Reconstructing event histories in standardized survey research:

Cognitive mechanisms and aided recall techniques
The Netherlands Organization of Scientific Research (NWO) is gratefully acknowledged for funding this project titled “Reconstructing event histories in standardized survey research: cognitive mechanisms and aided recall techniques” (Grant Number 400-03-331).
Reconstructing event histories in standardized survey research:

Cognitive mechanisms and aided recall techniques

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan de Vrije Universiteit Amsterdam, op gezag van de rector magnificus prof.dr. L.M. Bouter, in het openbaar te verdedigen ten overstaan van de promotiecommissie van de faculteit der Sociale Wetenschappen op vrijdag 29 april 2011 om 11.45 uur in de aula van de universiteit, De Boelelaan 1105

door

Tina Judith Glasner

geboren te Hadamar, Duitsland
promotor: Prof. dr. W. Dijkstra
copromotor: dr. W. van der Vaart
Acknowledgements

There are many people without whom I could not have written this dissertation.

First of all, I would like to thank my advisors Wil Dijkstra and Wander van der Vaart. Over the past six years, both have devoted countless hours to this project and have provided me with so much valuable feedback on my work. I will never forget our marathon dissertation sessions at various offices, homes, and even boats. Thank you also, Wil, for making me switch to Mac. I have seldom looked back.

When times got rough, I had colleagues who never failed to cheer me up (or at least distract me from my work). An honorable mention goes to my ‘perpetual roommate’ dr. Maarten Buis who introduced me to the wondrous world of PhD comics and made me feel less alone. I would also like to thank Hebe Verrest, Peter Lugtig, Ger Snijkers and all my other former and current colleagues at Utrecht University for their listening ear and their willingness to look at hundreds of blurry baby pictures. Special thanks also to Joop Hox and Edith de Leeuw for their endless patience and support, as well as for the chocolates.

I am greatly indebted to Annette Scherpenzeel, Corrie Vis, Josette Janssen, Marije Oudejans, Miquelle Marchand and many others at CentERdata, not only because without them I would not have any data to write about, but also for being fabulous in general. Thank you also to Bas Weerman at RAND, who made the web-based calendar instrument happen.

My dissertation committee, Bob Belli, Marcel Das, Yfke Ongena, Jan Smit, and Hans van der Zouwen provided very interesting and useful comments on the dissertation and helped me locate many typographical errors and other oddities in this book. Thank you.

Last but very certainly not least I would like to thank my friends, family (including the in-laws) and especially my wife Anna and my little daughter Maya for everything they have put up with over the last (couple of) year(s). You rock. I promise I will never write a dissertation again.
Contents

INTRODUCTION:

CALENDAR INSTRUMENTS IN RETROSPECTIVE SURVEYS 11

1.1. AREA OF APPLICATION: LIFE COURSE RESEARCH 12
1.2. RECALL ERROR IN SURVEYS 14
   1.2.1. Omission 14
   1.2.2. Dating error 16
1.3. AIDED RECALL TECHNIQUES IN SURVEYS 21
1.4. CALENDAR METHODS 24
1.5. RESEARCH QUESTIONS AND APPROACH 26

CHAPTER 2: AUTOBIOGRAPHICAL MEMORY 29

2.1. AUTOBIOGRAPHICAL MEMORY 29
   2.1.1. Representation and organization of events in autobiographical memory 30
   2.1.2. Life scripts 34
2.2. MEMORY FOR TIME 37
   2.2.1. TIME IN AUTOBIOGRAPHICAL MEMORY 37
      2.2.2. Reconstructing dates and durations 38
      2.2.3. Reconstruction strategies and dating accuracy 40
2.3. SUMMARY AND CONCLUSION 41

CHAPTER 3: APPLICATION AND EFFECTS OF CALENDAR INSTRUMENTS IN SOCIAL SURVEYS 43

3.1. INTRODUCTION 43
3.2. CALENDAR INSTRUMENTS: APPLICATIONS AND RATIONALE 44
   3.2.1. Overview: different names, similar instruments 44
   3.2.2. Theoretical rationale of calendar instruments in retrospective surveys 46
3.3. DESIGN FEATURES OF THE CALENDAR 48
   3.3.1. Time dimension 48
   3.3.2. Domain grid 49
   3.3.3. Landmark events 50
3.4. POTENTIAL CHALLENGES IN CROSS-CULTURAL SURVEYS 52
3.5. Experiences with different modes of data collection 54
3.6. Effects of calendar methods on data quality 56
   3.6.1. Indirect comparisons between calendar data and regular survey data 57
   3.6.2. Agreement between calendar data and external sources 59
3.7. Operational costs of calendar procedures: fieldwork and sampling 63
   3.7.1. Interview time and data entry 63
   3.7.2. Interviewer training 64
   3.7.3. Non-response 64
3.8. Evaluations of the interviewing process 66
3.9. Conclusion and discussion 67

CHAPTER 4: CALENDAR INSTRUMENTS AND THE RETRIEVAL OF EVENTS FROM AUTOBIOGRAPHICAL MEMORY 71

4.1. Introduction 71
4.2. Aim of the study 72
4.3. Method 73
   4.3.1. Sample 74
   4.3.2. Question-list 75
   4.3.3. Interviewing method 76
   4.3.4. Coding scheme and method of analysis 77
4.4. Results 82
   4.4.1. Retrieval of time-tagged events 83
   4.4.2. Event order cues 84
   4.4.3. Retrieval strategies based on event details, guessing, and temporal relationships 87
   4.4.4. Visual cues 87
   4.4.5. Data revision 88
4.5. Conclusions and Discussion 88
4.6. Limitations and suggestions for future research 89
CHAPTER 5: EVALUATION OF A CALENDAR RECALL AID IN A WEB-BASED LIFE COURSE SURVEY

5.1. Designing a calendar instrument for an online panel

5.2. Pilot study: Design of the questionnaire and calendar recall aid
   5.2.1. Sample
   5.2.2. Method
   5.2.3. Results
   5.2.4. Conclusion and recommendations

5.3. Study aim and hypotheses
   5.3.1. Break-off rates
   5.3.2. Completeness of retrospective reports
   5.3.3. Data consistency
   5.3.4. Interview duration
   5.3.5. Data revision
   5.3.6. Respondent evaluations

5.4. Method
   5.4.1. Experimental design
   5.4.2. Questionnaire
   5.4.3. Sample
   5.4.4. Data and operationalizations
   5.4.5. Analyses

5.5. Results
   5.5.1. Break-off rates
   5.5.2. Completeness of retrospective reports
   5.5.3. Data consistency
   5.5.4. Interview duration
   5.5.5. Data revision
   5.5.6. Respondent evaluations

5.6. Conclusions
CHAPTER 6: CONCLUSION AND DISCUSSION

6.1. SUMMARY OF FINDINGS
6.2. EFFECTS OF LANDMARK EVENTS AND VISUAL FEEDBACK GRID
6.3. RECOMMENDATIONS FOR THE USE OF VISUAL FEEDBACK AND LANDMARK EVENTS IN WEB-BASED SURVEYS
6.4. CALENDAR INSTRUMENTS AND THE SURVEY RESPONSE PROCESS
6.5. FINAL REMARKS AND DIRECTIONS FOR FUTURE RESEARCH

REFERENCES

APPENDICES

APPENDIX 4.1: RETROSPECTIVE QUESTIONNAIRE USED IN COGNITIVE INTERVIEWS
APPENDIX 4.2: CALENDAR USED ‘QUESTION-LIST PLUS CALENDAR’ CONDITION
APPENDIX 5.1: LIFE COURSE QUESTIONNAIRE USED IN WEB INTERVIEWS

SAMENVATTING
Introduction:

Calendar instruments in retrospective surveys

Researchers in many scientific disciplines use information on past behavior and the life events of individuals in their empirical research. Often, this information on past events is collected retrospectively by means of (semi) standardized survey questionnaires. The problem with retrospective self reports, however, is that their quality can be compromised by recall errors such as omissions, dating error, and biased retrieval (Haandrikman, Rajeswari, Hutter, and Ramesh, 2004; Schwarz & Sudman, 1994; Van der Vaart, 1996).

Supported by theories from cognitive psychology, a variety of aided recall techniques have been developed in order to address this problem. The current study focuses on the use and effects of timeline-based visual recall aids, or calendar instruments, which are based on the idea that retrospective questions about certain types of events or behaviors might be easier to answer, if the respondent can relate the timing of those events to other events (e.g. “I applied for that position shortly after our second child was born”). In recent years, calendars have been integrated into large-scale, longitudinal social surveys such as the German Life History Study (GLHS), the Survey of Health, Ageing and Retirement in Europe (SHARE), and the Panel Study of Income Dynamics (PSID). Previous research indicates that the application of these aided recall properties may have beneficial effects on data quality (Belli, Lee, Stafford, & Chou, 2004; Belli, Shay, & Stafford, 2001, Van der Vaart, 1996). Despite the growing interest in calendar methods, which has led to the development of a great variety of instruments, little is known about their effectiveness in reducing recall error, and most applications still seem to be characterized by a modest theoretical foundation.

This dissertation will explore the relationship between design features of calendar recall aids, cognitive processes involved in the retrieval of past behaviors, and the quality of the retrospective reports. The aim of the dissertation is to provide the heretofore ‘missing link’ between the Event History Calendar’s theoretical rationale and the research practice and to develop an optimized calendar recall aid, which can be used in web-based survey research.
1.1. Area of application: Life course research

One of the main areas of application for calendar instruments is life course research, which is apparent from the names of some of the most well known calendar instruments, such as the Life History Calendar (Freedman, Thornton, Camburn, Alwin, & Young-DeMarco, 1988) Life Events Calendar (Hoppin, Tolbert, Flagg, Blair, & Zahm, 1998), or Event History Calendar (Belli, 1998). In recent years, the life course perspective has become increasingly popular in a number of different scientific disciplines. Sociologists Giele and Elder (1998) describe the life course as “a sequence of socially defined events and roles that the individual enacts over time” (p. 22). Rather than focusing on isolated events or episodes in a person’s life, life course research provides scientists with information on the influence of past historical and biographical events or states on present or future life outcomes. Different varieties of life course research can be found in the social, behavioral and medical sciences and include life span psychology as well as life course sociology, epidemiology, and criminology. Sociological life course studies often focus on specific stages (e.g. childhood, young adulthood) and the timing of events (e.g. marriage, parenthood) in people’s lives. In epidemiology, life course research is often concerned with causal links between (early life) risk factors and later health outcomes (Kuh, Ben-Shlomo, Lynch, Hallqvist, & Power, 2003). The basic measurement concepts in life course research are transitions and trajectories (Elder, 1985). Transitions can be defined as “changes in state that are discrete and have an identifiable beginning and end” (Daaleman & Elder, 2007), such as marriage, divorce, entry to (un)employment, or changes in health status. Trajectories consist of series of transitions and represent a long time perspective on a certain aspect of life, such as work, relationships, health, or psychological state.

There are a number of different ways in which researchers can collect data on those transitions and trajectories. The most frequently used method to obtain the information required for life course research is to reconstruct past events by means of surveys. In most studies, interviewers administer (semi-)standardized questionnaires, either in person or using computer assisted telephone interviewing. In some cases, especially when sensitive information such as sexual histories or past drug use is involved, researchers might choose to use self-administered questionnaires.
In conducting life course surveys, one of the most important distinctions is between prospective versus retrospective data collection. Prospective life course surveys usually have a longitudinal design, in which respondents take part in multiple waves of a panel study. There are, for example, several ongoing birth cohort studies in the United Kingdom, in which data from the same sample has been collected from birth (e.g. the MRC National Survey of Health and Development, see Wadsworth & Kuh, 1997) and continues to be collected throughout the respondents’ lives. Retrospective life course surveys, on the other hand, can be single-round studies, in which the researchers collect data about a very long reference period (e.g. Ingersoll-Dayton, Krause, & Morgan, 2002; Wittebrood & Nieuwbeerta, 2000). In other cases, researchers combine retrospective and prospective data collection. At the start of new longitudinal panel studies, respondents are often asked to provide information on, for instance, employment and residential history. In a 1996 survey that was conducted as part of the German Life History Study, for example, Eastern German respondents answered retrospective questions about their lives before and after the country’s re-unification in 1989 (Reimer & Matthes, 2007).

Retrospective studies are sometimes regarded as inferior to study designs in which life course data is collected prospectively, as retrospective data is assumed to be more biased by recall error. However, prospective cohort studies also have significant drawbacks. Firstly, they are very time-consuming, since the researcher might have to follow a birth cohort for many years before trajectory data becomes available. Secondly, (selective) sample attrition can be a serious threat to the validity of panel data (Wadsworth et al., 2003).

In some cases, especially in the medical sciences, using a prospective study design would require a very long time frame and/or large sample size and is therefore not a viable option. The problem is very common in epidemiological studies of rare and/or newly discovered diseases. Shadish, Cook, & Campbell (2002) argue that for prospective population studies of rare diseases (i.e. diseases with a lifetime prevalence of less than 1 in 1,500 persons) very large sample sizes would be needed in order to find enough cases. For that reason, medical researchers often turn to quasi-experimental research designs, in which groups of patients with a specific disease are compared to a healthy control subjects, the case control study. Using this design, life course data have to be collected retrospectively after the subject has been diagnosed with the disease.
Apart from the discussed ‘true’ retrospective designs, there are often retrospective aspects to prospective panel designs. Even in prospective studies, researchers cannot monitor their subjects’ lives in a way that allows them to capture data in ‘real time’. By periodically asking panel members to report (changes in) their living situation, the retrospective component of the data collection process is not completely eliminated, but reduced to a shorter recall period i.e. the time between the previous and the present wave of the survey (Belli & Callegaro, 2009; Caspi et al., 1996). In many cases, such as the Panel Study of Income Dynamics, the reference period of the retrospective questions in prospective studies covers only a few years, although it can be substantially longer in longitudinal surveys that follow up with respondents less frequently.

1.2. Recall error in surveys

Despite differences in length of the reference period, both prospective and retrospective data collection can elicit inaccurate reports due to recall error (e.g. Mathiowetz, 1986; Reimer, 2004; Schwarz, 2005; Van der Vaart, Van der Zouwen, & Dijkstra, 1995). The most common types of recall error, omission of events and temporal displacement of events as well as their respective causes can lead to substantial underreporting and temporal bias in the data.

1.2.1. Omission

The first error, omission, occurs when a respondent forgets or otherwise fails to report entire events, episodes, or transitions. Schacter (1999) attributes omissions to three general causes, i.e. forgetting (which he calls ‘transience’), blocking, and absentmindedness. Whether forgetting is due to actual loss of information over time, or to access failure remains an unresolved issue in the cognitive literature. In a survey situation, however, information that is very difficult to access can be just as useless as information that is truly lost from memory (Bradburn, Rips, & Shevell, 1987). If the respondent cannot retrieve the information during the course of the interview, it is lost for the purpose of the study. This is especially true for standardized (telephone) surveys, when respondents have a very short
time-span to answer questions and the interviewer cannot go back and forth in the questionnaire.

While forgetting is defined as loss or permanent inaccessibility of information in memory, blocking is an issue of temporary inaccessibility. In survey situations, blocking occurs when information has been stored in long-time memory, but cannot be retrieved on demand. In the meantime, the respondent is aware of the memory block and knows that the information will probably be accessible again later (Schacter, 1999; p. 187). A well-known example of a memory block is the ‘tip-of-the-tongue state’, in which people are unable to retrieve a specific piece of knowledge on demand, such as a name, word, or tune, but recover the memory later, either spontaneously or in response to recall cues. Latencies between first (unsuccessful) retrieval attempt and successful recovery of the information can vary between a few seconds and several days (Brown, 1991). While the tip-of-the-tongue phenomenon is especially well documented for vocabulary tasks, there might also be implications for retrospective questions in social and medical research. It is possible, for instance, that respondents experience blocking for names of former employers, or for types of medication. Since most tip-of-the-tongue blocking experiences are resolved within a few minutes (Brown, 1991), procedures that allow respondents to take their time or to review their answer – e.g., a flexible interviewing approach, in which answers can be changed or filled in at a later stage, is recommendable. As the likelihood of temporary inaccessibility of information increases with age (James & Burke, 2000; Maylor, 1990), such flexibility is especially useful when working with elderly respondents.

Unlike forgetting and blocking, absent-mindedness is not an issue of inaccessibility of information. Instead, it can be regarded as a matter of respondent focus. While Schacter was mainly concerned with absent-mindedness as a source of prospective recall error, lack of focus is also likely to harm the accuracy of retrospective reports. Lack of focus in a survey can be due to lack of motivation to answer the questions as accurately as possible. In the cognitive literature this tendency of respondents to not optimize their answers is called ‘satisficing’ (Krosnick & Alwin, 1987). Using a satisficing strategy does not imply that the respondent willfully provides an incorrect answer, but rather that he or she does not allocate sufficient cognitive resources to arriving at the correct answer. In the context of retrospective questions satisficing could mean that the respondent either uses ineffective
retrieval strategies, such as very global estimation strategies, or simply does not try hard enough to retrieve events which do not come to mind readily. This can result in response biases, such as the disproportionate use of prototypical values (Friedman, 1993) or the underreporting of behavioral frequencies (if ‘difficult’ events are omitted).

Krosnick and Alwin (1987) found that when cognitive and motivational demands were high, i.e. when respondents were confronted with a difficult or tedious response task, respondents with ‘less cognitive sophistication’ were more likely than others to employ satisficing strategies. As an effect of survey mode, use of satisficing strategies has also been found to be more prevalent in telephone than in face-to-face surveys (Holbrook, Green, & Krosnick, 2003).

1.2.2. Dating error

Although completeness is obviously a very important component of data quality, there are more ways in which retrospectively reported data can be inaccurate. In the following, I will describe several types of dating error in retrospective surveys. As will be explained in more detail in the next chapter, only some very salient events are stored in memory with an exact date. For other autobiographical memories, dates and durations of events and episodes have to be reconstructed. The amount of dating error in the data will depend on the quality of this reconstruction i.e. on the successful use of correct recall strategies. Sometimes retrospective reports of dates will be randomly off the mark, due, for example, to wild guessing on the part of the respondent. In other cases, there will be systematic errors in retrospective reports that can be caused by phenomena such as rounding, scale effects, or telescoping.

Rounding

Rounding occurs when respondents are asked to estimate intervals or behavioral frequencies. When respondents cannot remember the exact interval or frequency, many tend to round off numbers to easily recognizable units, such as ‘a year’ instead of ‘ten months’, or ten instead of nine times. This process can lead to under- or overreporting of behaviors, as well as to temporal distortions. Huttenlocher et al. (1990) asked their respondents to report either the date of a previous interview, or how many days had elapsed since then. The re-
sults of this study clearly suggest that respondents have a strong preference for rounding their responses off to prototypical values. Even though respondents were asked “how many days ago” the interview took place, the authors observed that the size of the time unit to which respondents rounded off increased with time. If less than five weeks had elapsed since the interview, respondents usually rounded off to weeks (i.e. values of 7, 14, or 21 days). If the elapsed period had been longer, they rounded off to months (i.e. values of 30 or 60 days). The authors also found that more answers were rounded down than up, which led to a net underestimation of retrospective distance.

In ‘real-life’ surveys one can also expect to find a large number of episodes, which are reported in whole or half years. Examples can be found in fertility studies, when women are asked to report for how long they have been using a certain type of contraception or breastfed their children (Becker & Diop-Sidibe, 2003; Goldman, Moreno, & Westoff, 1989). In those studies, heaping occurred at prototypical values such as six, twelve, 18, and 24 months.

There are several potential reasons why responses are rounded off to prototypical values. One of those reasons could be that respondents use satisficing strategies and simply would not like to take the time or make the effort to use a more effective dating strategy. It is also possible that in some cases the recall task posed by the survey question is simply too difficult. Especially when the reference period is long, and less salient events (e.g. small purchases or donations) have to be dated, which are not part of a logical event sequence, it is quite likely that the respondent will not be able to retrieve the event accurately, even when a recall aid is used. In other cases, rounding could be the result of a mistake made during the comprehension stage of the response process (for a detailed discussion see Ongeña, 2010). Falsely assuming that the survey interview implies the same (relatively low) standards of accuracy as an everyday conversation, many respondents do not even realize that they are not supposed to round off numbers.

**Scale effects**

A number of studies on long-term autobiographical memory have found an interesting distribution of dating error, which has sometimes been referred to as the ‘scale effect’ (Friedman & Wilkins, 1985; not to be confused with the ‘frequency scale effects’ reported by
Schwarz, 1999), since its occurrence depends on the temporal scale (days, weeks, months, or years) that is used in the survey. It seems that when respondents were asked to date events from their lives, they often were quite accurate in reporting the month of the event. Reporting the right year, however, appeared to be more difficult. When Auriat (1993) compared her male respondents’ reports of their second move after marriage to data from the Belgian population register, she found that time differences in months between reported and actual dates were often multiples of 12, i.e. respondents misestimated dates by whole years. Similar error patterns were found for reports of starting dates of training courses (Van der Vaart, 1996), and in a German life course study, in which respondents were asked to date the birth of their children (Reimer, 2003). Again, respondents dated autobiographical events quite accurately in terms of months, but not necessarily in terms of years.

There are several – not mutually exclusive - explanations for this effect. First, it is possible that respondents who make scale errors infer dates from phenomenological experiences surrounding the event rather than from calendar information. In that case, respondents might still remember very high outside temperatures when they moved to their new address, and therefore conclude that it was probably July or August, but cannot remember the exact year. Secondly, scale effects might occur because respondents relate the target event to annually recurring events, such as birthdays, university terms, or public holidays. Thirdly, as was probably the case in Van der Vaart’s study (1996), some events such as university terms or summer holidays start around the same time every year. Therefore when reporting the month of the event, the respondent does not necessarily need to think of a specific instance or year, but can infer the information using general knowledge about institutional calendars.

Reimer (2003) reported that scale effects distorted her data in a symmetrical way. Scale errors lead to similar amounts of pre-dated and post-dated events. It should be noted that scale effects are especially likely to arise in very long-term memory data, i.e. reports about events that date back at least several years (Auriat, 1993). As our current study will focus on long-term recall, we should be aware of such scale effects.
Telescoping occurs when events or episodes are projected in memory to earlier or later points in time. The first phenomenon is called ‘backward telescoping’, while the second is called ‘forward telescoping’. Although survey researchers often focus on the effects of forward telescoping, both types of error occur regularly. In Auriat’s study of event dating (1993) the error distribution for reports of remote autobiographical events was almost symmetrical, with approximately equal amounts of forward and backward telescoping error. Also, a study about memory for unemployment indicated that recalled unemployment episodes are equally likely to be subject to backward as to forward telescoping (Mathiowetz, 1986). In these questions, respondents are asked, how often a certain type of event occurred within a specified reference period.

