential part of our future work in SHARELIFE.
References


# Discussion Paper Series

Mannheim Research Institute for the Economics of Aging Universität Mannheim

To order copies, please direct your request to the author of the title in question.

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Autoren</th>
<th>Titel</th>
<th>Jahr</th>
</tr>
</thead>
<tbody>
<tr>
<td>161-08</td>
<td>Karsten Hank</td>
<td>Generationenbeziehungen im alternden Europa</td>
<td>08</td>
</tr>
<tr>
<td>162-08</td>
<td>Axel Börsch-Supan, Karsten Hank, Hendrik Jürges, Mathis Schröder</td>
<td>Longitudinal Data Collection in Continental Europe: Experiences from the Survey of Health, Ageing and Retirement in (SHARE)</td>
<td>08</td>
</tr>
<tr>
<td>163-08</td>
<td>Martin Salm</td>
<td>Job loss does not cause ill health</td>
<td>08</td>
</tr>
<tr>
<td>164-08</td>
<td>Martin Salm, Daniel Schunk</td>
<td>The role of childhood health for the inter-generational transmission of human capital: Evidence from administrative data</td>
<td>08</td>
</tr>
<tr>
<td>165-08</td>
<td>Christina Benita Wilke</td>
<td>On the feasibility of notional defined contribution systems: The German case</td>
<td>08</td>
</tr>
<tr>
<td>166-08</td>
<td>Alexander Ludwig, Michael Reiter</td>
<td>Sharing Demographic Risk – Who is Afraid of the Baby Bust?</td>
<td>08</td>
</tr>
<tr>
<td>167-08</td>
<td>Jürgen Maurer, André Meier</td>
<td>Smooth it Like the “Joneses?” Estimating Peer-Group Effects in Intertemporal Consumption Choice</td>
<td>08</td>
</tr>
<tr>
<td>168-08</td>
<td>Melanie Lührmann, Jürgen Masurer</td>
<td>Who wears the trousers? A semiparametric analysis of decision power in couples</td>
<td>08</td>
</tr>
<tr>
<td>170-08</td>
<td>Jürgen Maurer</td>
<td>Who has a clue to preventing the flu? Unravelling supply and demand effects on the take-up of influenza vaccinations</td>
<td>08</td>
</tr>
<tr>
<td>171-08</td>
<td>Johannes Binswanger, Daniel Schunk</td>
<td>What Is an Adequate Standard of Living during Retirement?</td>
<td>08</td>
</tr>
<tr>
<td>172-08</td>
<td>Mathis Schröder, Axel Börsch-Supan</td>
<td>Retrospective Data Collection in Europe</td>
<td>08</td>
</tr>
</tbody>
</table>