Taking a question from a real-life survey as an example, I will illustrate in which way both types of telescoping can affect the accuracy of retrospective reports. In the health section of the 2003 questionnaire of the Panel Study of Income Dynamics (PSID), respondents were asked the following question: “How many nights were you in a hospital altogether in 2002?” This means that an adequate answer would be a report of the number of nights spent in hospital, within the reference period. Answering this question, the respondent can make a number of recall errors, which will lead to inaccurate reports. Firstly, he or she might not be able to recall some of the instances at all (see omission). This leads to underreporting of events. Secondly, respondents may think they remember an incident, which has not actually taken place. This would lead to overreporting of events. In case of rather distinctive events such as hospital stays, this type of error will probably not occur very often. Thirdly, reports can be affected by telescoping error, which entails that respondents erroneously shift events in time.

Figure 1.1 illustrates the ways in which telescoping can lead to under- and overreporting of past events. The white arrows show how telescoping can lead to underreporting, while the black arrows show how it leads to overreporting. ‘T0’ is the time of the interview, ‘t -2’ and ‘t -1’ are the beginning and the end of the reference period, respectively.
1. Events from the reference period, which in our example was 2002, can be projected into the more distant past, i.e. the time before January 1\(^{st}\), 2002.

2. Reversely, events, which happened before January 1\(^{st}\), 2002 can be projected into the reference period.

3. Events that occurred during the reference period can be perceived to have happened between the end of that period and the time of the interview (i.e. after December 31\(^{st}\), 2002).

4. Events that occurred later than December 31\(^{st}\), 2002 can be reported as having occurred within the reference period, which means they were subject to backward telescoping.

Errors of type A and C will lead to underreporting of events, while errors of type B and D will lead to overreporting of events. It must be noted that in most surveys errors of type A and B are far more likely to occur than errors of type C and D. Firstly, as memory decreases with time, the accuracy of reports is assumed to be higher for recent events than for more distant events. Secondly, the interval between the end of the reference period and the
time of the interview is usually short. Often, these two events even coincide ($t_{-1} = t_0$). In that case, the respondent will then be asked to report events, which occurred ‘during the last two years (or six months/thirty days et cetera)’.

In those situations, forward telescoping can only lead to overreporting, while backward telescoping must lead to underreporting. The problem with underreporting is, that even in record check studies, it is often not clear if omissions of events are due to backward telescoping or to other factors, such as forgetting. For this reason most authors focus on forward rather than backward telescoping. Forward telescoping has been shown to be especially strong for more salient events (Bradburn, Huttenlocher, & Hedges, 1994; Loftus & Marburger, 1983; Tourangeau, Rips, & Rasinski, 2000).

The net effect of forward and backward telescoping will depend on the length of the reference period and the saliency of the target information. On an aggregated level, it is important to know if forward and backward telescoping errors occur in equal amounts and - more or less - cancel each other out. If that is not the case, the retrieved event dates will be biased in one direction, leading to systematic forward or backward ‘net’ telescoping error.

1.3. Aided recall techniques in surveys

Survey researchers often try to maximize the value of retrospectively obtained data by using weighting and imputation techniques as part of their analyses. However, the advance of the CASM (Cognitive Aspects of Survey Methodology) movement, which started in the 1980s, has led to the development of methods that improve data quality from an earlier phase of the survey process. CASM research focuses on the question how insights from cognitive psychology can be used in order to advance the quality of survey research, specifically in the questionnaire design and data collection stages of the survey process.

In the following section, we will discuss a number of ways in which survey researchers have tried to stimulate and assist autobiographical memory in order to improve the quality of retrospectively collected data. These methods, which can be used in standardized surveys, are often referred to as ‘aided recall’ techniques. Examples include the use of multiple and supplementary questions, bounding techniques, and the use of calendar instruments, a visual recall aid.
Asking multiple questions instead of just one standardized item about the same event or behavior, allows the respondent to use multiple frames of reference when retrieving the answer from memory. One way of using multiple questions to measure the same concept is to decompose an item into smaller categories. For example, a question about the respondent’s total number of visits to health care practitioners in the last 12 months could be split into several items about visiting family doctors, registered nurses, dentists, gynecologists, or other medical specialists. In that way, the researcher provides more specific memory cues, thereby enabling the respondent to enumerate individual experiences and provide a more accurate retrospective account of their behavior (Means, Mingay, Nigam, & Zarrow, 1988). Other researchers have argued that while decomposition may lead to increased numbers of reported events, this increase could be due to overreporting rather than to higher recall accuracy (Belli, Schwarz, Singer, & Talarico, 2000). Supplementary questions work in a way that is similar to decomposition. By asking a few contextual questions, which can contain additional cues before the main item is addressed, the researcher can try to jog the respondents’ memory and give them more time to think about their answers. This can also be achieved by formulating longer questions or introductions, which can motivate the respondent to process a question more thoroughly and give a more accurate answer (Sudman & Bradburn, 1983).

Using multiple or longer questions is a reasonably uncomplicated aided recall technique that can be used in just about any retrospective survey regardless of the data collection mode and question context. Its main advantage over regular standardized survey questions lies in promoting data quality by giving the respondent more time to retrieve their answers and judge their accuracy before reporting them.

Another example of an aided recall technique is the use of so-called bounding cues. In order to decrease the amount of telescoping error in retrospective date reports, some survey researchers have provided respondents with temporal ‘anchoring points’. Usually, those anchoring points come in two types. The first variety, bounding by interview, can only be employed in longitudinal surveys. In these surveys, it has been a well-known problem that respondents report the same event as having occurred within the reference periods of different waves of the survey. For example: if a respondent has had a medical check-up shortly before the first interview, he or she will report that check-up as having
occurred recently. However, in the following wave the respondent might report that same incident again, but he or she would then think that it occurred in the beginning of the second reference period. In order to avoid these types of overreports in longitudinal surveys, researchers have experimented with ‘bounded interviews’ (Tourangeau et al., 2000). In those interviews, the interviewer is familiar with the answers the respondent gave during the previous interview. If the interviewer provides actual feedback about those answers to the respondent, this is called ‘dependent interviewing’, or - in case of computerized data collection - ‘pre-loading’ (Hoogendoorn, 2004). Bounding by interview has been shown to decrease overreporting of events at the beginning of the reference period (Neter & Waksberg, 1964), the so-called seam effect.

Unfortunately, there are some restrictions to this bounding method (Loftus & Marburger, 1983). Apart from the fact that it is relatively labor intensive, it cannot be used if the respondent has not participated in the previous wave of the survey. Loftus and Marburger (1983) therefore suggest another type of temporal anchoring which can be used with more respondents. In their study of crime victimization, they bounded the reference period by using a well-known public event as the reference point: “Since the eruption of Mt. St. Helens, has anyone beaten you up?” The authors found that using this “landmark event” reduced forward telescoping significantly.

One advantage of using landmark events over other bounding techniques is that multiple temporal anchoring points can be offered, not only to demarcate the beginning of, but also as landmarks within the reference period (e.g. “after Thanksgiving, but before Christmas”). Next to historical or public events, one can also use personal events such as weddings or relocations as landmarks (Baddeley, 1992; Means et al., 1988).

However, as pointed out by Loftus en Marburger (1983), there are some potential disadvantages to using landmark events as recall cues. First of all, landmarks can be incorrectly dated. Especially in the case of personal landmarks the researcher will not be able to verify their dates. Secondly, public events provided by the interviewer might not be equally relevant to every respondent in the sample. Thirdly, when public events are used as a demarcation points for the reference period, the length of that period will not be the same for all respondents. Especially if the fieldwork period is relatively long, the reference period of a retrospective question such as “Since the eruption of Mt. St. Helens, has anyone beaten
you up?” will be shorter for respondents who take part in the survey towards the beginning of the fieldwork period than for ‘late’ respondents.

### 1.4. Calendar methods

This dissertation focuses on the use and effects of timeline-based visual recall aids, or calendar instruments, which incorporate multiple bounding cues. The purpose of calendar instruments in retrospective surveys is to stimulate retrieval, facilitate judgment of the correctness of the retrieved information, and support response editing.

Calendar instruments consist of graphical time frames in which life history information can be recorded. They typically include a horizontal display of the reference period divided into smaller time units, temporal anchors in the shape of important public or personal events, and one or more themed timelines (see figure 1.2).

<table>
<thead>
<tr>
<th>2009</th>
<th>JANUARY</th>
<th>FEBRUARY</th>
<th>MARCH</th>
<th>APRIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>_______</td>
<td>48</td>
<td>______</td>
<td>______</td>
</tr>
<tr>
<td>ADDRESS</td>
<td>A’dam, Parkstr</td>
<td>______</td>
<td>A’dam, Churchstr</td>
<td>______</td>
</tr>
<tr>
<td>DOMESTIC SITUATION</td>
<td>partner and son</td>
<td>partner</td>
<td>______</td>
<td>______</td>
</tr>
<tr>
<td>JOBS &amp; EDUCATION</td>
<td>ING bank</td>
<td>______</td>
<td>______</td>
<td>Business French</td>
</tr>
<tr>
<td>PERSONAL LANDMARKS</td>
<td>Son moves out</td>
<td>Spain!</td>
<td>Sarah’s birthday</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1.2: Calendar instrument

24
Many different versions of calendar instruments are currently used in a variety of fields. Most were designed from a practical rather than theoretical point of view, focusing on the usability of life course data for statistical analysis. Early evaluation studies suggested that life course calendars had a number of advantages over traditional question-list surveys. The first reported calendar instrument, Balán et al.’s Illustrated Life History (1969), was found to improve the completeness and consistency of retrospective event histories by facilitating data quality checks, interviewer feedback, and data revision during the interview. The authors suggested that, apart from the editing features that it offered, the method also increased data quality by providing the respondent with recall cues.

Despite the seemingly wide-spread acceptance of the assumption that calendar instruments facilitate recall (Caspi et al., 1996; Freedman et al., 1988), it took almost thirty years before their effective mechanisms were formally linked to retrieval processes in autobiographical memory (Belli, 1998). Belli reviewed the Event History Calendar in the context of hierarchical models of autobiographical memory (Conway, 1992). He stated that calendar instruments encourage respondents to place events into a temporal context by relating them in several ways to other parallel or superordinate autobiographical events and episodes (see chapter 2).

Over the years, several authors have published methodological evaluations of various types of calendar methods. The body of research includes descriptions of individual instruments, non-experimental evaluations, and a small number of experimental comparisons between calendar instruments and regular question-lists (Belli, Shay, & Stafford, 2001; Van der Vaart, 2004; Van der Vaart & Glasner, 2007a). A comprehensive overview of methodological studies can be found in chapter three of this dissertation.

In their recently published compilation of research conducted with calendar instruments and time diaries Belli, Stafford, and Alwin (2009) highlighted a number of possible directions for future research on calendar methods. Their suggestions include cognitive testing of calendar instruments in order to provide empirical evidence for the proposed theoretical rationale, and the development of self-administered calendar recall aids for web-based surveys. This dissertation presents the results of an extensive methodological study that addresses both topics.
1.5. Research questions and approach

The study comprises a literature review, a cognitive pilot study, and a field experiment, all of which were aimed at answering the following questions:

1. What do we know about the effects of calendar instruments on data quality?
2. Which cognitive processes are involved in answering retrospective questions? Do these processes differ between interviews with calendar instruments and interviews with regular retrospective questionnaires?
3. Which components of calendar instruments are relevant to improving the quality of retrospective data in web-based life course surveys?
4. How can this knowledge be used to further improve the design of calendar techniques for web-based life course surveys?

The following chapters will start with a short review of the literature on autobiographical memory and retrieval strategies used to answer retrospective questions in survey research (chapter 2). This review addresses the second research question and provides a theoretical framework for the following chapters. The third chapter comprises a review of the methodological literature on calendar instruments currently used in medical and social research (research question 1). In the empirical part of the dissertation, we take a closer look at the way in which calendar instruments influence cognitive processes in survey interviews. Chapter four gives an account of a cognitive interviewing study in which a regular standardized retrospective questionnaire was compared to a questionnaire with a calendar recall aid with regard to the retrieval strategies that respondents in both types of interviews used when answering retrospective survey questions (research question 2). In chapter five, we describe the development and experimental evaluation of calendar recall aids for a web-based life course survey in the Netherlands. In this evaluation study, we tested three different versions of a calendar recall aid in order to examine the separate and joint effects of landmark events and visual feedback grids on data quality in web surveys (research question 3). The final chapter of the dissertation discusses the implications of the study for ret-
rospective web surveys (research question 4), the limits of our study, as well as suggestions for further empirical research.
Chapter 2:
Autobiographical memory

In this chapter, a number of current insights into autobiographical memory and their implications for retrospective survey research will be reviewed. First, general models about the organization of events in autobiographical memory will be discussed, followed by an overview of theories about the memory for dates of events. The final part of the chapter presents the theoretical rationale of calendar instruments in the context of those theories about autobiographical memory.

2.1. Autobiographical memory

Historically, memory research has mostly focused on laboratory experiments rather than the study of real life memories (Pillemer, 2003). Starting in the late 19th century with Ebbinghaus’ investigation of his own forgetting curve (Ebbinghaus, 1885/1964), a tradition developed in which memory for lists of words, nonsense syllables, and other stimuli was tested. The assumption behind these experiments was that memory was mainly a store of individual experiences and actions, and that it did not make a great difference if the stimuli one had to recall had been encountered in real life or in a laboratory situation (Neisser, 1988, p. 361). In recent years, however, the focus has shifted more and more towards studies of ‘real life’ autobiographical memories (Barsalou, 1988) and towards the narrative structures and autobiographical reasoning processes that underlie our understanding of ourselves and others (McAdams, 2001; Singer & Bluck, 2001).

There is an ongoing discussion among psychologists about what exactly constitutes autobiographical memory (Baddeley, 1992; Greenberg & Rubin, 2003). While a common-sense definition of autobiographical memory would be “the form of memory in which the events of one’s own life comprise the significant memoria” (Neisser, 1988), some authors apply far stricter criteria. Bluck and Habermas (2000) for example consider
only memories which are “linked to self through their emotional or motivational significance over one’s life” as being “truly autobiographical” (p.122). In order to resolve the conflict about definitions, Baddeley (1992) suggested that there should be two different types of research on autobiographical memory. Studies of the first type examine cognitive mechanisms and recall processes, while studies of the second type focus on construction of the self and the social and emotional aspects involved in that process.

The current study will draw mainly from the first line of research, although it will become clear later on in this chapter, that factors relating to construction of personal as well as social identities cannot be ignored when dealing with retrospective data. In the following section, the content and organization of autobiographical memory will be discussed. After that, we will examine the way in which this proposed structure of autobiographical memory shapes the outcome of the retrieval process.

2.1.1. Representation and organization of events in autobiographical memory

In its broadest sense, autobiographical memory contains information about events, that have taken place in a person’s life. There is no agreed upon definition of what exactly constitutes an event. This means that the term is used for describing a variety of remembered occurrences and behaviors, ranging from the very mundane, such as opening a can of cat food, to once-in-a lifetime events, such as graduating from high school. In traditional memory research, different kinds of information were supposed to be stored in and retrieved from memory in essentially the same way. Current views, however do distinguish at least two types of memory: episodic and semantic (Tulving, 1972). In the episodic memory system, memories of distinct autobiographical events are stored. The semantic memory system, on the other hand, contains general knowledge of language, non-autobiographical facts, et cetera. It also comprises knowledge about general features of certain types of events (such as reading the menu as part of having dinner at a restaurant). From this point of view, episodic memory feeds into semantic memory, where information from specific events is synthesized into a more general kind of knowledge. When retrieving information about an autobiographical event from memory, episodic knowledge is supported by semantic knowledge.
To continue with the restaurant example: even if we cannot remember asking for the menu the last time we visited a restaurant, we would still suppose that we did so, simply because that is what we would usually do on such occasions.

In survey research, this type of inference from semantic memory can lead to incorrect retrospective reports. Memory of recurrent events, such as regular medical visits, has been found to be less accurate than memory for distinctive events (Means et al., 1988). Especially when highly regular and similar behaviors are concerned respondents often do not “recall and count” all specific instances (Menon, 1994). Rather, they use estimation strategies, relying on information stored in semantic memory. Many survey researchers therefore provide respondents with recall cues, which are aimed at helping respondents retrieve information about specific events instead of general event knowledge.

However, the influence of our generic knowledge about the world on episodic memory is not restricted to regular or mundane behaviors. Instead, this knowledge can also affect the memory of more important life events and their dates. For example, culturally shared information about the ‘most suitable’ timing of life events such as marriage, or having children, can bias people’s estimates about when these events occurred in their own lives (Berntsen & Rubin, 2004).

Tulving’s model of autobiographical memory has inspired many authors. However, more complex models of autobiographical memory have replaced his rather minimalist view of two distinct memory systems. Especially the sharp distinction between episodic and semantic memory has been challenged by many authors (including Tulving himself in his later work). Barsalou (1988), for instance, concludes from his own studies that autobiographical memories are located on a continuum, ranging from very distinct to highly general. On one end of the continuum there are unique events, such as one’s own 21st birthday, which Barsalou calls specific events. Somewhat further, there are more abstract memories of summarized events, which have been deducted from the (episodic) memories of several specific events (e.g. one’s own birthday parties). Finally, there is relatively stable abstract knowledge about types of events (e.g. knowledge about parties in general), which is often culturally shared rather than stemming from individual episodic memories.

Barsalou describes two free recall studies, in which subjects were asked to describe events in which they participated during the past summer. He found that his subjects
had a preference for reporting summarized events rather than specific ones. He also observed that these summarized, ‘extended’ events seemed to be chronologically organized and appeared to form a ‘life narrative’. Based on these findings, Barsalou proposes a theory of autobiographical memories in which “extended-event time lines (sic) are the primary organizers” (p. 219). The model states that there are parallel timelines for all important life domains, such as school, relationships, living situation et cetera. Extended events from one’s life are located chronologically on those timelines. The relationship between the different timelines is defined by ‘logic of goal attainment’. This means that in autobiographical memory, events are stored in close proximity to other events that enabled or prevented them.

An extension and specification of Barsalou’s ideas can be found in Conway’s structural model of autobiographical memory (1992, 1996). Similar to Barsalou, Conway proposes three different types of memory structures, which are linked to each other causally and temporally: ‘lifetime periods’, ‘general events’, and event details, the ‘phenomenological record’. He further suggests that autobiographical memory might be organized hierarchically rather than in a ‘flat’ way.

On top of the hierarchy, there are very long-term extended events or ‘lifetime periods’, such as working for a certain employer or living in a certain city. Within those lifetime periods, different ‘themes’ (e.g. work, relationships etc) can be distinguished. One step lower in the hierarchy, there are the memories of summarized events, or ‘general events’, which took place during those lifetime periods. These events vary in specificity. The structure includes memory of more general events such as reading a lot of books, as well as more specific memories, such as working on a certain project, going on holiday, or meeting one’s partner for the first time. Memories of these general events are anchored in the ‘phenomenological record’, the memory structure in which very specific phenomenological experiences are stored. Those experiences include sensory (e.g. ‘it was cold that day’) as well as emotional information (e.g. ‘I was feeling disappointed’).
Conway (1992, p. 178) illustrates the connection between the three memory structures with the following example: “when I lived with X” (lifetime period), “difficulty in personal relationships” (theme within lifetime period), “holiday in Italy” (general event), “the fiasco at the airport” (specific event, which provides access to the associated phenomenological records).
2.1.2. Life scripts

In the recent literature about storage and retrieval of autobiographical memories, the role of personal and cultural “life scripts” in structuring autobiographical memory has received much attention (Berntsen & Rubin, 2004; Kirkegaard Thomsen & Berntsen, 2008). In relation to Conway’s structural model (1992), these life scripts or “life story schemas” (Bluck & Habermas, 2000) can be seen as a superordinate structure, which encompasses the lifetime periods. The underlying assumption is that people perceive and remember autobiographical events in the larger context of their lives as a whole. Life story schemas mold a person’s biography into a coherent life story. Coherence, in this context, has several aspects.

Temporal coherence means that there is some chronological order to the life story. Referring to Conway’s model, this implies that the life story is (roughly) composed of a temporal sequence of lifetime periods.

Thematic coherence allows individuals to tell a life narrative about one thematic domain in their lives, i.e. “the story of my relationships”. This type of coherence can also be found in Conway’s lifetime periods-structure.

Causal coherence is needed in life stories, in order for individuals to explain to themselves and others, why life events occurred, and what led them to certain decisions. In Conway’s model, this type of coherence is represented in the causal links between the different life domains or themes.

This view of autobiographical memories fitting into a life story schema can be regarded as a primarily functional approach to autobiographical memory. Rather than being concerned with the accuracy of autobiographical memories, i.e. their correspondence to the actual event, this approach explains why and how individuals remember events from their past (Bluck, 2003; Conway, Singer, & Tagini, 2004). In survey research this focus on coherence rather than correspondence can lead to serious errors in retrospective reports. The following example from the German Life History Study illustrates how life scripts or schemas can influence retrospective reports about unemployment.

In retrospective surveys, unemployment spells are very likely to be underreported or abbreviated (Dex, 1995; Jacobs, 2002; Jürges, 2004) because of their negative emotional
connotations. Also, respondents tend to re-interpret those episodes as other forms of labor market inactivity, such as ‘being a homemaker’ or ‘taking care of the children’. In this context, Reimer (2004) was rather surprised to find the opposite pattern in an analysis of employment episodes from the German Life History Study. When comparing reports about the same period from two waves of the longitudinal study, she discovered that unemployment spells are actually over-reported in the second wave. A closer look at those episodes revealed that, conversely, episodes of ‘labor market inactivity’ were underreported compared to the first measurement. The author concluded that those surprising findings could probably be explained by “fundamental changes in labor market context between the two interviews”. Respondents from the (ex-communist) former German Democratic Republic, who had little familiarity with the concept of unemployment at the time of the first survey, might have defined their situation as temporary inactivity or leave at first. However, with hindsight, they re-interpreted those episodes as unemployment. Viewed in light of a shared cultural concept of biography (see above), this suggests that the respondents’ cultural life script has changed between the first and the second wave of the survey. The typical biography in Eastern German society before 1990 did not include unemployment. Unfortunately, the labor market situation has changed dramatically in the years after German reunification, making unemployment spells a widely shared and therefore less stigmatized experience in the life course.

**Cultural life scripts and retrospective behavioral reports**

In writing about differences in cultural life scripts and their potential effects on development and performance of autobiographical memory (among other dimensions of psychological functioning) most authors so far have focused on the dimension of individualism versus collectivism (for an overview, see Oyserman, Coon, & Kemmelmeier, 2002). Guided by the assumption that European-American or Anglo-Saxon cultures promote more individualistic values than, for instance, East Asian cultures, members of those groups have often been compared to each other in cross-cultural psychological research. Wang and Conway (2004), for instance, compared the unique memories that Chinese and European Americans adults reported in a free recall task. Chinese participants were more likely to mention shared social-cultural events than Americans, while Americans tended to report
more personal memories. On a related note, and more specific to survey research, culture seems to affect the quality of behavioral reports (Ji, Schwarz, & Nisbett, 2000). Chinese respondents were less influenced by the format of response scales than American respondents when reporting frequencies of observable behaviors (going to the library, coming late to class, catching a cold) but not of relatively unobservable behaviors (telling a lie, faking agreement, having a nightmare). The authors offer the explanation that self-monitoring is more common and pronounced in collectivist cultures, such as the Chinese, which could influence the encoding and retrieval of behavioral frequencies observable by others. Thus differences in cultural life scripts, as well as in cultural values such as individualism versus collectivism can influence responses to retrospective survey questions.
2.2. Memory for time

In surveys, respondents are often asked to report not only if they have experienced certain events, but also when these events took place, and how long they lasted. In that case the respondent has to retrospectively report dates and durations of past events or episodes. With some questions, a reference period is provided (e.g. “When in the last 2 years did you buy that car?”). Other questions do not specify a reference period (e.g. “When did you first start school?” or “When did you last see a doctor?” etc). The question we would like to answer is which strategies respondents use to retrieve dates and durations of past events. In order to do so, we will first examine the way in which those dates and durations are stored in memory.

2.2.1. Time in autobiographical memory

In autobiographical memory only a few events are time-stamped or time-tagged, i.e. stored in memory with a specific date (Friedman, 1993). This form of representing temporal information is also called a calendar representation (Huttenlocher, Hedges, & Bradburn, 1990). Events that are stored with a time-tag are usually retrieved quite accurately (Burt, 1992). With regard to public events, time-tagging is evident for events that have been rehearsed so often that they are actually referred to by their date, such as ‘9/11’, the day of the terrorist attacks on the World Trade Center in New York, in 2001. The format of this particular time-tag suggests that in the long run, day and month of the event might be remembered more reliably than the year in which it took place (see p. 13, scale effects). Other events are time-tagged (at least with regard to day and month) because they are commemorated or celebrated on the same designated date every year, e.g. Christmas, or National Days. A large-scale online experiment in the Netherlands identified a third type of occasion on which public events are stored with a time tag. Janssen, Chessa, and Murre (2006) showed that Dutch respondents also dated a royal wedding, which took place on “easy to remember” February 2nd, 2002 (i.e. 02-02-02), very accurately. In an event dating study conducted among college students, holidays and special occasions (which probably
offer a combination of public and personal event cues) were the event category most likely to be dated exactly (Thompson et al, 1996).

In terms of personal memories, only some very salient events are time-tagged. Once again, events that are rehearsed and commemorated regularly are most likely to be stored with a specific date. These events include the birth of a child, which is likely to be celebrated on the same day every subsequent year, one’s own wedding (anniversary), or high school graduation (reunions). Most autobiographical events, however, personal as well as public, do not occur on memorable dates or are salient enough to be time-tagged. This means that dating them involves a certain amount of reconstruction (Friedman, 1993).

2.2.2. Reconstructing dates and durations

One method of reconstructing the dates of autobiographical memories is to relate them to time-tagged public or personal landmark events stored in autobiographical memory. An example from the personal domain would be ‘getting one’s driver’s license two weeks after one’s 18th birthday’. A personal memory tied temporally to a public landmark could be ‘visiting Czechoslovakia the summer after the collapse of Communism’.

A second way of making use of event sequence representations for dating autobiographical events is a technique called sequencing (Belli, 1998), in which the event is connected to thematically related preceding and/or succeeding events. In this case, the temporal framework used for dating the events is not provided by a calendar, but rather by an event sequence (Huttenlocher et al., 1990). In the context of Conway’s model of autobiographical memory, this means for example, that a general event like “working on project X” took place before working on the current project (project Y), but after the summer holiday. Apart from providing us with (somewhat global) temporal information, sequencing strategies also help us to contextualize events, and report them as a narrative, thereby reducing the risk of omitting events.

Thirdly, the date of an event or episode can be inferred from other domains within the structural model of autobiographical memory. This means, that if an event falls into a clearly defined lifetime period or other extended event, the start and end dates of that period form the temporal boundaries within which the other event is located. For example, if
a person started her university education in September 2002, the event ‘sophomore spring break’, most likely took place between February and April of 2004. In relation to Conway’s model, Belli (1998) calls this way of inferring dates “top-down retrieval”. He also suggests that this type of retrieval might be especially likely to occur, when respondents have to retrieve isolated events, which they cannot place in an event sequence.

Fourthly, the temporal boundaries within which the event is placed do not necessarily need to be derived from more abstract events. It is also possible that timing cues are obtained from other thematic structures within the same lifetime period or extended event (e.g. “When I called off my engagement, I lived in X and had just started working as a teacher”). This process is called parallel retrieval (Belli, 1998).

Finally, the timing of autobiographical events can also be based on temporal information inferred from the phenomenological record. If one remembers, for instance, that an event like “moving to Amsterdam” took place while it was freezing outside, one can usually assume that it was winter.

It must be noted that the event dating strategies mentioned above are not mutually exclusive. As a matter of fact, motivated respondents will often use multiple, if not all, strategies for reconstructing dates and durations of past events. This can be illustrated by an example: I know that somewhere in the past, I completed a 3-month internship, but I cannot remember the exact year or month. In order to date the event I might consider the following:

- The internship started shortly after my birthday, so it must have been February (= personal landmark event)
- It was part of my graduate studies, so it must have taken place in 2001 or 2002 (= deduction from higher structure / “top-down-retrieval”)
- Before I started the internship, I followed a methodology class. At the end of the internship there was some overlap with a social psychology class (= sequencing)
- I lived in XY Street with A (= parallel cueing).
- My compensation was paid in guilders, not Euros (= inference from phenomenological record in combination with public landmark event). As the Euro was introduced in January 2002, the internship must have started in February 2001.
In this way, I have reconstructed the timing of an autobiographical event by using information from different memory structures. It should be noted that the date could have been inferred using only part of the total information I was able to retrieve. However, the various bits of dating information that were used during retrieval probably affected different aspects of the date reconstruction process. While the personal landmark event (birthday) helped to determine the month of the event, the temporal information provided by for instance the parallel (living arrangements) and top-down (graduate studies) cues allowed me to narrow down the potential year to a relatively short period.

Generally speaking, if a survey respondent uses multiple dating strategies for the same event, this does not always mean he actually needs more than a single dating cue; it does not even necessarily mean that the event is not time-tagged. If the respondent is motivated to date an event as accurately as possible, he might retrieve dating cues even after providing a sufficient answer, in order to verify the accuracy of the reported event date.

2.2.3. Reconstruction strategies and dating accuracy

The reconstruction strategy used by the respondent can influence the accuracy of the estimated date. Based on the literature on dating accuracy (e.g. Thompson et al, 1996) we would expect that the use of multiple cues for dating the same event would increase accuracy, especially if these cues are consistent. We would also expect that more sophisticated dating strategies (i.e. those that include specific information about the time of events) yield more accurate results than more general strategies, such as rough estimates of time passed since the event (Tourangeau et al, 2000). In event dating studies among diarists, Burt (1992) and Thompson et al (1996) found that subjects remembered directly retrieved dates of autobiographical events more accurately than reconstructed dates. Furthermore, landmark- and detail-based date reconstruction strategies led to better results than ‘wild guessing’. Based on those results, our assumption is that researchers who collect retrospective data should favor answers based on direct date retrieval over reconstructed dates, which in turn should be preferred to guessed dates. Recall aids should only be offered if the researcher does not expect respondents to be able to access exact temporal information for all target events. A recall aid should stimulate the use of direct date retrieval, and – if a time-
tag is not available – date reconstruction strategies, and discourage the use of general estimation strategies.

2.3. Summary and Conclusion

This review of current ideas in the literature on autobiographical memory shows us that in order to enhance recall of specific life events, it is advisable to stimulate the recall of episodic rather than of semantic information concerning these events. If the respondent relies too much on semantic information in memory, this increases the risk of using heuristics such as social and cultural norms and life scripts. Those strategies are especially likely to occur if the target event or behavior is recurrent or otherwise indistinctive. The use of heuristics can affect the accuracy of retrospective survey data in several ways (e.g. use of prototypical values), which were discussed in the introduction of the dissertation. Recall aids used in survey research should therefore encourage the respondent to establish an episodic context of the target behavior in two ways. Firstly, they should stimulate the respondent to use temporal and causal relationships with other events as sequencing cues, especially landmarks and events that had a goal relationship with the target event, enabling or preventing (aspects of) it. These cues can be from the same life domain, or from other domains. Secondly, recall aids can install an episodic context by helping the respondent access information from the ‘phenomenological record’ when reconstructing event dates. This information will distinguish specific events from other, similar events stored in memory.

Hierarchical models of autobiographical memory show that information is stored in memory on different levels of abstraction, from specific to increasingly global. Our questioning procedures should be attuned to the level of information that we are interested in. Since we focus on life history information collected in social surveys, we are mostly interested in collecting data on events within certain life themes (e.g. work and health). The target data, which will be used for event history analysis concerns specific events (e.g. specific jobs or illnesses) within these life themes.

This study focuses on the recall of autobiographical events and their dates. Since most events are not time-tagged, dates will often have to be reconstructed rather than re-
trieved directly. Recall aids can prompt the respondent to report very salient personal events that are time-tagged in the respondent’s memory and might be used as memory landmarks. Furthermore event sequences and events from other thematic domains can be used to aid recall. It is very important that the (landmark) events from which the date information is derived are dated accurately themselves. If respondents base their temporal estimates on faulty information, this could lead to more dating error. While it is true that the dates of public landmark events can be verified more easily by the researcher than dates of personal events, public events are far less likely than personal events to be embedded in autobiographical event sequences by logic of goal attainment. Furthermore, if ‘true’ autobiographical memories are only those memories that are connected with the self (Bluck & Habermas, 2000), the great majority of public events will not qualify as landmark memories. Therefore it is not clear to which degree those events are useful as cues for retrieving personal events and their dates from autobiographical memory.

The next chapter will present an overview of studies in which different types of calendar instruments were used as recall aids.
3.1. Introduction

Calendar and timeline methods are aimed at helping respondents gain better access to long-term memory by providing them with a graphical time frame in which life history information can be represented (Van der Vaart, 2004). This stimulates the respondent to relate, visually and/or mentally, the timing among several kinds of events. Inconsistencies in reports are more easily discovered and one event may prompt the recall of another. Additionally, detailed sequences of events are easier to record since they can be marked graphically in the time frame. There are some small (historical) differences between the concepts ‘calendar’ and ‘timeline’, as will be pointed out later, but here we will use the word ‘calendar’ for both methods.

Even though the general assumption is that calendar methods improve data quality, there has been limited methodological research into their effectiveness. Since using these aided recall procedures might increase operational costs of the survey, more methodological insights are needed. In the past decades, several authors (e.g. Freedman et al., 1988) have described the method in detail, most of the time focusing on the specific type of calendar, which they used in their own study. Still, little is known about the effects of calendar instruments on data quality, measured in terms of completeness and consistency of the data and the occurrence of dating errors. Only recently, a number of experimental stud-

---

ies have been conducted, the results of which indicate that calendars can indeed be beneficial to data quality.

This chapter presents an overview of calendar instruments currently used in different fields of research and their effects on the accuracy of retrospective reports. Firstly, I will provide an overview of the most important areas of application and present the instrument’s rationale. Secondly, design features and the suitability of the instrument for different modes of data collection will be described. Thirdly, there will be a detailed discussion of the effects of several types of calendar instruments on data quality. Fourthly, I will present the consequences for operational costs and a summary of interviewer and respondent evaluations of those instruments. Finally, I will draw conclusions from our review and offer suggestions for further research into specific features of calendar procedures.

3.2. Calendar instruments: applications and rationale

3.2.1. Overview: different names, similar instruments

Reviewing the literature on calendar techniques used in social and medical research reveals that neither the instrument nor the terminology has been standardized so far. In the social sciences different names for calendar instruments include Life History Calendar (Freedman et al., 1988), Timeline (Van der Vaart, 1996), Life History Matrix and Time Axes (Brückner & Mayer, 1998), Event History Calendar (Belli, 1998), Life Events Calendar (Hoppin, Tolbert, Flagg, Blair, & Zahm, 1998), Illustrated Life History (Balán, Browning, Jelin, & Litzler, 1969) and month-by-month calendar (Becker & Sosa, 1992). In the medical sciences simple calendar instruments, which measure only one behavioral domain during a relatively short reference period are usually called ‘timelines’ (Sobell et al., 1988). The reason for the division between (medical) ‘timelines’ and ‘calendars’ is foremost a historical one.

In principle, all calendar or timeline instruments are based on the same idea, which is to enhance autobiographical recall by providing the respondent with event cues. In the following, the term 'calendar instrument' will be used to refer to all of these instruments. Although different versions of calendar instruments have been developed relatively inde-
pendently from each other in different fields of research, they usually take the form of two-dimensional grids (see chapter 1, figure 1.2), which all share three important characteristics:

1. The instrument includes a graphical display of the time dimension (often on the horizontal axis of the grid). This reference period is divided into smaller time units, such as years, months or days. The size of the time units mainly depends on the length of the reference period.

2. The graphical display encompasses one or more thematic axes, representing the domains (e.g. employment, relationships, or accommodation) concerning which the data is collected. Those thematic domains are usually presented as parallel timelines.

3. The respondent is provided with temporal bounding cues, such as public or personal landmark events.

The number of domains and the length of the reference period can vary to a great extent, but most calendar instruments comprise multiple life domains and reference periods that are longer than one year. The landmark domain may contain personal landmark events or public event cues (e.g. Hoppin et al., 1998), as well as a combination of the two.

During fieldwork, calendar instruments can be used either as a separate memory aid or as a data collection device. The latter usually applies to the complicated, interviewer-applied Event History Calendars (Brückner and Mayer, 1998). As an alternative, uncomplicated calendars can be used as separate aided recall devices in addition to (often standardized) survey questionnaires (Van der Vaart, 1996, 2004).

Calendar instruments are currently used in a variety of fields, including life course research (Axinn, Pearce, & Ghimire, 1999; Caspi et al., 1996; Reimer, 2004; Smith & Thomas, 2003; Van der Vaart, 1996), epidemiology (Colt, Engel, Keifer, Thompson, & Zahm, 2001; Engel, Keifer, & Zahm, 2001; Hoppin et al., 1998; Wingo, Ory, Layde, & Lee, 1988), family planning studies (Becker & Diop-Sidibe, 2003; Becker & Sosa, 1992; Goldman et al., 1989; Rosenberg et al., 1983), health behavior (Wiebe & Landis, 2000), sexual risk taking (Martyn & Martin, 2003) and domestic violence (Yoshihama, Gillespie, Hammock, Belli, & Tolman, 2005).
The wide range of topics is also reflected in a great diversity of studied populations. Calendar instruments have been administered to young adults (Freedman et al., 1988; Van der Vaart, 1996), teenage girls (Martyn & Martin, 2003), and demographically diverse samples (e.g. Belli et al., 2001). They have also been used in many countries, including the United States (Belli et al., 2001, 2007; Colt et al., 2001; Engel, Keifer, Thompson et al., 2001; Engel, Keifer, & Zahm, 2001; Freedman et al., 1988; Hoppin et al., 1998; Kominski, 1990; Martyn & Martin, 2003; Yoshihama et al.), Mexico (Balán et al., 1969), Peru (Goldman et al., 1989), Costa Rica (Becker & Sosa, 1992), Nepal (Axinn et al., 1999), New Zealand (Caspi et al., 1996), Switzerland (Eisner et al, 2009), and the Netherlands (Van der Vaart, 2004; Van der Vaart & Glasner, 2007a), as well as internationally (Dijkstra & Ongenaa, 2007a; Börsch-Supan, Brandt, Hank, & Schröder, in press).

3.2.2. Theoretical rationale of calendar instruments in retrospective surveys

As discussed in section 1.4, calendar instruments were originally designed to meet practical research needs, which means that in the early days relatively little attention was paid to their theoretical foundations. This situation changed with the advance of the CASM movement (see section 1.3), which focuses on the relationship between ‘cognition’ and ‘survey methods’ (see Tourangeau, Rips, & Rasinski, 2000). CASM studies demonstrated that retrospective data suffer from recall error like omission, underreporting, dating errors, and biased representations, all aspects that are central to calendar techniques (Schwarz & Sudman, 1994; Van der Vaart, 1996).

According to Belli (1998), calendar instruments offer a variety of event cues. Conventional survey questionnaires, on the other hand, stimulate the use of effective retrieval techniques only to a limited degree. Conventional questionnaires are usually structured in a way that prompts the use of top-down and sequencing cues, but does not explicitly stimulate respondents to make parallel or logic-based inferences from other thematic domains.

In both regular as well as calendar interviews, survey respondents could be asked for example, to name in forward or in backward order, all employers they have worked for...
and date those employment episodes (= sequencing), or to indicate their career development while working for each of these employers (= top-down retrieval). Even though sequencing and top-down strategies can be used with other question-list formats, the advantage of the calendar over a regular survey questionnaire lies in the fact that event sequences can be displayed graphically and that the respondent can use that graphical display to detect overlap and gaps between episodes.

Belli argued that the Event History Calendar also encourages respondents to place events into a temporal context by relating them in different ways to other parallel or superordinate events than regular survey questionnaires. Firstly, the landmark events used in calendar applications serve as temporal anchoring points, or bounding cues. Secondly, the visual properties of the calendar make it possible for respondents and interviewers to link episodes across life domains, thereby potentially encouraging top-down and parallel retrieval. Especially the option of applying parallel crosschecks is a distinguishing feature of calendar instruments. Example questions could be: “Did the incident occur while you were living in X?” or “Did you start working for Y before or after you finished college?” et cetera. Compared to sequential probing (e.g. “What happened before/after [episode]?”), parallel probing requires a more flexible interviewing approach, as the interviewer will have to choose the most suitable parallel domain(s) to refer to, that is the domains which provide the most relevant retrieval cues for the respondent.

Apart from providing survey respondents with retrieval cues, calendar instruments have been argued to improve the accuracy of retrospective reports by giving respondents as well as interviewers the opportunity to review and revise answers during or after the interview (Schwarz & Oyserman, 2001). A visible gap in one of the calendar’s timelines, for example, could indicate that the respondent either forgot to report an episode, or dated the start and/or end dates of other episodes incorrectly.
3.3. Design features of the calendar

As can be expected from an instrument that is used in so many different fields, calendar techniques come in many different versions. The following sub-sections describe the three main design features of calendar instruments (as identified in section 3.2.1.): time dimension, domain grid, and landmark events.

3.3.1. Time dimension

The standardized time dimension in calendar instruments is represented by a timeline on which the reference period is displayed in calendar time, sometimes supplemented by a second timeline, on which the age of the respondent at several points of that period is noted. As calendars are often used in life course studies, some applications encompass many years of a respondent's life history and can range from the year the respondent was born to the time of the interview (Hoppin et al., 1998). However, calendar instruments are also used in surveys with far shorter reference periods. Belli et al. (2001), for example, included a calendar in a biennial social survey, in which respondents answered retrospective questions about the two-year period preceding the interview. As a general rule, calendars with a shorter time-range allow for more detailed recording of events, since the temporal axis can be split into shorter time units (e.g. months instead of years). Some calendar instruments used in the medical field even measure behaviors in terms of 24-hour intervals, so that daily fluctuations in for example alcohol consumption (e.g. Sobell & Sobell, 2003) can be recorded with the instrument. At the other end of the spectrum, there are calendar instruments in which retrospective life histories are recorded in year units (e.g. Axinn, Barber, & Ghimire, 1997).

When collecting retrospective information about the life courses of a demographically diverse sample with a wide age range, the problem occurs that the reference period of the calendar instrument is different for respondents of different ages. For paper-and-pencil calendar instruments, Axinn and his colleagues (1999) suggested to either use a fixed reference period, thereby left-truncating the collected data for most respondents, or to provide
calendars large enough to record the entire life course of the oldest respondent in the sample, and to start recording data for the other respondents in the center rather than at the left end of the calendar. Another way of dealing with age differences in the sample would be to leave the time dimension blank so that the interviewer can customize the age and calendar year timelines for each respondent at the start of the interview and cut off or fold back the rest of the calendar. While this approach leads to a more manageable calendar size for most interviews, writing down calendar year as well as age takes extra time. If possible, this should either be done before the interview, or adjustable year and age lines could be used which the interviewer can tape onto the calendar in the appropriate place.

Computerized calendar instruments (e.g. Dijkstra, Smit, & Ongena, 2009) are usually designed in a way that the visual feedback provided by the calendar automatically adjusts to the respondent’s age so that the calendar does not comprise any empty columns on the left of the display.

### 3.3.2. Domain grid

The number of thematic domains reflected in the calendar instrument varies widely between studies. While some calendar instruments measure only one life domain such as work (e.g. Engel, Keifer, Thompson, & Zahm, 2001) or gambling (Weinstock, Whelan, & Meyers, 2004), most calendars used in the social sciences include multiple thematic domains. If the instrument is used as a recall aid next to a standardized questionnaire, the order of thematic domains in the calendar will often correspond to the question order in the survey, although not all details and themes may be displayed in the calendar. Flexible calendar instruments that are used as data collection tools (i.e. Belli et al’s Event History Calendar, 2001) often allow respondents more freedom in deciding in which order they answer the questions, including the possibility to go back and forth between domains during the interview.

In terms of the theoretical rationale of calendar instruments, this use of multiple timelines should enable respondents to infer event dates by taking parallel memory cues from other thematic domains. Given this purpose, researchers should aim to start calendar interviews by collecting information from thematic domains that can serve as retrieval cues.
later during the interview. The most suitable thematic domains to start the interview with would therefore be those domains for which dates of events can be retrieved relatively easily, i.e. domains that comprise a relatively large number of time-tagged events (see section 2.2.1). For example, for most people the birth dates of their own children will be time-tagged events. However, as was mentioned before, the structure of autobiographical memories and retrieval pathways is highly idiosyncratic, and depends not only on the autobiographical experiences of a person (e.g. whether or not that person has a complicated job or relationship history) but also on individual cognitive abilities and preferences (e.g. whether one is good at remembering dates). This makes it difficult for the researcher to determine which domain order best facilitates retrieval processes and provides relevant cues to the respondent. It is therefore advisable to pay attention to this issue when pre-testing questionnaires for retrospective surveys. During such a pre-test, one could for example directly ask respondents to indicate in which order they would like to answer the questions, if they were completely free to choose any question (or domain) order. Alternatively, different versions (with regard to domain order) of the same questionnaire could be evaluated using behavior coding or cognitive interviewing techniques and coding the number and quality of instances in which respondents use parallel retrieval cues.

### 3.3.3. Landmark events

As an additional design element, most calendar instruments include a timeline on which landmark events can either be specified by the researcher, or written down by the interviewer during the interview. These landmarks are personal or public events from the reference period, which are meant to help the respondent date other events as he progresses through the interview. By definition, those landmark events need to be time-tagged, so that they can function as reliable dating cues for the respondent. Usually, time-tagged landmarks are very salient autobiographical events, which ‘stand out’ in memory and have been retrieved and rehearsed relatively often. Shum (1998) defines landmark memories as playing an active as well as a passive part in autobiographical memory. Not only do landmark events serve as indexes that (actively) help people organize and access other autobiographical memories, they are also stored in memory in a more detailed way than other events.
In some studies landmark events are recorded in the timeline of a thematic domain, especially when information about only one life domain is collected (Engel, Keifer, & Zahm, 2001; Hoppin et al., 1998; Rosenberg et al., 1983). However, most studies include a separate timeline on which to write down landmark events (Belli et al., 2001; Van der Vaart, 2004). In order to make the calendar instrument more attractive and more easily understandable for populations with limited literacy, some researchers used icons and toy figures (Engel, Keifer, Thompson et al., 2001), or adhesive pictures (Hoppin et al., 1998) instead of written cues. Instructions as to which type of events the respondent might use as landmarks differ per study. In general, the researcher will be interested in those landmarks from a functional point of view, that is: if the landmark helps the respondent date other events more accurately, it does not matter what type of event the respondent uses. Nonetheless, most surveys in which calendars are used will give a number of examples, often including vacations, family events, major health events and the like. Two analyses of personal landmark events reported by respondents in calendar studies (Van der Vaart & Glasner, 2011; Van der Vaart & Glasner, 2007b) indicated that the wording of the instructions to report personal landmark events might steer the respondent’s choice of landmarks. In both studies, respondents tended to report relatively large numbers of personal landmarks from the thematic categories that were listed as examples in the instructions.

Next to the personal events mentioned by the respondents, many instruments also offer public event cues (e.g. Hoppin et al., 1998) such as natural disasters (Loftus & Marburger, 1983) and other memorable historical events. The Neighborhood History Calendar used by Axinn and his colleagues (1997) in Nepal contained memorable public events from national (e.g. deposition of the king), regional (e.g. natural disasters) and local levels (e.g. accidents, or neighborhood level changes, such as household electrification), which the researchers thought might be more useful dating cues than the standard calendar time for many of their respondents. Furthermore, institutional, economical, and educational calendars might provide respondents with temporal reference points, such as school terms and public holidays.

So far, few researchers have studied landmark memories with regard to their content (e.g. Shum 1998, Van der Vaart & Glasner, 2011). In a study conducted by Yoshihama
and colleagues (2005), respondents were asked which landmark cues they thought were helpful, in this case in recalling partner violence. Respondents mentioned cues from their relationship history and self-generated landmarks (‘significant life events’) as the most helpful ones. This finding indicates that events from thematically related domains as well as other very salient autobiographical events can offer useful memory cues.

3.4. Potential challenges in cross-cultural surveys

Until now, calendar instruments have seldom been used in international surveys. Presently, we are only aware of two studies, in which (semi) standardized calendars were applied in multiple countries: the SHARELIFE survey which was conducted in 13 European countries (Börsch-Supan et al, in press) and an epidemiological case-control study among 26,000 patients from hospitals and clinics in Germany, France, Italy, Greece, and Slovenia described by Dijkstra, Smit and Ongena (2009). Given the popularity of calendar methods, however, it is not unlikely they will be used in more international studies in the future. In this context, we would like to point out a number of potential challenges that researchers could encounter when applying calendar methods cross-culturally.

First of all, there are some potential problems with the calendar instrument’s standardized time dimension. The time units of the Gregorian calendar might not be equally meaningful to all populations. In many Asian countries for example, local calendars, which can provide important temporal reference points in surveys (Haandrikman et al. 2004; Axinn et al., 1999), are used instead of or next to the Gregorian calendar (Wong, Neoh, Lee, and Thong 1990). Also, respondents in populations that are relatively unconcerned with standardized calendar time might not be able to retrieve and report any time-tagged personal events, which can be used as memory landmarks. In some (but not all) non-industrialized regions, a considerable proportion of survey participants might not be able to report, for example, in which year or at what age they were married (Singh & Samara 1996), or in which year and month their first child was born (Gage 1995). Furthermore, the calendar instrument’s time dimension is usually displayed horizontally. While a left-to-right horizontal order of events might be the most intuitive way of displaying temporal information to users of, for instance, the Latin and Cyrillic writing systems, this might not be
the case for all populations. East Asian scripts can often be written vertically as well as horizontally, and some other languages, such as Hebrew and Arabic, are written right-to-left. If reading and writing direction affect optimal display of the time dimension, then it will also influence the direction in which event sequences are recorded in the calendar. Cross-linguistic psychological studies (e.g. Dobel, Diesendruck, & Bölte 2007) have shown that writing direction influences spatial representations of events, which indicates that event sequences might also be more meaningful to respondents if displayed in the same direction as their day-to-day reading and writing.

Finally, cultural differences might also interact with the logic underlying the representation of events in a calendar. One of the main purposes of graphically representing domain data in life history calendars is the reduction of incompleteness and inconsistencies in the data (Balán et al. 1969). Experimental (i.e. split-ballot) comparisons between calendar methods and traditional questionnaires have shown that calendars or timelines reduce the superposition of (supposedly) mutually exclusive behaviors (Becker & Sosa 1992). However, events or states, which are mutually exclusive in one cultural context, do not necessarily reflect response inconsistency in another culture. For their research in Nepal, Axinn et al (1999) had to modify American calendar instruments in order to make it possible to code multiple simultaneous marriages. If this modification had not been made, respondent reports of multiple spouses would have been treated as inconsistencies and been removed from the analysis.

As a consequence of the potential problems listed above, calendar instruments for international surveys should be pre-tested very carefully in all participating countries. Researchers should be aware of cultural differences in the use of and reaction to the recall aid in order to avoid introducing (additional) cultural bias into the results of their survey.
3.5. Experiences with different modes of data collection

Some calendar instruments are completed by the interviewer, while others are filled in by the respondents themselves. The traditional life history calendar (LHC) is an interviewer-administered paper and pencil questionnaire (Balán et al., 1969; Freedman et al., 1988; Wingo et al., 1988). However, it has been used in personal interviews as well as in telephone administrations. Freedman and her colleagues (1988) administered almost half of their LHC-interviews via the telephone, in which case the calendar was not visible to the respondent. It could only help the interviewer to detect inconsistencies in the respondent's reported employment history, and prompt the respondent with personalized cues in order to resolve these inconsistencies. Rather surprisingly, the authors conclude, that “the two modes produced almost the same degree of consistency” (p. 65) between the collected retrospective reports and the control data. Unfortunately, the authors do not elaborate on these results. Given their earlier assumption that the respondent uses the calendar as a visual recall aid, one would expect the instrument to be less beneficial to data quality during a telephone interview.

Next to these paper-and-pencil calendar instruments, computerized versions of calendar methods have been evaluated (Belli, 2000, Belli et al. 2007; Wiebe & Landis, 2000). Similar to paper-and-pencil calendars, the computerized instruments are used in personal interviews as well as with CATI applications. The Panel Study of Income Dynamics (PSID), for instance, includes a computerized Event History Calendar (EHC) together with a CATI questionnaire (Belli et al., 2001, 2007). A version of this computerized EHC was used during personal interviews of the Los Angeles Family and Neighborhood Survey (LAFANS) (Pebley & Sastry, 2004). In the first wave of the study, the calendar instrument was used for collecting data about a reference period of two years previous to the survey. In the second wave, the reference period of the EHC was six to seven years. The EHC part of the LAFANS interview was interviewer-administered, although interviewers were instructed to review (and if necessary revise) the completed EHC together with the respon-
dent. In the case control study reported by Dijkstra and his colleagues (2009), computerized calendar applications were used, which were displayed on two separate laptop screens, one for the interviewer and another one on which the respondent could see the Event History Calendar during the interview.

Instead of using interviewer-administered calendar instruments, some questionnaires include calendars, which are completed by the respondents themselves. These calendars can be part of either paper-and-pencil questionnaires (Martyn & Martin, 2003), or they can be used as visual recall aids during personal (Van der Vaart, 2004) or telephone interviews (Van der Vaart & Glasner, 2007a). In the latter study, the calendar instrument was mailed to respondents prior to the interview. In the advance letter and the accompanying written instruction respondents were asked to fill out the simplified life history calendar, which spanned a period of over seven years. During the telephone interview, respondents would use the calendar as a visual recall aid.

At this point, we are not aware of any study describing computerized calendar instruments, which are self-completed by respondents. This is likely to be due to the fact that the current computerized versions are quite complicated to fill out, and cannot be used without training. Nonetheless, at least one electronic EHC has been used as a separate recall aid in a self-administered survey (Wiebe & Landis, 2000). Before the interview, interviewer and respondent filled out the EHC together. It was subsequently displayed on a laptop screen, during both, the CAPI and a short ACASI section of the interview.

From what is known so far about the relative effectiveness of these different calendar applications, it cannot be concluded that any specific modes of data collection are more suitable for using calendar applications than others. It is also not obvious if those instruments work better as recall aids or as data collection devices. There are some indications that respondents might be more motivated to make active use of the calendar if they can see the instrument during the interview. This suggestion is based on the observation that respondents prefer calendar techniques to regular question-list surveys in personal interviews (Freedman et al., 1988) but are largely indifferent to them in CATI surveys (Belli et al., 2001), where they cannot make use of their visual qualities.
3.6. Effects of calendar methods on data quality

As early as in the late 1960s Balán et al. (1969) concluded, in what was probably the first (non-experimental) evaluation of calendar methods, that the calendar instrument had the following advantages over traditional question-list surveys:

- It improved the completeness of reports by enabling the interviewer to detect 'gaps' in the data provided by the respondent.
- Inconsistencies in the account could be detected by the interviewer or by the respondent himself. The respondent could then correct his original account.
- It facilitated recall for distinct events, by displaying those events as part of a sequence. This (supposedly) led to a reduction of omissions.
- It improved timing of recalled events by allowing the respondent to relate events and dates from different life domains to each other.

Although the study by Balán and his colleagues did not have an experimental design, their observations are still valid today. The expected positive effects of calendar methods on completeness, consistency, recall and timing – as well as the implied effective mechanisms of the calendar – are main issues in evaluations of calendar methods.

Over the years several authors, though not explicitly referring to the Balán-study, have tested one or more of the four statements mentioned above. Many have also included more general observations about the data collection process, such as experiences with different modes of data collection (see previous section), respondent-interviewer rapport, and consequences for the duration of the interview. The body of research on calendar methods also includes a few psychometric studies on reliability and/or validity of data collected with calendar instruments.

Our review of the methodological literature reveals that the quality of the data collected with calendar instruments has been evaluated in multiple ways. The studies can be grouped into the following categories:
1. Comparisons of calendar data with similar data collected with more traditional questionnaires (in a split-ballot or otherwise), but without the availability of an external standard of comparison (Becker & Diop-Sidibe, 2003; Becker & Sosa, 1992; Engel, Keifer, & Zahm, 2001; Goldman et al., 1989; Yoshihama et al., 2005);

2. Studies in which the agreement between data collected with a calendar instrument and external data sources is measured. These studies can be divided into two subgroups. In the first subgroup of studies, agreement with external data is measured, but no comparisons are made with regular questionnaires. External data sources include physicians’ records (Rosenberg et al., 1983; Wingo et al., 1988) or reports from earlier waves of longitudinal surveys (Freedman et al., 1988). In the second subgroup, the authors assess the agreement between calendar data and external data, and also include a control condition, in which a traditional questionnaire is used (Belli et al., 2001; Belli et al, 2007; Van der Vaart, 1996, 2004; Van der Vaart & Glasner, 2007a).

We will start with the first category and present some findings based on indirect comparisons between calendars and traditional questionnaires; after that results from the ‘agreement studies’ will be presented.

**3.6.1. Indirect comparisons between calendar data and regular survey data**

The focus of the first group of studies is mainly restricted to indirect measures of data quality, in particular consistency of the data based on logical arguments (e.g. in most societies, there should be no overlap between marriages), completeness of the data (e.g. the detection of “gaps” in employment history) and patterns in recalled dates (such as the use of prototypical values, i.e. “heaping”). Since these studies do not include an external standard of comparison, they cannot provide direct evidence for the superiority of calendar data in terms of accuracy. However, as will be illustrated below, they do provide some indications
that the calendar method overall performs better in collecting recall data than the traditional question-list.

A split-ballot comparison between a calendar method and a traditional questionnaire in a fertility study (Becker & Sosa, 1992) indicated that the use of the calendar resulted in more consistent reports. It demonstrated that the calendar method resulted in less superposition of (supposedly) mutually exclusive behaviors: significantly less overlap of advanced pregnancy and contraception use was reported in the calendar condition (1.3%) than in the traditional interview (10.3%). Also supporting positive calendar effects, an interaction was found between the recency of the behavior and the effect of the calendar (Goldman et al., 1989). Goldman and her colleagues note that the calendar instrument was especially effective in enhancing recall of contraceptive use in the beginning of the reference period. A similar effect was found in a study of domestic violence victimization (Yoshihama et al., 2005). The results indicate that higher lifetime victimization rates in the calendar condition were caused by the fact that more respondents reported incidents, which took place in the distant past.

Studies that evaluated retrospective data in terms of completeness mostly conclude that the calendar method performs better than the traditional question-list. Calendars were found to be more helpful in reducing the amount of time unaccounted for in the respondent's life course (Engel, Keifer, & Zahm, 2001; Goldman et al., 1989). This reduction is likely to be due to the visual nature of the calendar, which makes it easier for the interviewer to detect those left-out periods and ask the respondent about them (Balán et al., 1969). Overall, calendars appear to result in higher numbers of reported events and episodes, which is usually interpreted as a positive effect (Becker & Sosa, 1992; Engel, Keifer, & Zahm, 2001).

Regarding the heaping of reported event dates - which occurs when respondents report prototypical values (e.g. courses starting in September, or “the accident happened two years ago”) instead of the actual values – only few studies are known to evaluate calendar effects (see also the next section). In an experimental evaluation Goldman et al. (1989) found that the calendar method significantly reduced heaping in reports of contraceptive use. While in the traditional questionnaire condition a disproportionate number of women rounded durations (measured in months) to prototypical values of 6, 12, 24, 36, and
48 months of use, this hardly occurred in the calendar condition. It should be noted however, that this difference was probably enhanced by the coding protocol. While in the questionnaire condition, interviewers could record durations in either months or years; in the calendar condition, interviewers were instructed to code durations in months always.

3.6.2. Agreement between calendar data and external sources

The second main category of evaluation studies focuses on direct assessments of agreement between the recalled information and the external information: in particular concerning the number of events, their characteristics and the duration or dates of events. Some authors turned to data sources such as doctors' records (Rosenberg et al., 1983), purchase records (Van der Vaart & Glasner, 2005) or population registers (Auriat, 1993) to validate the retrospective reports. In the absence of this type of validating information, authors compared calendar data with respondents' earlier (concurrent) reports from the same longitudinal study (Belli et al., 2001; Freedman et al., 1988; Van der Vaart, 1996). It can be argued that comparisons of the latter type are an assessment of (test-retest) reliability rather than of validity (Dex, 1995). Nevertheless, it seems reasonable to assume that the amount of error is smaller in concurrent than in retrospective reports, since the former are less affected by memory bias. As illustrated below, the results of these both types of studies generally suggest that the calendar method has beneficial effects on data quality.

Non-experimental validation studies

While non-experimental agreement studies do not compare the performance of the calendar method to the performance of other methods, they do give an indication of the quality of calendar data. In this line, Rosenberg et al (1983) performed a record check study, which did not include a comparison with another type of questionnaire. Using doctor’s records as validation measures the authors report an agreement of 90% between the calendar data and the records for month-specific use of oral contraceptives. The mean duration of the reference period was 33 months. The agreement between physicians' records and self-reports decreased when brand and dose of contraceptive were also considered.
High levels of data quality were also reported in non-experimental longitudinal studies. In their evaluation of calendar questionnaires Hoppin et al. (1998) report very high test-retest reliability of pesticide use when respondents were contacted by telephone one to three weeks after the original interview. A more detailed study of test-retest reliability of the calendar method - the time between the interviews being eight to fourteen months - resulted in very high agreement for reported life event anchors such as marriages, or immigration (Engel, Keifer, Thompson et al., 2001). Freedman et al. (1988) compared respondents' self-reports from two waves (1980 and 1985) of a longitudinal study. In the 1985 wave, a calendar instrument was used. The authors found an 87% agreement between school attendance reported concurrently in 1980 and retrospectively in 1985. Part-time school attendance was remembered less well than either full-time attendance or no attendance. Responses about work in 1980 were less consistent. Here, the agreement between waves was 72%. The general tendency to underreport unemployment in retrospective surveys was not fully compensated for by the calendar.

Thus, several life course studies that applied calendar instruments report relatively high correspondence between retrospective calendar data and matching responses or collateral reports obtained beforehand. Similar results are found in small-scale medical studies on health timelines (e.g. Searles, Helzer, & Walter, 2000). Although these results suggest positive effects of the calendar procedures on recall accuracy, they lack an experimental design: since there is no control condition, it has not been demonstrated whether these results would have been different in a study without aided recall procedures.

**Experimental studies**

So far, only four studies (Belli et al., 2007; Belli et al., 2001; Van der Vaart, 1996, 2004; Van der Vaart & Glasner, 2007a) have combined the use of external control data and comparisons between calendars instruments and regular questionnaires in an experimental design. The authors conducted split-ballot experiments in which they used a calendar instrument in one condition and a traditional questionnaire in the other condition. Belli et al. (2001; 2007) and Van der Vaart (1996) then validated the data from the two conditions with earlier reports from the longitudinal studies. Van der Vaart and Glasner (2007) used
purchase records as validation data. Given the relevance of these studies we will discuss their results in detail below.

In his 1996 study, Van der Vaart (1996; 2004) developed and tested a calendar method (in these studies called a ‘timeline’) that was filled out by the respondents during a face-to-face interview and subsequently was used as a visual recall aid. The calendar was tested in a field experiment on educational careers during the second wave of a longitudinal social survey, comparing the retrospective reports with reports during the first wave four years before (the recall period was four to eight years). As compared to the regular questionnaire procedure, adding the calendar enhanced data quality with respect to the number of educational courses followed, the starting year of the courses, and the entire sequence of types of courses taken. Although the calendar reduced recall error in the dates of courses, it did so for absolute error only. It did not affect telescoping (i.e., the direction of the net error in dates) and neither did it diminish the heaping effect in reported dates. The calendar was shown to be most effective if respondents had to perform relatively difficult retrieval tasks in terms of recency, saliency, and frequency of the target behavior (e.g., for respondents who had followed a great number of courses).

Comparable results were found by Belli et al. (2004; 2001) who evaluated an Event History Calendar by means of a field experiment integrated into a longitudinal household study on social and economic behaviors. All (CATI) interviews in that study were conducted in 1998 and elicited retrospective reports on the number and the duration of events that occurred in 1996. The quality of the 1998 reports using either a calendar - that was visible to the interviewer only – or a question-list, was assessed using data from the same respondents collected one year earlier on events in 1996. Compared to the question-list survey the calendar instrument resulted in significant difference scores, indicating positive effects, for three out of six topics: the number of moves, the number of jobs and the number of persons entered residence. No differences in data quality were found regarding the number of persons left residence, whether having received children aid and whether having received food stamps. Regarding four out of six continuous measures the calendar method led to significant higher correlations with the 1996 reports than the question-list. This applied to ‘income’, the durations of periods ‘being unemployed’, periods ‘missing work due to illness’, and periods ‘missing work due to illness of others’. No differences in
correlations were found for the duration of periods ‘working’ and periods ‘on vacation’. In spite of the effects on correlations, hardly any differences in mean errors were found between both conditions.

In a second methodological comparison that was implemented within the PSID panel, Belli et al (2007) compared a CATI Event History Calendar to a regular retrospective questionnaire with respect to a wide range of topics (marriage, cohabitation, residences, employment, unemployment, and smoking behavior) and for reference periods of up to 30 years. Again, the EHC led to better data quality with regard to some, but not all measures. While it increased agreement measures for employment histories and decreased overreporting of smoking behavior, the EHC did not improve retrospective reports of residential changes and led to decreased data quality with regard to the number of reported marriages. In the employment section of the interview, the EHC was especially effective in improving agreement measures for temporally remote employment spells.

Finally, the experimental record check study by Van der Vaart and Glasner (2007) generally confirmed the findings of the experiments presented above. In this study a calendar was employed as a visual recall aid for respondents in a telephone survey. Unlike most calendar instruments used in the social sciences, the calendar was designed to enhance the recall of singular events (the purchase of pairs of glasses) instead of episodes. The respondents’ retrospective reports on a recall period of over 7 years, were compared to database information with regard to three issues: the price and the date of the latest purchase of pairs of glasses and the number of purchases. Due to a lack of variance in the number of purchases no effects of the recall aid on this measure could be established. Regarding both the price and the date of the purchase this study demonstrated that:

1. The calendar had positive effects on recall accuracy, although it did not affect telescoping (net error in dates; see section 1.2.2).

2. A more difficult recall task - i.e. the respondent had to recall less salient and/or less recent events – led to greater recall errors.

3. Employing the calendar was especially effective in enhancing recall accuracy when the respondents’ recall task was relatively difficult in terms of saliency and recency, that is: for less salient and less recent purchases.
As will be discussed in more detail below, a downside of this procedure was that the response rate in the calendar condition was quite low. Sending respondents the calendar instrument beforehand probably increased the risk of refusal. Overall the results of these experimental studies - that compared the calendar method and the question-list method by using external validating data - are mixed but quite promising. They demonstrate that the positive effects of calendar methods on recall accuracy generally outweigh their potential negative effects.

### 3.7. Operational costs of calendar procedures: field-work and sampling

While calendar applications usually have beneficial effects on data quality, it is also important to judge their operational costs in terms of interview time, data entry, interviewer training, and so forth, as well as the effects on sampling. Regarding these issues several potential disadvantages of the use of calendar instruments emerged from the literature.

#### 3.7.1. Interview time and data entry

One rather consistent finding is that calendar instruments take longer to administer than traditional questionnaires. Engel, Keifer and Zahm (2001) report that personal interviews, in which the calendar method was used for data collection took nearly twice as long as the traditional question-list interviews. Belli et al (2007) found that CATI interviews in which an Event History Calendar was used were approximately 10% longer than CATI interviews with a conventional questionnaire. In a study in which the calendar was used as a recall aid, the face-to-face interviews in the calendar condition took twelve minutes longer on average than interviews without a calendar (Van der Vaart, 1996). The mean length of all interviews in that study was approximately two hours, which means that using a calendar recall aid increased interview durations by about 10%. When a similar calendar application was sent to respondents before a CATI interview, the interview took only marginally longer than in the control condition without calendars (Van der Vaart & Glasner, 2007a). How-
ever, respondents indicated that preparing the recall aid had taken them 12.5 minutes on average (with reported durations ranging from 5 to 30 minutes). As the mean interview duration in that study was only about 22 minutes, the calendar instrument increased the demands made on the respondents’ time substantially.

While the paper-and-pencil telephone administered calendar application evaluated by Belli et al. (2001) did not increase total interviewing time, data entry did take significantly longer than for regular questionnaires. Freedman and her colleagues (1988) made a similar observation and noted that their calendar data was more difficult and expensive to code than a conventional questionnaire.

### 3.7.2. Interviewer training

Estimates of the extra amount of interviewer training required for administering calendar instruments vary widely by study. Some studies have stated that training time was tripled compared to earlier waves of their survey, when no calendar instruments were used (Freedman et al., 1988). Eisner et al (2009) report that half a day of their two-day interviewer training dealt exclusively with the application of the Event History Calendar. Other authors report that interviewers who had to use calendar instruments received the same amount of training as interviewers who administered the traditional questionnaire (Belli et al., 2001; Van der Vaart, 1996; Van der Vaart & Glasner, 2007). For an international health study in which computer assisted EHCs were used, Dijkstra and Ongena (2007b) conducted three-day interviewer trainings. However, due to the complexity of the data gathered in this study, it is not unlikely that a question-list version of the calendar instrument would have required a similar amount of interviewer training (W. Dijkstra, personal communication).

### 3.7.3. Non-response

Only a few studies give some indications concerning the impact that calendar methods may have on response rates. Furthermore, the exact impact can be hard to determine since calendars sometimes are administered as just one of the elements of the data collection (for example during one section of an interview). This latter situation is found in the study by
Van der Vaart (1996), where the calendar was part of the second wave of a large-scale data collection process; apparently the calendar had no effects on the response rate (71% in this second wave). The study by Belli et al. (2001) involved a separate data collection for the calendar condition and the question-list condition and the response rates - that could be established independently – appeared to be equal (84%) in both conditions. Also in a study in which a calendar was used as a visual recall aid during a personal interview, no effect on the non-response was found (Kominski, 1990). However, the authors note that interviewers and respondents were not required to actively use the calendar.

A different picture emerges from a split-ballot consumer survey in which the calendar – which was used as a recall aid during the interview - was sent to the respondents prior to the telephone interview (Van der Vaart & Glasner, 2007a). This procedure led to a response rate that was substantially lower in the calendar condition (39%) than in the regular condition (67%). Quite possibly respondents were scared off by the extra work they had to put in when a calendar instrument was used. Yoshihama et al (2005) found an even larger difference in response rates between the calendar condition and the question-list conditions. In spite of the fact that both samples were selected from one (listed) operational population, using the same criteria, the response rate was only 18% in the calendar condition versus 78% in the control condition. All in all it seems clear that more studies are needed to determine potential effects of the use of calendar methods on response rates.

In conclusion, our findings indicate that researchers who are planning on using calendar instruments should be aware that this could affect operational costs in a number of ways. Several calendar instruments were found to increase interview durations; others increased coding time. Additionally, a number of studies report higher non-response rates in calendar interviews. The extent of the increase in interview durations relative to the total interview time depends on the complexity of the calendar instrument and the questionnaire. Obviously, asking respondents to fill out a very complex calendar recall aid, when the actual interview takes only a few minutes should be avoided, unless this is justified by a significant and relevant increase in data quality.
3.8. Evaluations of the interviewing process

Since the calendar method affects the questioning procedure and the tasks of both the interviewer and the respondent substantially, it probably has motivational effects next to memory effects. However, whether these motivational effects prove to be positive or negative is hard to say. On the one hand, a calendar procedure may have positive motivational effects since it is a less well-known method and asks for a more active approach of the interviewer and/or respondent. This may suggest to the interviewer and respondent that their task is important and needs precision, as is the case for aided recall methods in general (Sudman & Bradburn, 1983). But on the other hand, these very same characteristics might also cause fatigue, meaning that the task for the interviewer and/or respondents may be too burdensome (Billiet, Loosveldt, & Waterpas, 1984). While researchers usually appreciate the positive influence of calendar instruments on data quality, they also note that the application of a calendar method often increases the length of the interview and sometimes of the interviewer training time (see the previous section). However, reports of experiences with calendars as instruments for data collection are generally quite positive. This is especially true for interviewers’ and respondents’ subjective evaluations of the instrument (Martyn & Martin, 2003; Belli et al., 2007).

Results of evaluation surveys among interviewers who worked with calendar instruments suggest that these instruments are perceived as being interesting to work with (Belli et al., 2004, 2007; Freedman et al., 1988). Among others, the interviewers' preference for calendar instruments over regular questionnaires was attributed to the fact that inconsistencies in the data can be removed while the respondent is still available for clarification (Balán et al., 1969; Belli et al., 2001; Goldman et al., 1989). Generally calendar instruments were also perceived to yield better data quality than standard questionnaires. Interviewers felt that calendar instruments help respondents to understand questions better (Belli et al., 2001, 2007). Interviewers also thought that calendars made the interview "more enjoyable" for the respondent (Freedman et al., 1988). Similarly interviewers found that the calendar helped them discuss sensitive issues with their respondents (Martyn & Martin, 2003).
Interviewer perceptions that respondents prefer calendar instruments to traditional questionnaires are confirmed by respondents' feedback on the interviewing process. In the study by Martyn & Martin (2003), mentioned above, respondents reported that the calendar made it easier for them to discuss sensitive issues with the interviewer because they could refer to the information they had written down in the calendar. In addition they stated that the calendar did 'jog their memory' by helping them relate events from several life domains to each other. Respondents are also reported to enjoy filling in the calendar (Hoppin et al., 1998). In a comparison of a calendar-based questionnaire with a traditional questionnaire, respondents appeared to be more patient and cooperative in the calendar condition (Engel, Keifer, & Zahm, 2001). They were more concerned with data quality when a calendar was used, and sometimes even asked for a copy of the completed calendar to take home with them. Caspi and his colleagues (1996) note that, in a pilot study, respondents actively corrected the information in the interviewer-administered paper-and-pencil calendar. Literature on the use of satisficing strategies by respondents suggests that this higher involvement with the interviewing process and objective should have a positive effect on data quality by reducing satisficing (e.g. Holbrook, Green, & Krosnick, 2003).

The respondent feedback about calendar instruments was usually positive. This was especially true for interviews in which respondents could see the calendar and use it as a visual recall aid (e.g., Engel, Keifer, & Zahm, 2001; Hoppin et al., 1998). When calendars were used mainly as a tool for the interviewer, interviewer evaluations of the procedure tended to be more positive than respondent evaluations (Belli, 2000).

3.9. Conclusion and discussion

This review illustrates that applications of calendar instruments in social research have been growing rapidly during the last decades. Calendar instruments have been used in a wide variety of research fields and with very diverse populations, although there have been only two documented instances of calendar instruments being used in international (i.e. multi-site) life course surveys. Calendar instruments are used in personal as well as in telephone interviews, and they serve either purely as a recall aid or as an instrument of data collection. In some studies, respondents filled in calendars according to written instructions...
and without the assistance of an interviewer. Computerized versions of calendar instruments are available for both, CATI and CAPI applications, but not (yet) for self-completed web questionnaires.

Methodological studies on the use of calendar techniques show mixed results with regard to effects on data quality. Effects were found for some topics, such as educational courses and unemployment, but not for others. Sometimes those effects were strong, but mostly they were modest. Most calendar instruments were found to increase the completeness of respondents' accounts. This was especially true for the reduction of 'gaps', i.e. time unaccounted for. Beneficial effects on the accuracy of the number and characteristics of reported events were also reported. The available empirical evidence indicates that calendars tended to enhance the consistency of responses. Additionally, they were reported to lead to a reduction of dating errors, even though results on telescoping and heaping were mixed.

Furthermore, several studies reported interaction effects of the calendar application with the difficulty of the recall task, indicating that calendars might be especially helpful for recalling less recent, more frequent and less salient events. However, the number of methodological studies is limited and most of the supposed beneficial effects of the calendar reported in this review need more empirical evidence.

The operational costs of calendar instruments in terms of increased interview duration and additional interviewer training are not entirely clear yet. If differences between regular questionnaires and calendar instruments were reported, those differences usually indicated that the calendar increased interview duration and/or coding time. Calendars were also reported to increase interviewer training time in some cases, although it remains largely unclear if training times for the calendar interview were compared to ‘complete’ or to abbreviated interviewer trainings for the administration of regular survey questionnaires (see section 3.7.2). However, the magnitude of the effect varied strongly between studies. The interviewer response to calendar instruments is generally good, while respondent evaluations seem to depend on the degree to which respondents can see and experience the calendar. This ‘user evaluation’ is important. Since working with calendar instruments requires some extra effort, it is crucial to keep interviewer and respondent motivated to finish
the interview and ensure good data quality. Again, more research is needed and then may lead to ‘best practices’.

After studying the literature on calendar instruments, it remains largely unknown which kind of effects can be attributed to which different design characteristics. Most calendar instruments include public and/or personal landmark events as well as a grid, which provides visual feedback to respondent and/or interviewer. While public landmark events can be pre-specified in the calendar quite easily, collecting personal landmark events may take a considerable amount of time during the interview, as does recording the respondent’s answers in a visual grid and coding them after the interview. If we were able to identify the most effective components of calendar instruments, and the circumstances under which those components work best, we could limit the use of relatively ineffective cues and decrease operational costs.

When designing calendar instruments for international surveys, researchers will need to test how the central features of those calendars, spatial lay-out and landmark events translate to other cultural contexts. Apart from potential differences in preferences for reading direction and time units, there may be considerable variation across countries in the way in which respondents generate personal memory landmarks and use them as temporal anchoring points. More generally, the type of memory cues that lead to better data quality may differ not only between countries, but also between subpopulations within the same culture (i.e. between men and women, young and elderly respondents et cetera). Aided recall techniques that lead to more accurate retrospective reports for one group of respondents might not be very helpful for other groups.

Researchers should also be aware of the fact that in some instances, applying calendar procedures can be counterproductive. They may create consistency and completeness in the data that does not represent higher validity but biased reconstructions instead. This artifact may arise especially if respondents' behavior is less consistent than is assumed by the researcher. It is possible, for example, that women use contraceptive methods during pregnancy (Becker & Sosa, 1992), or that somebody holds a fulltime and a part-time job at the same time. Likewise, in one culture certain events or states might be mutually exclusive and thus reflect response inconsistency – e.g., multiple simultaneous marriages - while they
are not in another culture (Axinn et al., 1999). While in many cases removal of such inconsistencies will result in more valid data, it may lead to error in other cases.

Correspondingly, when it comes to 'time unaccounted for' in the respondent's life history, asking the respondent to 'fill in the gaps' might not always be a good decision. When the gap stands for an episode that the respondent cannot recall during the interview, he or she might stretch previous or subsequent episodes in order to reduce the amount of time unaccounted for, leading to decreased data quality (Auriat, 1993). Another potential drawback of calendar methods lies in their aforementioned potential negative effect on response rates, especially if the respondent perceives the instrument as time-consuming and cumbersome.

The literature review revealed that the growing interest in calendar methods has resulted in many different applications with many different names, which suggests that researchers do not take enough advantage of each other’s efforts. It also appeared that most applications are characterized by only a modest theoretical foundation.

A number of topics that have emerged from this review will be addressed in the empirical part of the dissertation. In order to further develop and adjust the rationale of the calendar methodology, the theoretical ideas that underlie calendar applications will be tested in a cognitive interviewing study. After that, we will conduct a methodological experiment in which the effects of landmark events and visual feedback on different aspects of data quality are measured in the context of a web-based retrospective life course survey.
4.1. Introduction

The use of calendar instruments originates in the research practice. Early reports suggested that calendar instruments improved data quality by increasing the completeness and consistency of retrospectively reported autobiographical episodes (Balán et al., 1969). Nonetheless, there was a clear lack of theoretical foundations for the new method. This situation changed when Belli (1998) linked the Life History Calendar to the structure of autobiographical memory, thereby providing a rationale for its assumed (and partially proven) effectiveness. Building onto the work of Conway (1992) and Barsalou (1988), Belli suggested that calendar instruments enhance autobiographical recall by offering the respondent sequential, top-down, and parallel cues (see section 3.2.2). While this assumption has often been cited and seems to be generally accepted by researchers who use calendars tools, it has never been confirmed in a systematic way (Belli, Stafford, & Alwin, 2009).

In the following sections, we intend to answer the question if date retrieval and data revision processes in survey interviews with a calendar recall aid differ from those elicited by regular retrospective question-lists without a recall aid (research question 2). This chapter presents the results of a think-aloud study that we conducted in order to gain more insights into the effective mechanisms of calendar instruments.
4.2. Aim of the study

Belli, Lee, Stafford, & Chou (2004) made a first effort to identify cognitive processes in real-life EHC interviews by coding interviewer and respondent verbal behaviors in 216 EHC interviews and 197 conventional question-list interviews. The authors found that respondents in the calendar condition were more likely to spontaneously use sequential and parallel dating strategies than respondents in the question-list condition. Moreover, positive associations were found between these recall strategies and data quality. Interviewer probes also differed between conditions with interviewers using more parallel and public holiday cues in the calendar condition. Nonetheless, in total numbers, far more top-down cues than EHC specific cues were used in both conditions.

Apart from this behavior coding study, few attempts have been made at finding empirical evidence for the theoretical ideas behind calendar instruments. Supposedly, this is due to the general difficulty of obtaining data on cognitive processes, which can obviously not be observed directly by a researcher. This means that we have to rely either on the observation of outwardly observable indicators of cognitive processes (as is the case in behavior coding studies) or on respondents’ verbal reports of these thought processes. While providing valuable observations of interviewer-respondent interactions and behaviors, behavior coding is limited to describing observable, outward behaviors. This means that insights into the respondent’s thought processes depend on the likelihood of the individual respondents spontaneously voicing their thoughts in a ‘natural interview setting’. Behavior coding studies might therefore miss important clues regarding the retrieval strategies used by respondents in response to calendar recall aids.

Cognitive interviewing, another tool that can be used for pre-testing survey questions, can often be a valuable addition to behavior coding. Unlike behavior coding, cognitive interviews are more of a ‘laboratory’ approach, and they are usually analyzed in a qualitative way. Their advantage over behavior coding is that respondents are asked explicitly for information regarding cognitive processes. We will use cognitive interviews to investigate which retrieval strategies survey respondents use when answering retrospective questions in interviews with regular question-lists and with calendar recall aids. In these
interviews, information on retrieval processes will be obtained by asking the respondent to think aloud while answering retrospective survey questions. Usually, cognitive interviews are focused on questionnaire design properties and aimed at detecting potential cognitive problems that respondents might experience during the interview (Willis, 2005). The goal of our current study was somewhat different from that of a regular pre-test. Rather than identify problems, we were aiming to describe the cognitive processes respondents engaged in during the retrieval of autobiographical events in survey interviews. We were interested in the types and numbers of cognitive processes involved in retrieving event dates, and in the potential differences in retrieval strategies between standardized retrospective interviews with and without a calendar recall aid. Our study adds to Belli et al’s research by shifting the focus from interviewer-respondent interactions to cognitive processes, and by limiting interviewer influences on retrieval strategies.

4.3. Method

Our study was carried out in cooperation with research institute CentERdata (Tilburg University, The Netherlands), who were also going to provide the research infrastructure for the methodological experiment described in chapter 5. We conducted 22 cognitive interviews in two conditions. In the first condition, we imitated the most conventional survey situation, in which the respondents fill in a retrospective question-list without any type of visual recall aid. This condition was meant to be a control group, which would give us an impression of the cognitive processes occurring in ‘regular’ retrospective surveys. In the other condition, a semi-standardized calendar instrument was used as a visual recall aid next to the retrospective question-list. In this condition, the subjects in our study also completed a standardized question-list, but they could use personal landmark events and visual timeline feedback of their earlier answers which were both recorded in the calendar during the interview as recall cues. We were aiming to imitate a survey situation in which respondents self-complete a web-based questionnaire, and a calendar recall aid is gradually building up on the computer screen, since that was the type of instrument we were going to use in the field experiment described in chapter 5 of this dissertation. This question-list plus calendar condition was further split into two subgroups, in which different calendar formats
were tested. In both of these subgroups, subjects were asked to provide any number of memorable events from the twenty-year reference period, that they could date with some confidence, in other words ‘time-tagged’ events that could be used as personal landmarks. Additionally, the subjects in one of the subgroups could consult a second landmark timeline with pre-specified public events from the reference period. Table 4.1 shows the number of subjects in each group.

Table 4.1: Number of subjects per condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Number of respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Question-list only</td>
<td>8</td>
</tr>
<tr>
<td>2. Question-list plus calendar</td>
<td>14</td>
</tr>
<tr>
<td>a) Personal landmarks only</td>
<td>(5)</td>
</tr>
<tr>
<td>b) Public and personal landmarks</td>
<td>(9)</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
</tr>
</tbody>
</table>

4.3.1. Sample

The subjects for this study were recruited from CentERdata’s existing database of web panel members, which was the same panel population that we used for the large scale experiment presented in chapter 5. CentERdata contacted 40 members of their web panel who lived in the Amsterdam area and had been very consistent in participating in earlier surveys. 37 of those 40 respondents indicated that they were willing to be included in the study. We invited 22 respondents to take part in our cognitive interviews and recruited the other 15 respondents for another pilot study described in chapter 5. All subjects were paid 25 Euros plus travel expenses for their participation in the cognitive interviews, which lasted about 45 minutes on average, but could take longer than an hour. Higher educational categories were overrepresented in our sample, which nonetheless included a number of subjects with little formal education. The mean age of our sample was 54.7 years (with a minimum of 34 and a maximum of 80 years). Unfortunately, we were not able to match
subjects in the two main conditions on background variables such as age or education. This resulted in a somewhat unequal distribution of age, gender and educational categories across the two conditions. The mean age of the respondents in the questionnaire only condition was 60.5 years, compared to a mean of 52.2 years in the questionnaire plus calendar condition. In the questionnaire only condition, 4 out of 8 respondents were female, against 9 out of 14 respondents in the question-list plus calendar condition. The main difference between the sub-samples lay in the representation of educational categories. Three out of eight subjects in the question-list only condition versus two out of fourteen subjects in the question-list plus calendar condition had only completed primary education or lower-level vocational education, while only one subject in the question-list only condition but eight of the subjects in the question-list plus calendar condition were college/university educated. The middle category (high school or medium level vocational education) was represented by four respondents in each of the categories.

4.3.2. Question-list

All respondents filled in a retrospective question-list, with a reference period of almost twenty years, starting from January 1st, 1987 and ending at the time of the interview, in November 2006 (see appendix 4.1). The experimental question-list covered four subject areas, which were chosen in cooperation with CentERdata, as these were life domains likely to be included in a retrospective life course interview for a new web panel. The chosen domains were housing, work, education, and health (split into ‘general health’ and ‘hospitalizations’). The question-list was structured in a chronological way, with respondents first answering questions about their status on the respective domains on January 1987 and then indicating if that status/situation had changed since then, i.e. if a life course transition had taken place. Subsequent to that question, the respondent was asked in what year and month the first transition took place. The same set of questions was repeated for each transition, until we reached the respondent’s present situation. While the first group of respondents (n=8) answered those questions without a recall aid, the second group (n=14) could use a 78 by 30 cm paper-and-pencil calendar instrument with nine horizontal timelines during the interview (see appendix 4.2). The order of the domains in the calendar was
identical to the question order in the standardized question-list. The calendar was filled in by the interviewer, while the respondent recorded his/her answers in the question-list.

All respondents in the question-list plus calendar condition were asked to provide any number of memorable events from the twenty-year reference period, that they could date with some confidence, in other words landmark events. Based on the findings of two landmark studies (Van der Vaart & Glasner, 2011; Van der Vaart & Glasner, 2007b) that instructions to report landmark events can steer the respondents in their choice of landmark type, we deliberately kept those instructions neutral. Respondents were encouraged to report any kind of event from the reference period, as long as it was “memorable” and could be dated with confidence. In the same study, we also found that respondents tended to report more events from the most recent part of the recall period. We included the instruction that if respondents only reported events from a limited period of time, the interviewer could ask them to report events from other periods, again without specifying the type of event.

4.3.3. Interviewing method

We used a think-aloud technique with relatively restrictive interviewer instructions concerning the use of verbal probes. Permitted probes were essentially restricted to non-directive general encouragements, such as “Would you please think-aloud while you answer the question?”, “Can you tell me what your thoughts were right now?”, or “How did you arrive at that answer?” Apart from those probes, the interviewer could use more general follow-up probes in order to stimulate the respondent to think out loud, such as humming, nodding, repeating or summarizing the respondent’s answer, or saying “Please go on, I’m listening”, (for an extensive overview of probing techniques in cognitive interviews see Snijkers, 2002).

In order to prepare respondents for the cognitive interview, the researchers asked two warming up questions, in which respondents could practice the think-aloud technique. The first one was a general question used in many cognitive interviews, which asks respondents to count the number of windows in their house or apartment and think aloud as they do so (Willis, 2005). The second question was more specific to retrospective life course
data, although it was deliberately chosen to relate to a subject (personal finances), which was not part of the interview: “When did you first open a bank account?” Since we were interested in the cognitive processes occurring during self-completed interviews, the subjects in our study were asked to fill in the question-list themselves. However, we decided to have the interviewer read out the questions to them. In that way we were hoping to stimulate the verbal exchange between interviewer and respondent, thereby motivating the respondent to use the think-aloud technique. The interviews were conducted in November 2006 by the author, one of her supervisors, and a research assistant. All interviews were recorded using digital voice recorders, and assistants transcribed the verbal protocols after the interview.

4.3.4. Coding scheme and method of analysis

As the main objective of our study was to identify cognitive subject processes involved in answering retrospective questions, our coding scheme was developed accordingly, with the great majority of codes focusing on the retrieval of dates of autobiographical events (for an overview of date retrieval and reconstruction strategies see section 2.2.2). The first category of codes denoted the retrieval of time-tagged events (see table 4.2).

Table 4.2: Time-tagged events

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date tag</td>
<td>a) Respondent indicates that he/she ‘just knows’ the date</td>
</tr>
<tr>
<td></td>
<td>b) Respondent mentions specific date (dd/mm/yy)</td>
</tr>
<tr>
<td>Age tag</td>
<td>Respondent uses his or her age as a dating cue</td>
</tr>
<tr>
<td>Recently retrieved</td>
<td>The event was retrieved recently and that is why the respondent still knows the date at the time of the interview</td>
</tr>
<tr>
<td>Copy</td>
<td>The event was mentioned as a personal landmark, and the date is copied from the landmark domain during the interview</td>
</tr>
</tbody>
</table>

In the second category of our coding scheme (see table 4.3), we specified codes that pertained to retrieval processes during which the subject used event order information to reconstruct event dates. This category included sequential, top-down, and parallel cues, as
well as cues taken from personal landmark events, public landmark events, and from specific events within the same thematic domain.

Table 4.3: Cues from relative order information

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequencing</td>
<td>Respondent explicitly refers to preceding or subsequent episodes as part of the dating strategy</td>
</tr>
<tr>
<td>Top-down</td>
<td>Higher-level information (especially lifetime episodes) is used to date a specific event</td>
</tr>
<tr>
<td></td>
<td>a) Shorter episode is mentioned as part of a longer episode (e.g. several jobs with the same company)</td>
</tr>
<tr>
<td></td>
<td>b) A specific date is located within a longer episode</td>
</tr>
<tr>
<td>Parallel domain event</td>
<td>Respondent uses cue from one of the other substantive domains, i.e. housing, work, education, general health and hospitalizations</td>
</tr>
<tr>
<td>Personal landmark</td>
<td>Respondents uses personal event from personal landmark domain or other personal event as temporal reference point for dating other events</td>
</tr>
<tr>
<td>Public landmark</td>
<td>a) Reference to public events pre-specified in the public event domain</td>
</tr>
<tr>
<td></td>
<td>b) Reference to public events mentioned in the personal/memorable event domain</td>
</tr>
<tr>
<td></td>
<td>c) Reference to public events not mentioned before</td>
</tr>
<tr>
<td>Same domain event</td>
<td>Respondent refers to specific event from the same thematic domain in order to date another event or transition</td>
</tr>
</tbody>
</table>

The third category of codes included cues that were taken from event details, such as weather, or psychological state (see table 4.4), the so-called phenomenological record.

Table 4.4: Cues from event details

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event detail</td>
<td>All temporal cues taken from lower level phenomenological information including physical circumstances, activities, feelings and moods and other phenomenological cues</td>
</tr>
</tbody>
</table>

If the subject mentioned retrieval processes that included durations, temporal distances between events (including recurrent events), or elapsed time since the event, this fell into the
fourth category of our coding scheme, ‘Retrieval cues from temporal relationships’ (table 4.5).

Table 4.5: Retrieval cues from temporal relationships

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal relationship</td>
<td>Subject mentions retrieval process that includes durations, temporal distances between events, or elapsed time since the event took place</td>
</tr>
<tr>
<td>Recurrent event</td>
<td>Annually or otherwise regularly recurrent event</td>
</tr>
</tbody>
</table>

Codes in the category ‘Estimates’ were assigned to general date estimates or guesses. Verbal reports of retrieval attempts fell into this fifth category if the subject indicated that he/she was not sure about the date but was willing to provide a rough estimate (‘I am not quite sure, but it must have been sometime in 1992. Don’t know why, just an estimate’) or if the event date could only be narrowed down to a season but not to a specific month (see table 4.6).

Table 4.6: Estimates

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>General estimate</td>
<td>Respondent indicates that he/she is not sure about the date but is willing to provide a rough estimate (‘I am not quite sure, but it must have been somewhere in 1992. Don’t know why, just an estimate.’)</td>
</tr>
<tr>
<td>Season</td>
<td>Respondent mentions only year and season, but either does not specify a month at all, or indicates that the month is arbitrarily chosen from the months within a season (e.g. ‘I know it was spring, so let’s just say April’)</td>
</tr>
</tbody>
</table>

Apart from coding retrieval processes, we were also interested in finding out how often subjects referred to the visual properties of the calendar when dating events and if the calendar stimulated respondents to revise their answers more often than they would do in regular survey interviews. Therefore we coded corrections and use of the calendar’s visual properties to date events in category six of our coding scheme (see table 4.7).
Table 4.7: Corrections and visual cues

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correction</td>
<td>Respondent corrects her/his earlier answer</td>
</tr>
<tr>
<td>Visual</td>
<td>Respondent explicitly uses the calendar’s visual features as a recall aid (e.g. “Mmh, there seems to be a gap… that cannot be right.”)</td>
</tr>
</tbody>
</table>

Finally, we specified a number of process codes, in order to identify potential interviewer influences on the subject’s thought process as well as on the subject’s adherence to the think-aloud protocol (table 4.8).

Table 4.8. Process codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Think-aloud</td>
<td>Interviewer encourages the respondent to verbalize their thoughts (e.g. “Please think aloud”, or “Please don’t forget to tell me what you are thinking while you answer the next question”)</td>
</tr>
<tr>
<td>Probe</td>
<td>Interviewer uses retrospective probe (e.g. “How did you remember that date?”). If multiple general probing questions were asked after each other, they were coded as separate probes</td>
</tr>
</tbody>
</table>

**Coding rules**

Except for the general process codes (interviewer probes, et cetera), only the subjects’ utterances were coded. We coded all utterances connected to the retrieval of dates, even if those utterances were made after the subject had already produced a date for the event. Codes were assigned to ‘utterances’, which were defined as ‘strings of text referring to a cognitive process’.

While the codes in the first three categories of our coding scheme were mutually exclusive (within as well as across the categories), all of them could be combined with a ‘temporal relationship code’. For example, if a subject mentioned that an event had taken place “two weeks after my 50th birthday”, the codes ‘personal landmark’ and ‘temporal relationship’ were assigned to this utterance. The codes in the category ‘estimates’ could not be combined with other codes. The ‘correction’ and ‘visual’ codes were not mutually exclusive and could also be combined with codes from other categories. If the subject reit-
erated a previous comment, the reiteration was not coded again, unless new components were added. Because utterances were the coding unit, mutually exclusive codes could very well be assigned to separate strings of text referring to the same retrieval attempt. Dating a change of residence, for example, could be coded as follows:

“I remember the date because I had just started seeing my ex-boyfriend again when I got the new apartment (utterance 1, coded as ‘personal landmark cue’)./ It was almost exactly a year after I got my job at the hospital (utterance 2, coded as ‘parallel event’ + ‘temporal relationship’)/ I remember because my season ticket was about to run out and I needed to change the address when I renewed it. The person at the counter made a mistake and it took her weeks to change it (utterance 3, coded as ‘event detail’).”

All interviews were transcribed by assistants and the transcripts were checked by the author. At this stage, one interview in the question-list plus calendar condition had to be discarded, because the interviewer had not appropriately encouraged the subject to verbalize her thoughts while retrieving the dates of events.

Software

The remaining interviews were coded in Sequence Viewer 4.4 (Dijkstra, 2008), a program best known for its application in behavior coding studies. Recently added text analysis features allowed us to assign text keys to our verbal protocols, and compare their occurrence across conditions. The program also enabled us to correlate text keys with a number of socio-demographic background variables, such as age, gender, and the number of life events and transitions that subjects reported during the interview.
4.4. Results

In total, we coded 642 verbal reports of retrieval strategies, 26 correction strategies, and 28 instances in which visual cues were used. Our 21 subjects reported an average of 30.6 retrieval strategies (min = 5, max = 91, SD = 23.9) and 1.2 corrections per interview (min = 0, max = 5). Since the average number of reported transitions for the twenty-year reference period was 9.9, this means that subjects tended to combine multiple cognitive strategies when trying to date autobiographical events. Subjects in the question-list plus calendar condition (M = 11.54, SE = 1.81) reported more transitions (i.e. status changes) than subjects in the question-list condition (M = 7.25; SE = 1.77), but this difference was not significant (t(19) = 1.585; p = .129) and became much smaller (10.44 versus 9.04) when we statistically corrected for age differences between the two groups. There were strong negative correlations between age and the number of reported retrieval strategies (Pearson’s r = -.589, p < .01), and between age and the number of reported transitions (Pearson’s r = -.794, p < .01). A point of concern was that the existing differences in mean age between the groups (60.5 versus 52.2 years; t(19) = 1.384, p = .182) could have biased the observed numbers of retrieval strategies. However, when entered into a multiple regression model, the number of reported transitions (t(18) = 3.950, p = .001) was a much stronger predictor of the number of reported retrieval strategies than age (t(18) = 0.608, p = .551). For this reason, we statistically corrected for the number of transitions during the analysis of differences in the types of retrieval strategies between the two conditions. We used an ANCOVA model with the experimental condition as the independent variable and the number of reported retrieval strategies as the dependent variable. The number of reported transitions was entered as a covariate. When the data was statistically corrected for the number of reported transitions, subjects in both conditions reported almost the same number of retrieval strategies (30.14 in the question-list only condition versus 30.84 in the question-list plus calendar condition). Table 4.9 shows the estimated mean numbers of the different types of retrieval strategies reported by subjects in both conditions.
Table 4.9: Reported use of retrieval strategies per interview by condition

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Question-list only (n=8)</th>
<th>Question-list plus Calendar (n=13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-tagged events</td>
<td>3.30*</td>
<td>3.58*</td>
</tr>
<tr>
<td>Event order strategies</td>
<td>14.71*</td>
<td>17.02*</td>
</tr>
<tr>
<td>Detail based strategies</td>
<td>6.65*</td>
<td>3.77*</td>
</tr>
<tr>
<td>Estimate</td>
<td>1.67*</td>
<td>2.43*</td>
</tr>
<tr>
<td>Temporal relationships</td>
<td>3.83*</td>
<td>4.11*</td>
</tr>
<tr>
<td>Total</td>
<td>30.14*</td>
<td>30.84*</td>
</tr>
</tbody>
</table>

(*Estimated means of reported retrieval strategies per interview, corrected for number of reported transitions)

The numbers of reported retrieval strategies based on event order, event details, guessing, and temporal relationships were all positively correlated with each other (p < .05). The retrieval of time-tagged events, on the other hand, was not significantly correlated with the use of any type of date reconstruction strategy.

We found no significant gender differences with regard to the reported use of the various types of retrieval strategies. Associations between age and the numbers of different types of retrieval strategies appeared to be confounded by age differences in the number of reported transitions.

In the following sections we will discuss the findings presented in table 4.9 in more detail.

### 4.4.1. Retrieval of time-tagged events

We found only small and non-significant differences between the conditions in the numbers of verbal reports of retrievals of time-tagged events. In both conditions, about one third of all transitions were time-tagged (= number of time-tagged transitions divided by the total number of transitions). Because the number of time-tagged events quite likely depends on the number of transitions, table 4.10 shows the average numbers of the four types of time-tagged events, corrected for group differences with respect to the number of reported transi-
tions. The adjusted means of the numbers of reported date-tagged events appeared to be virtually the same in both conditions. For the other subcategories in this group of codes, total numbers were quite small and did not differ significantly between conditions.

Table 4.10.: Time-tagged events

<table>
<thead>
<tr>
<th></th>
<th>Question-list only</th>
<th>Question-list plus calendar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date tag</td>
<td>2.27*</td>
<td>2.29*</td>
</tr>
<tr>
<td>Age tag</td>
<td>0.75*</td>
<td>0.39*</td>
</tr>
<tr>
<td>Recently retrieved</td>
<td>0.28*</td>
<td>0.44*</td>
</tr>
<tr>
<td>Copied landmark</td>
<td>-</td>
<td>0.46</td>
</tr>
<tr>
<td>Total</td>
<td>3.30</td>
<td>3.58</td>
</tr>
</tbody>
</table>

* Estimated means, corrected for number of reported transitions

On average, respondents in the question-list plus calendar condition determined 0.46 event dates by copying date information from the personal landmark they had provided earlier (i.e. the target event had been mentioned as a personal landmark).

4.4.2. Event order cues

Based on the theoretical rationale of calendar instruments (see section 3.2.2), subjects in the question-list plus calendar condition could be expected to report more use of event order cues than subjects who filled in the regular retrospective question-list. However, only non-significant differences in use of event order cues were found between the conditions (see table 4.11).
Table 4.11.: Event order cues

<table>
<thead>
<tr>
<th></th>
<th>Question-list only</th>
<th>Question-list plus calendar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequencing</td>
<td>2.70*</td>
<td>3.03*</td>
</tr>
<tr>
<td>Top-down</td>
<td>1.18*</td>
<td>1.50*</td>
</tr>
<tr>
<td>Parallel domain event</td>
<td>2.94*</td>
<td>3.50*</td>
</tr>
<tr>
<td>Personal landmark</td>
<td>3.25*</td>
<td>2.23*</td>
</tr>
<tr>
<td>Public landmark</td>
<td>0.60*</td>
<td>1.63*</td>
</tr>
<tr>
<td>Same domain event</td>
<td>1.96*</td>
<td>3.18*</td>
</tr>
<tr>
<td>Total</td>
<td>12.63</td>
<td>15.07</td>
</tr>
</tbody>
</table>

* Estimated means, corrected for number of transitions

Most of these relatively small differences between the conditions pointed in the expected direction, with subjects in the question-list plus calendar condition reporting slightly more sequencing, top-down, and parallel retrieval strategies, as well as more retrieval strategies involving specific events from the same thematic domain.

However, we found no evidence that subjects in the question-list plus calendar condition used more retrieval strategies based on dating cues from personal landmark events than subjects in the question-list condition (for a more detailed discussion of this finding see section 4.5). Furthermore, there was no association between the number of reported landmark events in the question-list plus calendar condition and the number of times that subjects reported using personal landmark-based retrieval techniques ($r = .039$, $n = 13$, $p = .898$). Subjects in the question-list plus calendar condition ($n = 13$) reported a total of 184 personal landmark events, varying from two to 71 events over the twenty-year reference period of the interview. Contrary to earlier results (Van der Vaart & Glasner, 2011; Van der Vaart & Glasner, 2007b), where a strong negative influence of temporal distance on the number of reported events was found, we noticed that the distribution of reported events over time was remarkably stable in our study, despite the long reference period. For the first decade of the reference period (1987 - 1996), 81 landmark events were reported, versus 103 landmark events for the second decade (1997 - 2006). This suggests that our instructions to report landmark events from all parts of the reference period were quite effective. Respondents tended to report older events first, often proceeding in forward chronological order.
Relationship events such as marriage, separation, or divorce were used as cues the most often, followed by vacations, birthdays, births and deaths, and references to other people’s health that had an impact on the subject’s life. 18 out of the 25 personal event cues used in the question-list plus calendar condition had been mentioned and recorded as personal landmarks by the subjects before the start of the interview. Even though this indicates that at least some of the reported landmark events were relevant to the retrieval task, we can tell from table 4.12 that the types of landmark events that were used as retrieval cues did not differ substantially between regular question-lists and question-lists plus calendar interviews.

<table>
<thead>
<tr>
<th>Event type used as cue</th>
<th>Question-list only (n = 8)</th>
<th>QL plus calendar (n = 13)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relationship events</td>
<td>5</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>Vacations</td>
<td>3</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Birthdays</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Births/Deaths</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Partner’s health</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Migration</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Own health</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>19</strong></td>
<td><strong>25</strong></td>
<td><strong>44</strong></td>
</tr>
</tbody>
</table>

Slightly different results were found with regard to reports of retrieval strategies involving public landmark cues. While subjects in the question-list plus calendar condition used more of those strategies than subjects in the question-list only condition (1.63 versus 0.60; F(1,18) = 3.00, p = .10), providing subjects with a timeline with public landmarks from the reference period did not increase that number any further. Only one out of a total of 26 reported public landmark-based retrieval strategies was based on one of the public landmarks specified in the calendar. More popular public event cues were related to cyclic events in institutional calendars such as school years, semesters et cetera (mentioned eleven times in the verbal protocols), or public holidays and festivals, most notably Christmas and New Year’s Eve (mentioned ten times). Apart from those categories, four subjects derived re-
trieval cues from changes in legislation or policy, which had directly affected their personal situation.

4.4.3. Retrieval strategies based on event details, guessing, and temporal relationships

Subjects in the regular question-list condition appeared to rely on retrieval cues from event details more often than subjects who used a calendar recall aid (6.7 versus 3.8 times). However, this difference was mainly caused by a small number of outliers in the question-list condition. Also, providing subjects with a calendar recall aid did not lead to a reduction in reports of general date estimates, i.e. guessing, which occurred 1.7 times per interview in the question-list only condition, and 2.4 times per interview in the question-list plus calendar condition. The use of temporal relationships between events as a dating strategy was equally unaffected by whether or not the subject could use a calendar recall aid ($F(1,18) = .01, p = .92$).

4.4.4. Visual cues

On average, subjects in the question-list plus calendar condition referred to visual information in the calendar 2.15 times per interview ($n = 13$, $SD = 2.27$), either when dating transitions or in order to correct earlier statements. There was a positive but non- significant association between the number of reported transitions and the number of times that subjects reported taking visual cues from the calendar ($r = .34, p = .13$). The use of visual cues was positively associated with sequencing strategies i.e. sequential retrieval of subsequent events from the same domain (see section 2.2.2) in particular ($r = .589; p < .05; n = 13$), but there was no significant correlation with any other type of retrieval strategy.
4.4.5. Data revision

Subjects in the question-list only condition corrected themselves less often than subjects who used a calendar but the difference was not statistically significant when the number of transitions was held constant (0.60 versus 1.63 corrections per interview; F(1,18) = 2.848, p = .109). Unsurprisingly, there was a strong positive association between the number of reported transitions and the number of corrections (r = .617, p < .01).

4.5. Conclusions and Discussion

In this chapter, we tried to answer the question if retrieval processes in survey interviews with a calendar recall aid differ from retrieval processes elicited by regular retrospective question-lists. We used the respondents’ concurrent verbalizations of their thought processes as an indicator of the retrieval strategies they used to answer the questions. Apart from a slight difference in the use of public events as retrieval cues, no significant differences were found in the reported use of direct date retrievals and of date reconstruction strategies based on event order, event details, and temporal relationships between events. The calendar recall aid also did not reduce the number of reported date guesses. All in all, the results of our comparison indicate that cognitive processes involved in answering retrospective questions could be remarkably similar for regular question-list interviews and for interviews in which calendar recall aids are used next to a regular retrospective question-list.

The most interesting finding of this study pertains to the subject of landmark events. The inclusion of a timeline with personal landmarks from the reference period is one of the main components of calendar instruments. However, no indication was found in the cognitive interviews that the calendar recall aid had an effect on the use of those personal landmark events as retrieval cues. Notwithstanding the fact that many subjects drew on personal event cues when reconstructing the dates of other events, we did not find any evidence that the personal landmarks specified in the calendar had an effect on the number of times that respondents used personal landmark events as retrieval cues. In fact, respon-
dents in the question-list only condition used more personal event cues than respondents in the question-list plus calendar condition.

Results were somewhat different for the use of public events as retrieval cues. Respondents in the question-list plus calendar condition tended to use more public event cues than respondents in the question-list only condition. However, the public event cues used during the interviews were not of the same type as the public events that were specified in the calendar. If public events were used as retrieval cues, those events tended to be either recurrent (such as public holidays) or to have had an impact on the subject’s personal situation (such as changes in labor law).

All in all, the results of our analysis point towards the possibility that the use of landmark events as retrieval cues may not depend on whether or not those events are specified in the calendar. Personal memories that are salient enough to be reported as landmark events might be ‘chronically’ available as retrieval cues, no matter if they are written down in the landmark domain of the calendar or not.

4.6. Limitations and suggestions for future research

An obvious limitation of this pilot study lies in the small sample size, which is not unusual for cognitive interviews, but led to a very low level of statistical power in our analyses. Nevertheless, many of the differences we found between interviews with and without a calendar recall aid were very small (and not always in the expected direction), which makes it unlikely that the lack of statistically significant results can be fully attributed to type II error. If feasible, similar studies should be conducted on larger samples.

When interpreting the results of our study, we need to take into account some potential influences of the research method (i.e. cognitive interviewing) on our findings. Compared to, for example, behavior coding, cognitive interviews are one of the more ‘intrusive’ methods of evaluating survey question-lists. Asking respondents to elaborate on retrieval strategies might lead them to think about answers more thoroughly than they would have done in a regular survey interview. In a non-laboratory situation, calendar instruments might stimulate the respondent to think harder about answers than they would in an interview with a regular question-list (either because landmarks are cognitive ‘warming-
up’ questions, or because the use of the relatively unusual questionnaire format has a motivating effect on respondents as it may be perceived as more ‘interesting’). It is possible that using cognitive interviews led us to underestimate differences between the two conditions, as the method would also cause more thorough cognitive processing in interviews with regular survey questionnaires (Willis, 2005). Unfortunately, there are very few alternatives to conducting cognitive interviews if we want to find about retrieval techniques used in survey interviews. As was mentioned in section 4.2, the informativeness of behavior coding studies depends on the respondents’ propensity to report their thoughts during a regular survey interview.

The results of this pilot study show again how important it is to gain more insight into the effective mechanisms of calendar instruments. Are the positive effects of the calendar method actually due to enhanced retrieval of events, or rather to the facilities for cross checking answers between domains? In that case greater accuracy and completeness of answers could be an effect of the visual feedback that calendar instruments provide to interviewers and respondents, which can be used to edit earlier answers or to check for overlap before filling in responses.

Further studies into the effectiveness of calendar instruments should focus on the question, which components of the calendar (domain grid, personal and public landmark events) are actually necessary to achieve improvements in the quality of retrospective reports. If it is possible to maintain the positive effects of the Life History Calendar while simplifying the method, these insights can be used in order to develop more cost-effective calendar instruments. We suggest three types of studies that could give us more insight into the mechanisms in which calendar instruments work. First of all, behavior coding can be used to study interviewer-respondent interactions in calendar interviews, active use of memory cues, and the effect of those behaviors on data quality. Secondly, it would be interesting to utilize eye-tracking techniques in order to determine if respondents actually use the landmark events and domain grids during self-administered online surveys. Finally, the current study should be followed up by large-scale methodological experiments in which different formats of calendar instruments are compared, so that the effects of the different components on data quality can be measured. In chapter 5, the results of such a large-scale experiment will be presented.
Chapter 5:
Evaluation of a calendar recall aid in a web-based life course survey

5.1. Designing a calendar instrument for an online panel

Calendar instruments are currently used in a wide variety of research fields within the social and medical sciences. So far, applications have been developed for personal as well as telephone interviews, and computerized versions of calendar instruments are available for both CATI and CAPI interviews, but not – to our knowledge – for self-completed web surveys. Although calendar instruments have been shown to increase the quality of retrospectively collected life course data in offline research, there is very little information about their applicability in web surveys.

The goal of this study was to investigate if calendar recall aids can also improve the quality of web survey data in general and - more specifically - to develop a cost-efficient recall aid for a life history questionnaire to be used in the Advanced Multi-Disciplinary Facility for Measurement and Experimentation in the Social Sciences (MESS) project, hosted by research agency CentERdata at Tilburg University. For that project, a fresh sample of approximately 8000 respondents within 5000 households was recruited from the Dutch population. In an early stage of the panel, respondents will be asked to provide retrospective information on several central life domains, such as health, education, work, housing, and family.

A couple of characteristics of the questionnaire and sample had to be taken into account when designing the calendar recall aid. First of all, the calendar instrument was going to be part of a web-based questionnaire. This means that the questionnaire had to be self-completed by respondents, so instructions had to be clear, and the instrument needed to
be easy to use without further clarification. Secondly, a web-based mode of data collection sets limits to the amount of information that can be clearly represented in the calendar as the instrument has to fit regular computer screens. Ideally, there should be no need for the respondent to scroll vertically or horizontally in order to see different parts of the memory aid. In cooperation with the software designer, we chose a horizontal layout over a vertical layout not only because most existing calendar instruments use the horizontal display, but also based on the results of a small-scale study conducted by Wiebe and Landis (2000). When pre-testing calendar instruments for a computerized questionnaire, the authors found that respondents slightly preferred horizontal display of the visual feedback, and were more likely to detect errors in this layout compared to a vertical design. Since the survey was aimed at gathering information about the respondents’ entire life course, the size of the calendar instrument had to be flexible and depended on the respondent’s age (see section 3.3.1).

Finally, the sample included respondents of all ages and educational levels. As we expected specifically some of the older and less highly educated respondents to have limited experience with computer use and the Internet, the calendar instrument had to be very easy to use.

As part of the development of the calendar recall aid for the new panel, we first carried out a pilot study among respondents recruited from CentERdata’s existing online panel.
5.2. Pilot study: Design of the questionnaire and calendar recall aid

The goal of the pilot study was to determine the ideal question order for the web-based interviews. We were specifically interested in two aspects, namely which order of the thematic domains (housing, health, work et cetera) respondents preferred, and whether they should be asked to report events in ‘forward’ or in ‘backward’ chronological order. We asked 15 subjects to enter information about their life course into a ‘blank’ calendar recall aid. By observing the way in which respondents completed the calendar if they were completely free to chose in which order they reported events from different thematic domains, we were hoping to gain information on how to design the questionnaire and recall aid in order to facilitate the recall processes in our retrospective life course interviews.

At the time of designing this pilot study, the final decision about the domains that were to be included in the eventual retrospective life course calendar had not been made yet, which led us to work with a number of thematic domains that had been used in calendar-aided life course surveys before, and were likely to be included in the MESS survey: housing, number of persons in household, general health, employment, and education. Also, the leaders of the MESS project had not decided whether the retrospective survey should cover the respondent’s whole lifetime, or a fixed reference period (e.g. ten or twenty years). For practical reasons regarding the layout of the test recall aid, we decided on a fixed reference period of twenty years for all respondents.

5.2.1. Sample

As we recruited the subjects for our pilot from CentERdata’s existing online panel (for details about the recruitment process see section 4.3.1), all of them had at least some experience with filling in web questionnaires. We interviewed eight male and seven female panel members, ranging in age from 33 to 79 years (mean age = 52.9 years). Eight of the fifteen
subjects held college and/or university degrees, which indicates that subjects with higher levels of educational attainment were overrepresented in our sample.

5.2.2. Method

The calendars we used covered a reference period of twenty years (1987-2006) and were each printed on two white A3 sheets (landscape format), which were attached to each other with adhesive tape, thereby creating a calendar sized 78 by 30 centimeters. Our calendar had no pre-specified order of domains. Instead, the domain names (general health, housing, hospitalizations, jobs, and education) were printed on paper labels, which were spread out next to the calendar in random order. The subjects could then choose which domain they wanted to talk about first and were free to go back and forth in time and between domains. Subjects were asked to provide any number of memorable events from the reference period, that they could date with some confidence, in other words ‘time-tagged’ events. Instructions were deliberately kept open with regard to what kinds of events the subjects were allowed to report, stating only that the event should be “memorable”.

As in the study described in chapter 4, we used a cognitive interviewing method, which combined think-aloud and verbal probing techniques (for a more detailed description see section 4.3.3). Subjects were asked to “think aloud” while answering the retrospective questions or filling in the calendar. If subjects had trouble verbalizing their thought processes or forgot to do so, interviewers used non-directive probes.

5.2.3. Results

We looked at the order in which the 15 subjects chose to report about the different thematic domains. Table 5.1 shows the mean and median rankings of the domains in the calendar. We also included the mean numbers of transitions reported for the respective domains, in order to gain more insight into whether the domains that subjects chose to complete first actually provided them with many recall cues that they could use during the rest of the interview.
Table 5.1.: Domain ranking and number of transitions

<table>
<thead>
<tr>
<th>Domain</th>
<th>Domain Ranking</th>
<th>Number of transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housing</td>
<td>Mean: 2.4</td>
<td>Mean: 2.1</td>
</tr>
<tr>
<td></td>
<td>Median: 2</td>
<td></td>
</tr>
<tr>
<td>Hospitalizations</td>
<td>Mean: 2.9</td>
<td>Mean: 0.7</td>
</tr>
<tr>
<td></td>
<td>Median: 2</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>Mean: 3</td>
<td>Mean: 2.8</td>
</tr>
<tr>
<td></td>
<td>Median: 3</td>
<td></td>
</tr>
<tr>
<td>Work</td>
<td>Mean: 3.3</td>
<td>Mean: 4.7</td>
</tr>
<tr>
<td></td>
<td>Median: 3</td>
<td></td>
</tr>
<tr>
<td>General health</td>
<td>Mean: 3.4</td>
<td>Mean: 2.1</td>
</tr>
<tr>
<td></td>
<td>Median: 4</td>
<td></td>
</tr>
</tbody>
</table>

**Housing**

Subjects tended to report their housing histories towards the beginning of the interview, i.e. among the first three domains. Only one subject chose housing as his fourth topic, and none of the subjects reported about it last. This finding suggests that housing is a central theme in subjects’ autobiographical memory and might be a good domain to start the interview with. Together with general health, it was also one of the two topics in our pilot that were applicable to all subjects at all times. However, most subjects reported less than two residential moves for the twenty year reference period, which means that the housing domain may not have provided a large number of retrieval cues.

**Hospitalizations/specific health events**

The majority of the subjects did not report any hospitalizations during the reference period. There was a tendency among subjects to either start with this domain or save it for last. This was true for subjects who had experienced hospitalizations (mean ranking = 2.8; median ranking = 2) as well as for subjects who had not been hospitalized (mean ranking = 3; median ranking = 2.5). Several subjects who had not experienced hospitalizations indicated that they started with that domain, because it was “an easy one”, which they wanted to get
out of the way before they started working on the more difficult sections. Nonetheless, during debriefing, at least one of the subjects regretted her choice of picking hospitalizations first: “[I should have started with] housing. (…) Because I connect many memories to where I lived at the time.” Even though general health and hospitalizations were thought to be closely related thematic domains, only 7 out of 15 subjects reported them in adjacent domains.

**Education**

In 10 out of the 15 interviews, work and education were placed in adjacent domains. However, even though most subjects preferred to fill in the questionnaire in forward chronological order (see below), only eight of the subjects reported education before work. We must note though, that due to the advanced age of some of the subjects in this pilot study, a substantial number of subjects (5) had no educational transitions to report for the 20-year reference period, which means that the results probably cannot be generalized to studies in which information about the whole life course is collected. Since full-time education typically precedes (full-time) employment, and our pilot study found that respondents are inclined to proceed through the questionnaire in forward chronological order, we suggest that questions about education should be asked before questions about employment.

**Employment**

In the employment domain of the calendar more transitions were reported for the 20-year reference period than in any other section. Age had a strong negative correlation with number of reported job transitions, as most of the older subjects had been retired for a while and reported relatively stable work situations before retirement. No correlation was found between age and ranking of the job domain. The majority of subjects chose to record their work history towards the middle part of the interview (as the third or fourth domain). There was a significant negative correlation between number of job transitions and domain rank, which might indicate that subjects who had changed jobs more frequently used the work domain as an organizing mechanism in their autobiographical memory. At first, this might seem counter-intuitive, especially given the earlier observation that some subjects liked to
start the interview with ‘easy’ questions. According to the theoretical rationale of the calen-
endar, more complicated life themes should be mentioned last, in order to make more use of memory cues from other domains. However, our interviews offer a possible explanation. Several subjects who had reported multiple job transitions told the interviewer that they had recently updated their resumes, and had therefore ‘rehearsed’ their own job histories, which made them easier to remember. This may have caused them to report their job history relatively early during the interview.

*General Health*

With regard to the general health domain, subjects again had the tendency to report about events either in the beginning or at the end of the interview. Those who had experienced many transitions in their perceived general health were inclined to report about those changes at a later point during the interview. We could not find any significant relationship between domain rank and number of transitions in this domain. We noticed that when asked for changes in their subjective health, many subjects reported that there had been a change, but still rated their general health the same before and after the change. This raises the question if general subjective health is indeed a thematic domain that can be assessed retrospectively, over a long period of time. It is possible that specific health events are reported somewhat more reliably. If researchers nonetheless decide to include retrospective questions about subjective health, general and specific questions (if represented on separate visual domains) should probably be kept together spatially, so that respondents can use specific health events as retrieval cues for changes in their general health status. Also, questions about health should be asked towards the end of the questionnaire, when more memory cues from other domains are available in the calendar.
**Forward versus backward order of completion**

Subjects were instructed to use the order of completion they felt most comfortable with. The calendar could be completed sequentially or non-sequentially, in forward or backward direction. The subject also had the opportunity to go back and forth between domains. The great majority of subjects (14 out of 15) chose to complete most of the calendar in forward sequential order. Usually, subjects would choose one of the domains, start at the beginning of the reference period, and fill in all episodes and events on that domain before moving on to the next domain.

**5.2.4. Conclusion and recommendations**

Based on the results of the pilot, we decided that the calendar domains in the methodological experiment should be structured as follows: Housing, Education, Employment, Specific health events, and General health. It remained unclear from our pilot if respondents can answer retrospective questions about their general health in a reliable and valid way. Secondly, the fact that the great majority of subjects chose to fill in the questionnaire in forward sequential order implies that standardization of recall order in the retrospective interview probably would not interfere too much with respondents’ spontaneously used retrieval strategies. The tendency to report events chronologically might stem from the convention to tell (life) stories in chronological order, and might be especially pronounced for thematic domains that respondents are used to relate to others in that way (e.g. residential or educational history). Our results indicate that survey respondents may have reasonably separate (often causally related) narratives for each life domain, and use the other domains and life events mainly for cross-checking the timing of specific episodes (see also chapter 4).
5.3. Study aim and hypotheses

We were aiming to develop an optimized calendar instrument for a new online panel by identifying the recall aid’s most effective components and eliminating redundant or ineffective ones (see section 1.5, research questions 3 and 4). Calendar instruments combine two important elements, namely temporal bounding cues (or ‘landmark events’) and visual feedback of the respondent’s answers (i.e. domain ‘timelines’). We evaluated the effects of adding those elements to our questionnaire with regard to response/break-off rates, completeness and consistency of retrospective reports, interview duration, data editing, and respondent evaluations. Based on the literature review and the cognitive interviews we expected to find a number of effects that will be discussed in the following subsection.

5.3.1. Break-off rates

There is little information about the effects of visual recall aids on break-off rates in web-based surveys. As has been mentioned above, generally speaking, there seems to be little effect of visual elements on break-off rates in web surveys (Couper, 2008). However, calendar instruments have occasionally been found to increase non-response rates in offline surveys (e.g. Van der Vaart & Glasner, 2007a).

Hypothesis 1: Break-off rates in web surveys with a calendar recall aid will differ from break-off rates in web surveys without a calendar recall aid.
5.3.2. Completeness of retrospective reports

By lack of control information against which it would be possible to check the accuracy of the numbers of retrospectively reported episodes, the completeness of reports will be measured in terms of the number of reported events in the three major thematic timelines of the questionnaire. Since we deemed it unlikely that respondents would overreport employment episodes, relationships, and residences, data quality was thought to be better if the respondent reported more events.

Hypothesis 2a: In retrospective web surveys with a calendar recall aid, respondents will report higher numbers of specified events and episodes than in retrospective web surveys without a calendar recall aid.

Additionally, we will determine if the calendar was especially effective in improving reports of events that could be classified as being difficult to remember. Van der Vaart (1996) distinguished recency, frequency, and saliency of target events as components of task difficulty. He found that his calendar recall aid had a greater positive effect on the retrieval of ‘difficult’ (i.e. less recent, more frequent, and less salient) events than on the retrieval of ‘easy to remember’ events. Two other evaluations of paper-and-pencil calendar instruments (Goldman et al, 1989; Yoshihama et al, 2005) found that those recall aids specifically increased reports of temporally remote events, i.e. events that took place towards the beginning of the reference period. In the following, we will focus on temporal distance as an indicator of recall task difficulty.

Hypothesis 2b: In retrospective web surveys with a calendar recall aid, respondents will report higher numbers of temporally remote events than in retrospective web surveys without a calendar recall aid.
5.3.3. Data consistency

Offline calendar instruments have been shown to increase the chronological consistency of retrospectively reported data. We formulated the following hypotheses:

Hypothesis 3a: In retrospective web surveys with a calendar recall aid, respondents will make fewer chronological errors (in terms of event order and consistency of start and end dates) than in retrospective web surveys without a calendar recall aid.

Hypothesis 3b: In retrospective web surveys with a calendar recall aid there will be lower numbers of inconsistencies between reported and inferred episode durations than in retrospective web surveys without a recall aid.

5.3.4. Interview duration

In offline research, interviews in which calendar recall aids were used tended to take more time than interviews without a calendar recall aid (Glasner & Van der Vaart, 2009). It is likely that part of this increase in interview duration was due to the additional information about landmark events that the respondent had to provide in calendar interviews (i.e. additional questions), and that another part was due to having to fill in the calendar grid (i.e. visual feedback). In computerized surveys, the visual answer feedback can be provided automatically by the survey software and should not increase interview duration. We therefore expect that using calendar recall aids in web surveys will only affect interview durations if the respondent has to provide personal landmark events.

Hypothesis 4: The use of personal landmark events in retrospective web surveys will increase interview durations.
5.3.5. Data revision

Collecting our data in a web-based survey enabled us to collect survey paradata on the number of times our respondents used the survey’s ‘back’ button and how often they edited their answers to earlier questions. In the cognitive evaluation described in chapter 4, we found that respondents who used a calendar recall aid edited their responses more than twice as often as respondents who completed a regular retrospective questionnaire without a recall aid. Even though the difference (1.6 versus 0.6 edits per interview) was not significant in our small sample, it was in line with our expectations from the literature review, so we take this result as an indication that calendar instruments may indeed stimulate respondents to review and - if necessary - edit their answers to the questionnaire.

Hypothesis 5: Respondents who are provided with a calendar recall will edit their responses more often than respondents who are not provided with a calendar recall aid.

5.3.6. Respondent evaluations

In addition to the data quality measures mentioned above, our study comprises respondent evaluations of the method. In our review of the literature on calendar instruments (see chapter 3 of this dissertation), we found that respondent evaluations of (offline) calendar methods were generally positive. We expect similar evaluations for the web-based recall aid.

Hypothesis 6: Respondents will prefer web questionnaires that include a calendar recall aid to web questionnaires without a calendar recall aid.
5.4. Method

The goal of our evaluation study was to optimize the recall aid by measuring the effects of landmark events, visual feedback, and the combination of these two elements on data quality in retrospective web surveys. We conducted a methodological experiment with a 2 x 2 factorial design. When opening the questionnaire, the respondents were randomly assigned to either the control group or one of the three experimental groups (see figure 5.1).

<table>
<thead>
<tr>
<th>Landmark events</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
</tr>
<tr>
<td>Visual feedback of answers</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
</tr>
</tbody>
</table>

Figure 5.1: Study design

The retrospective questionnaire and the recall aids, which were used in this experiment, were developed in cooperation with CentERdata taking into account the results of the pilot study discussed in section 5.2.

5.4.1. Experimental design

Respondents in the experimental conditions read an introduction to the recall aid they were going to see during the interview (translated from Dutch):

“We use a calendar with this questionnaire. At the top of the calendar you can see 10-year periods with year numbers and below that you see your age in that year. We did not take into account the month in which you were born (otherwise there would be two ages for each year in the calendar for everybody who was not born on January 1st). When you answer the questions, colored boxes will appear in that year of the calendar for some responses. In the top left corner of the calendar, you can find a
magnifying glass with a "+" and a "-". If you click on a specific 10-year period, a colored border will appear around that period. If you then click on "+" you will see that period magnified on the screen. You can see the complete calendar again by clicking on "-".

Respondents in the visual feedback only condition would see their answers to some of the questions in the children, relationship, accommodation and job sections of the questionnaire displayed in an event grid that was shown on top of the questions in each screen of the interview (see figure 5.2). The specific questions for which visual feedback was provided were the names, years of birth and, if applicable, death of children, start and end years of relationships, accommodation episodes, and jobs and some additional information about each of those episodes.

Figure 5.2: Visual feedback grid

The size of the recall aids (in years) was based on preloaded year of birth data collected during earlier interviews.

Respondents in the landmarks only condition were not shown any visual feedback of the answers they gave during the main interview, but only saw a simple timeline with their age and the landmarks which they had filled in before starting the main questionnaire (see figure 5.3).

Figure 5.3: Landmark timeline

Respondents in the fourth condition (visual feedback plus landmark events) saw the landmarks displayed in the top row of the answer grid (labeled ‘gebeurtenissen’), a combination of the two recall aids used in conditions two and three (see figure 5.4).
After reading the introduction screen, respondents in the landmark conditions were asked to report six memorable events from their lives. This number of landmark events was arbitrarily chosen, based on what the researchers imagined would be reasonable to ask of respondents. The instruction read as follows (translated from Dutch):

“First of all, we would like to ask you to report six events from your own life, which you can remember well and of which you still know in which year they happened. Would you please try to spread those events over your life as much as possible? We start with an important event that you would like to report first.”

Although it was possible to leave the answer to each landmark question blank by clicking on a “forward” button, this was not explicitly pointed out in the instructions. Every time a respondent would skip a landmark question (or any other question) they received a reminder to fill in an answer. Every question about landmark events and their dates had to be skipped individually.

All respondents, including those in the control condition in which no recall aid was used, then read a short introduction to the questionnaire:

“This questionnaire is about your life course. We will ask you about the children you have (had), your relationships, your jobs, residential moves, your health, and a few extra questions about certain time periods in your life. We will start with the children that you may have, or may have had”.

In the experimental conditions, the recall aid was displayed in this introduction screen.
5.4.2. Questionnaire

The questionnaire we used in the methodological experiment was designed by CentERdata, taking into account our recommendations from the pilot study with regard to domain order and forward order of recall. Unlike the experimental questionnaire we used in the pilot, the new questionnaire also included questions about children and relationships, which were asked early in the interview. The retrospective life course questionnaire was completed online. The first section of the questionnaire assessed the number and birth years of biological and adopted children as well as duration of and income during maternity leave. The second section included questions about the start (and end) dates of significant partner relationships, gender of the partner, reasons for separation, and whether or not the respondent and her/his partner were living together during the relationship. In the third section, we assessed residential history starting from the time the respondent left his/her parents’ household with regard to locations, dates, types of residence, and the prices of owned properties. The fourth section covered the respondent’s employment history, i.e. year of graduation, occupation(s), industry, start and end dates of employment episodes, full-time or part-time employment, reasons for working part-time, reasons for leaving jobs, occurrence of employment gaps, and sources of income during those employment gaps. In the fifth section, respondents answered a number of questions about their health history, which covered general health, illness and hospitalization during childhood, menstrual history, physical injuries and disabilities, periods of ill health and type of illness, work and finances during periods of illness, and long term effects of health problems. The sixth section included some general questions about periods of happiness, stress, poor health, financial hardship, and discrimination. In the final section, respondents were asked which sources of information they had used while filling in the questionnaire, and how certain they were of their answers to three items (year moved from first to second residence, year finished full-time education, and year finished first job). Respondents in the experimental conditions were also asked if they preferred the questionnaire they had just filled in to traditional online questionnaires. The sections about children, residential history and job history were completed in forward chronological order. Questions about important personal relationships could be answered non-chronologically.
5.4.3. Sample

In cooperation with CentERdata, 2251 invitations were sent to members of the institute’s existing online panel in March 2009. 1887 respondents subsequently logged in on the survey, which equals an initial response rate of 83.9%. Since we thought there was a risk that the questionnaires for multiple members of the household had been filled in by the same person, we selected only those respondents who were the first person in their household to complete the questionnaire. There were some significant differences between those primary respondents and those who filled in the questionnaire after somebody else in the household had already done so. Primary respondents were more likely to be male than secondary respondents (57.5% versus 38.8%; $\chi^2 = 47.518; \text{df} = 1; p < .001$). The selection of primary respondents left us with 1451 respondents who were randomly allocated to the four conditions when they opened the link to the questionnaire. For all respondents who participated in the study, we had access to a number of socio-demographic background variables, such as age, gender, income, and level of education (see table 5.2).
Table 5.2: Sample descriptives

<table>
<thead>
<tr>
<th></th>
<th>Control (n = 357)</th>
<th>Visual feedback (n = 356)</th>
<th>Landmarks (n = 367)</th>
<th>Visual FB &amp; Landmarks (n = 371)</th>
<th>Total (n = 1451)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>44.3% (158)</td>
<td>41.6% (148)</td>
<td>42.0% (154)</td>
<td>42.6% (158)</td>
<td>42.6% (618)</td>
</tr>
<tr>
<td>Male</td>
<td>55.7% (199)</td>
<td>58.4% (208)</td>
<td>58.0% (213)</td>
<td>57.4% (213)</td>
<td>57.4% (833)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean age (in years)</td>
<td>53.11 (SD = 14.99)</td>
<td>53.75 (SD = 14.67)</td>
<td>52.20 (SD = 13.86)</td>
<td>53.82 (SD = 14.39)</td>
<td>53.22 (SD = 14.48)</td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>43.0% (153)</td>
<td>43.0% (153)</td>
<td>41.7% (153)</td>
<td>37.7% (139)</td>
<td>41.3% (598)</td>
</tr>
<tr>
<td>Medium</td>
<td>31.7% (113)</td>
<td>27.2% (97)</td>
<td>30.5% (112)</td>
<td>30.6% (113)</td>
<td>30.0% (435)</td>
</tr>
<tr>
<td>Low</td>
<td>25.3% (90)</td>
<td>29.8% (106)</td>
<td>27.8% (102)</td>
<td>31.7% (117)</td>
<td>28.7% (415)</td>
</tr>
<tr>
<td><strong>Household income</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; € 2600</td>
<td>45.1% (161)</td>
<td>46.6% (166)</td>
<td>43.6% (160)</td>
<td>42.6% (158)</td>
<td>44.5% (645)</td>
</tr>
<tr>
<td>€1801-2600</td>
<td>31.1% (111)</td>
<td>29.2% (104)</td>
<td>27.2% (100)</td>
<td>30.2% (112)</td>
<td>29.4% (427)</td>
</tr>
<tr>
<td>€1151-1800</td>
<td>17.1% (61)</td>
<td>18.0% (64)</td>
<td>20.2% (74)</td>
<td>21.0% (78)</td>
<td>19.1% (277)</td>
</tr>
<tr>
<td>&lt; € 1150</td>
<td>6.7% (24)</td>
<td>6.2% (22)</td>
<td>9.0% (33)</td>
<td>6.2% (23)</td>
<td>7.0% (102)</td>
</tr>
</tbody>
</table>

Randomized allocation to the four conditions appeared to have been effective as we found no significant differences between the conditions with regard to gender ($\chi^2 = .616; \text{df} = 3; p = .893$), income ($\chi^2 = 6.488; \text{df} = 9; p = .690$) and mean age ($F(3, 1447) = .990; p = .396$). Respondents on the lowest level of educational attainment seemed to be slightly underrepresented in the conditions without visual feedback, particularly the control condition, but those differences were not significant ($\chi^2 = 5.878; \text{df} = 6; p = .437$).
5.4.4. Data and operationalizations

The experiment had a two-by-two design, with personal landmarks and visual feedback as independent variables. Both independent variables were coded dichotomously (1 = present/0 = not present).

Control variables

Even though the differences between conditions with respect to age, gender and education were not significant, we thought that those differences could nevertheless (slightly) affect the results of significance tests of the effects of the independent variables. Therefore age and gender were entered as covariates, and educational level was included as a third factor. We recoded the education variable in our dataset from the original six educational categories (‘basisonderwijs’, ’vmbo’, ’havo/vwo’, ’mbo’, ’hbo’, and ’wo’) into three categories: low (‘basisonderwijs’ and ‘vmbo’), medium (’havo/vwo’ and ‘mbo’), and high (’hbo’ and ‘wo’). Age was measured in years, and gender was a dichotomous variable (male = 0, female = 1).

Dependent variables

Break-off rates

Break-off was measured dichotomously. All respondents who had completed the last question of the web questionnaire were coded as ‘1’ (or ‘finished interview’); all others were coded as having broken off the interview (‘0’ = ‘did not finish interview’). The latter group included respondents who did not complete any items as well as respondents who broke off the interview at a later point.

Completeness of retrospective reports

Using the definitions from the questionnaire provided by CentERdata, our measure of data completeness, the number of reported episodes, was calculated for all long-term accommodation episodes (> 6 months), reported relationships, long-term employment episodes (> 6
months), long-term unemployment episodes (> 6 months), and long-term family leave episodes (> 6 months).

Unemployment episodes and long-term family leaves were coded as follows: Respondents were asked to report all jobs, which they had held for more than six months since finishing full time education in forward chronological order. After each completed employment episode, respondents were asked if they had held another job. If that was the case, the next question was “Did your next job start immediately after you quit your job as [TITLE OF JOB] or was there a period of more than six months before you started your next job?” Respondents who indicated that their new job had not started immediately were then asked “How would you describe your situation before you started your next job”? Response options included ‘unemployed’, ‘retired’, ‘temporary job’ (less than 6 months), ‘sick’, ‘homemaker/taking care of family’, and ‘military service’. We examined the two largest categories, ‘unemployment episodes’ and ‘homemaker/taking care of family’ in our analyses. We computed the sum variables and ‘number of unemployment episodes’ and ‘number of family leaves’.

Data consistency

Inconsistencies in reported start and end dates (i.e. end dates that preceded start dates) and inconsistencies in event order (i.e. the second job has an earlier starting date than the first job, or the birth of the second child precedes the birth of first child et cetera) were coded as chronological errors. Inconsistencies in event order were only computed for events that had to be reported in chronological order, namely birth and adoption of children, employment episodes, and accommodation episodes. Inconsistencies in start and end dates were computed for all accommodation, employment, and relationship episodes. We computed the sum variable ‘number of chronological errors’. Due to the low total number of errors, we also computed the dichotomous variable ‘chronological consistency error’ (1 = ‘one error or more’/0 = ‘no errors’).

Furthermore, we examined the number of inconsistencies between reported and inferred episode durations by using data from the accommodation domain. For each completed residential episode, the respondent was asked “In which year did you stop living at [address X]?” This was followed by the question “Did you move into your next residence
the same year?” If the answer to the latter question was ‘no’, the respondent then had to report in which year they had moved to their next residence. We selected only those respondents who reported to have lived in more than one residence (n = 1073) after moving out of their parents’ home. If a respondent answered the question “Did you move into your next residence the same year?” with ‘no’, and then indicated at the next question that he/she had started living at their next residence the same year they had moved out of their previous home, this inconsistent answer pattern was defined as an ‘overreported short accommodation episode’. We computed the sum variable ‘number of overreported short accommodation episodes’.

**Interview duration**

The duration of the interview was calculated as the number of seconds between the moment at which the respondent opened the questionnaire, and the moment when the respondent completed the last question.

**Data revision**

Calendar instruments enable interviewers and respondents to revise their answers based on the visual feedback in the calendar. During our experiment, we collected paradata on the number of times respondents clicked on the ‘back’ button, and on the number of times they changed their earlier answers during the interview. Our comparison between the experimental conditions excluded the changes respondents in the landmark conditions made during the first part of the interview when they had to report the landmark events.

**Respondent evaluations**

As a measure of respondent feedback about the method, respondents in the experimental conditions, who had filled in a questionnaire with visual feedback, landmark events, or both recall aids, were asked if they preferred the type of questionnaire they had just completed to the type of questionnaire they were used to filling in as members of an online panel. In both visual feedback conditions the instruction started with “You have just completed a survey in which your answers were displayed in a calendar”, whereas in the condition with only
landmarks, the instruction started with, “You have just completed a survey in which you could use a timeline of personal events as a recall aid”. In all three conditions, the instruction continued with “If we were going to conduct a similar survey in the future, would you prefer this new format, or would you rather see the questions displayed in the usual way?” The respondents’ preference was coded as a dichotomous variable (1 = prefers new format/0 = does not prefer new format).

5.4.5. Analyses

We used a 2x2x3 factorial ANCOVA to test the effects of landmark events and visual feedback on the continuous outcome measures. The ANCOVA model included the main effects for landmarks, visual feedback, and education level, the two-way interaction effect between landmarks and visual feedback, and the covariates age and gender.

In order to examine the effects of our independent variables on the dichotomous outcome variables we performed logistic regressions in which we used landmarks, visual feedback, the interaction term ‘landmarks x visual feedback’, education, age, and gender as the predictor variables.
5.5. Results

5.5.1. Break-off rates

We tested the hypothesis that break-off rates in web surveys with a calendar recall aid will differ from break-off rates in web surveys without a calendar recall aid (hypothesis 1). In total, we found a break-off rate of 14.7% during the course of the interview. This dropout of respondents was especially pronounced in the conditions in which respondents were asked to report landmark events before the interview. Most of the dropout in these landmark conditions occurred during the first part of interview, when respondents had to fill in the personal landmark events. 13.1% of the respondents in the ‘question-list plus landmarks’ condition and 17.5% of the respondents in the ‘question-list plus landmarks plus visual feedback’ did not complete any of the landmark questions. Another 5.2% and 7% respectively stopped the questionnaire later during the landmark section. The high dropout rate in condition 4 indicates that the combination of visual feedback and landmarks had a detrimental effect on response rates and that there might be a possible interaction effect between the two components.

We undertook a logistic regression analysis with ‘finished interview (yes/no)’ as the outcome variable (see table 5.3). Having to report landmark events had a strong negative effect on response, while the effect of visual feedback was negligible. In this model, no significant effect of the interaction term ‘landmarks x visual feedback’ was found. Age was a significant predictor of finishing the questionnaire (older respondents were less likely to drop out), while being in the lowest education category and/or the lowest net income category decreased the odds of finishing the questionnaire significantly. Female respondents were more likely to finish the questionnaire than male respondents.
Table 5.3: Logistic regression of drop out

<table>
<thead>
<tr>
<th></th>
<th>Exp. B = OR of finishing (1 = finished interview)</th>
<th>95% Confidence interval of Exp. B</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Design elements</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landmarks (1 = present)</td>
<td>0.342</td>
<td>0.21 – 0.56</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Visual feedback (1 = present)</td>
<td>1.028</td>
<td>0.58 – 1.84</td>
<td>p = .925</td>
</tr>
<tr>
<td>Landmarks x Visual Feedback (interaction effect)</td>
<td>0.62</td>
<td>0.31 – 1.23</td>
<td>p = .17</td>
</tr>
<tr>
<td><strong>Gender (1 = female)</strong></td>
<td>1.56</td>
<td>1.13 – 2.15</td>
<td>p = .007</td>
</tr>
<tr>
<td><strong>Age (in years)</strong></td>
<td>1.02</td>
<td>1.01 – 1.03</td>
<td>p = .001</td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>0.85</td>
<td>0.58 – 1.24</td>
<td>p = .396</td>
</tr>
<tr>
<td>Low</td>
<td>0.57</td>
<td>0.38 – 0.84</td>
<td>p = .004</td>
</tr>
<tr>
<td><strong>Household income</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; € 2600</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>€1801-2600</td>
<td>0.95</td>
<td>0.66 – 1.38</td>
<td>p = .787</td>
</tr>
<tr>
<td>€1151-1800</td>
<td>1.05</td>
<td>0.68 – 1.62</td>
<td>p = .840</td>
</tr>
<tr>
<td>&lt; € 1150</td>
<td>0.41</td>
<td>0.24 – 0.71</td>
<td>p = .001</td>
</tr>
</tbody>
</table>

The increased dropout in the landmark categories did not lead to systematic between-group differences in sample composition. Univariate analyses found no significant differences in gender ($\chi^2 = .626; df = 3; p = .890$), level of income ($\chi^2 = 7.961; df = 9; p = .538$), education level ($\chi^2 = 5.364; df = 6; p = .498$) or relationship status ($\chi^2 = 1.492; df = 3; p = .684$), and no significant differences in mean age between the resulting four subsamples ($F(3, 1234) = .564; p = .639$). Since most of the dropout in our experiment occurred shortly after the beginning of the interview, we limited our further analyses to the respondents who completed the whole questionnaire.
5.5.2. Completeness of retrospective reports

Completeness of the retrospective data was analyzed with regard to the number of reported relationships, accommodation episodes, employment episodes (> 6 months), unemployment episodes, and family leaves. We tested the following hypotheses:

Hypothesis 2a: In retrospective web surveys with a calendar recall aid, respondents will report higher numbers of specified events and episodes than in retrospective web surveys without a calendar recall aid.

Hypothesis 2b: The effects of a calendar recall aid on the number of reported events will be especially strong in case of less recent (i.e. more difficult to remember) events and only weak or non-existent in case of recent (i.e. easier to remember) events.

Number of reported episodes

In total, the 1238 respondents who finished the interview reported 1506 relationships, 4259 residences, and 3630 employment episodes of more than 6 months. Hypothesis 2a was not supported for any of those three categories of events, as we found no significant main effects of landmarks and visual feedback and no significant interaction effects between those two factors on the number of reported relationships, accommodations, and (long-term) employment episodes. It is, however, possible that these major episodes were relatively easy to remember and that the accuracy of reports was high in all experimental conditions. We examined effects of the recall aid on the retrieval of unemployment episodes, an event type that is often underreported in retrospective surveys, and on the retrieval of family leave episodes.

Reports of unemployment episodes and family leave were only available for respondents who had held at least one paid job in the past. Out of a total of 1238 respondents, 1226 indicated that they either had a job at the time of the interview or had held one in the past.

In line with hypothesis 2a, we found a significant main effect of visual feedback (F(1, 1222) = 4.517; p = .017; one-sided) on the total number of unemployment episodes respondents had experienced (see table 5.4). Respondents who received visual feedback of their answers reported significantly higher numbers of unemployment episodes. In the visual
feedback conditions, 15.1 % of respondents reported at least one unemployment episode, versus 11.8% in the conditions without visual feedback. The main effect of landmark events was non-significant ($F(1, 1222) = 2.537; p = .056; \text{one-sided}$), as was the interaction effect between visual feedback and landmarks ($F(1, 1222) = 1.305; p = .254$). Educational attainment, age, and gender did not have significant effects on the number of reported unemployment episodes and were therefore excluded from the model.

Table 5.4: Number of unemployment episodes

<table>
<thead>
<tr>
<th>Landmark events</th>
<th>No</th>
<th>Yes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>.127 (n = 331)</td>
<td>.138 (n = 297)</td>
<td>.132 (n = 628)</td>
</tr>
<tr>
<td>Yes</td>
<td>.151 (n = 324)</td>
<td>.219 (n = 274)</td>
<td>.182 (n = 598)</td>
</tr>
<tr>
<td>Total</td>
<td>.139 (n = 655)</td>
<td>.177 (n = 571)</td>
<td>.157 (n = 1226)</td>
</tr>
</tbody>
</table>

We found similar results when analyzing the numbers of family leaves (table 5.5). There was a significant main effect of visual feedback on the numbers of reported family leaves ($F(1, 1215) = 3.141; p = .038, \text{one-sided}$). Respondents in the visual feedback conditions reported significantly higher numbers of family leaves. Again, the effect of landmark events was non-significant ($F(1,1215) = .098; p = .38, \text{one-sided}$), as was the interaction between landmarks and visual feedback ($F(1,1215) = .016; p = .899$).

Table 5.5: Number of family leaves

<table>
<thead>
<tr>
<th>Landmark events</th>
<th>No</th>
<th>Yes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>.099* (n = 330)</td>
<td>.103* (n = 297)</td>
<td>.101* (n = 627)</td>
</tr>
<tr>
<td>Yes</td>
<td>.131* (n = 324)</td>
<td>.140* (n = 272)</td>
<td>.135* (n = 596)</td>
</tr>
<tr>
<td>Total</td>
<td>.115* (n = 654)</td>
<td>.121* (n = 569)</td>
<td>.115 (n = 1223)</td>
</tr>
</tbody>
</table>

* Estimated marginal means, controlling for gender and age
The covariate age had a significant effect on the number of reported family leaves, with older respondents reporting higher numbers of family leave episodes ($r = .08; p < .01$). Unsurprisingly, female respondents reported significantly more family leave episodes than male respondents (.25 versus .01; $t = 11.112; p < .001$).

*Interaction of aided recall with temporal distance*

Based on the results of earlier methodological studies, we expected that the calendar recall aid would especially increase the number of reported temporally remote events whereas the effect on recent events is expected to be much lower (hypothesis 2b). We examined the potentially differential effects of the recall aid on the retrieval of recent and temporally remote unemployment episodes and family leave episodes respectively. Since respondents who reported unemployment episodes or family leaves were not explicitly asked for the end date of those episodes, we had to take the start date of the episode as an indicator of temporal distance. Episodes that started before 1/1/1990 were coded as ‘remote’, all other episodes were coded as ‘recent’.

With regard to unemployment, we found that the positive effect of the visual feedback on the number of reported episodes was only significant for reports of temporally remote ($F(1, 1215) = 3.667, p = .028$, one-sided), but not for reports of recent episodes ($F(1, 1215) = 1.265, p = .131$). Hence, hypothesis 2b was supported for the effect of visual feedback on the number of reported unemployment episodes. Landmark events did not increase the number of reported recent or temporally remote unemployment episodes significantly, therefore hypothesis 2b could not be confirmed for the effect of landmarks.

While the visual feedback also led to small increases in the numbers of reported temporally remote family leave episodes, the effect was not large enough to be significant ($F(1, 1215) = 1.376, p = .12$, one-sided). The same was true for more recent family leave episodes ($F(1, 1215) = 1.749, p = .093$, one-sided). Again, landmark events did not have a significant effect, neither on the numbers of reported temporally remote family leaves, nor on the number of reported recent family leaves.
5.5.3. Data consistency

Chronological consistency

We tested hypothesis 3a: ‘In retrospective web surveys with a calendar recall aid, respondents will make fewer chronological errors (in terms of event order and consistency of start and end dates) than in retrospective web surveys without a calendar recall aid.’

The overall numbers of errors made with regard to chronological consistency in our sample were relatively low. Only 3.6% of the respondents reported at least one relationship, accommodation episode, or employment episode with inconsistent start and end dates (i.e. an end date that preceded the start date). 3.5% of the respondents made at least one mistake with regard to event order. There was a substantial overlap between the groups of respondents who made either mistake, so that the total percentage of respondents who made at least one chronological error was 4.8%. A logistic regression with ‘chronological error’ (1 = one error or more/0 = no errors) as the dichotomous outcome variable found neither landmarks nor visual feedback to decrease the odds of reporting at least one chronologically inconsistent episode significantly. The respondent’s age was the only predictor in our model that significantly affected the odds of making a chronological error (Exp B = 1.02; 95% CI: 1.01 – 1.03). In the youngest age group (25 to 34 years) only 0.6% of respondents made such an error as opposed to 9% of all respondents in the oldest age group (65 years and older).

Consistency of reported and inferred episode durations

Partially supporting hypothesis 3b (“In retrospective web surveys with a calendar recall aid there will be lower numbers of inconsistencies between reported and inferred episode durations than in retrospective web surveys without a recall aid.”), a three-way ANCOVA found a significant main effect of visual feedback on the number of consistency errors with regard to reported and inferred durations (F(1, 1068) = 5.266, p = .010; one-sided). The mean number of consistency errors in the visual feedback conditions was 0.019 versus 0.056 in the condition without visual feedback. There was no significant main effect of
landmark events (F(1, 1068) = .021, p = .443; one-sided) and a non-significant interaction effect of visual feedback and landmarks (F(1, 1068) = .263, p = .608). The control variables educational level, age, and gender had no significant effect on the number of errors made and were left out of the model.

5.5.4. Interview duration

We tested the hypothesis that the use of personal landmark events in retrospective web surveys will increase interview durations.

As the survey was self-completed, respondents could fill in the questionnaire at their own convenience, which included the possibility to complete the questionnaire in multiple sessions. In a substantial number of cases, this led to very long completion times, which inflated the mean and standard deviation of the interview duration. While the median data collection completion time was 2002 seconds, the mean completion time was 18754 seconds, i.e. more than five hours. For 16 cases, no completion time was computed. We filtered out those missing cases, as well as all interviews with completion times of more than two hours. Since the distribution of interview durations was still positively skewed, we carried out a logarithmic (ln) transformation of interview duration for our multivariate analysis (Bland & Altman, 1996). The transformed variable ‘ln(interview duration)’ was normally distributed (Kolmogorov-Smirnov Z = 0.661; p = .775).

Table 5.6: Data collection duration in seconds (back transformed means)

<table>
<thead>
<tr>
<th>Visual feedback of answers</th>
<th>Landmark events</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>1461.6 (n = 283)</td>
<td>1937.0 (n = 233)</td>
</tr>
<tr>
<td>Yes</td>
<td>1651.9 (n = 280)</td>
<td>1889.6 (n = 207)</td>
</tr>
<tr>
<td>Total</td>
<td>1553.4 (n = 563)</td>
<td>1914.5 (n = 440)</td>
</tr>
</tbody>
</table>

Our findings support hypothesis 4. A three-way ANCOVA found a significant main effect of landmark events on the natural logarithm of interview duration (F (1, 992) = 55.644, p <
.001), with interviews in the landmark conditions taking significantly longer than interviews in which no landmarks had to be reported (see table 5.6). The main effect of visual feedback on ‘ln (interview duration)’ was non-significant (F (1, 992) = 2.340, p = .126). The interaction effect between landmarks and visual feedback was found to be significant (F (1, 992) = 5.820, p = .016), although we should be careful in interpreting this effect, since a relatively large number of respondents in the landmark conditions were excluded from the analysis. The covariate age had a strong positive effect on the transformed interview durations (F (1, 992) = 239.868, p < .001). The back transformed mean interview durations were 36.7 minutes for respondents who were 65 years or older at the time of the interview, versus 19.7 minutes for respondents under the age of 35. The effect of gender on ‘ln (interview duration)’ was small but significant, with male respondents taking only slightly longer than female respondents (28.9 versus 27.5 minutes) to complete the interview.

5.5.5. Data revision

Using paradata from the experiment, we tested the hypothesis that respondents who are provided with a calendar recall will edit their responses more often than respondents who are not provided with a calendar recall aid (H5).

We found that only 10.7% of all respondents did not use the ‘back’ button at all, and 34.7% did not make any corrections to their earlier answers. On average, respondents went back 7.90 (median = 4) times during the interview, excluding the landmark section, and made 5.34 (median = 1) corrections. As the distributions of both measures (‘number of times clicked back button’ and ‘number of times edited response’) were severely skewed to the right, we decided to apply a logarithmic transformation to those variables. The measures ‘ln (times clicked back-button)’ and ‘ln (times edited response)’ were strongly correlated (Pearson’s r = .908, p < .001). The back transformed (geometric) means were 4.05 for the number of ‘back’ clicks, and 1.94 for the number of corrections made.

A three-way ANCOVA showed no significant effects of visual feedback (F(1, 1227) = .854, p = .356) and landmark events (F (1, 1227) = .139, p = .709) and no significant interaction
effect of the components (F (1, 1227) = .065, p = .799) on the natural logarithm of the number of times that respondents went back in the questionnaire.

However, there was a significant main effect of visual feedback on the transformed number of corrections made to earlier answers (F (1, 1232) = 2.960, p = .043; one-sided). Respondents in the visual feedback condition made more corrections to their earlier answers than respondents who received no visual feedback. The back transformed mean numbers of corrections were 2.12 in the visual feedback conditions versus 1.79 in the conditions without visual feedback. Again, we found no significant main effect of landmark events (F (1, 1232) = .354, p = .276; one-sided) and no significant interaction effect between visual feedback and landmarks (F (1, 1232) = 2.078, p = .15). The control variables age, gender, and education level had no significant effect on the dependent variable and were therefore excluded from the model. Hypothesis 5 was only supported with regard to the effect of visual feedback, but not to the effect of landmark events.

5.5.6. Respondent evaluations

Of all the respondents who were given a recall aid and finished the interview, 72.2% indicated that they preferred the new format to conventional questionnaires. Approval of the new format was the highest in the condition in which respondents did not have to report landmark events (see table 5.7).

<table>
<thead>
<tr>
<th></th>
<th>Visual feedback plus landmarks</th>
<th>Landmark events only</th>
<th>Visual feedback only</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefers QL with recall aid</td>
<td>70.4% (195)</td>
<td>66.8% (199)</td>
<td>78.5% (260)</td>
<td>72.2% (654)</td>
</tr>
<tr>
<td>Prefers regular question-list</td>
<td>29.6% (82)</td>
<td>33.2% (99)</td>
<td>21.5% (71)</td>
<td>27.8% (252)</td>
</tr>
<tr>
<td>Total</td>
<td>100% (277)</td>
<td>100% (298)</td>
<td>100% (331)</td>
<td>100% (906)</td>
</tr>
</tbody>
</table>
It should be noted that the percentages presented in table 5.7 could be somewhat biased by non-response error. Given the substantial differences in response rates between conditions, it is likely that many of the respondents who broke off the interview would not have indicated that they liked the new format better than regular questionnaires without recall aid (although, of course, we cannot be sure that they would not have enjoyed the rest of the interview had they not broken it off). We conducted a logistic regression (table 5.8) in which we compared the odds of preferring the new format between the three conditions in which the evaluation question was asked.

Table 5.8: Logistic regression model of respondent preference

<table>
<thead>
<tr>
<th>Condition</th>
<th>Exp. B = OR preferring new format (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>QL &amp; VF &amp; LM</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>QL &amp; LM</td>
<td>.82 (.57 – 1.17)</td>
<td>p = .263</td>
</tr>
<tr>
<td>QL &amp; VF</td>
<td>1.55 (1.07 – 2.24)</td>
<td>p = .022</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>0.98 (0.97 – 0.99)</td>
<td>p &lt; .001</td>
</tr>
</tbody>
</table>

Compared to respondents in the ‘question-list plus visual feedback and landmarks’ category, respondents who received only visual feedback were significantly more likely to prefer the questionnaire with the recall aid to regular questionnaires. For respondents who received only landmark events as a recall aid, the odds of a positive evaluation did not differ significantly from the ‘visual feedback plus landmarks’ group. Gender and education were not found to be significant predictors of respondent evaluations. Age was a significant predictor variable, with older respondents having lower odds of preferring the new format than younger respondents.
5.6. Conclusions

In the current study, we found some indications that adding calendar instruments to web-based survey questionnaires influences data quality in online surveys. We also found that the different components of those calendar instruments (landmarks and visual feedback) had different effects on various quality measures.

First of all, compared to the control condition, break-off rates were much higher in the conditions in which landmark events had to be reported. Visual feedback had no effect on break-off rates. All in all, our results indicate that the high break-off rates in our study may have been caused by adding (personal landmark) questions to the survey, which some of the respondents found difficult to answer and which could not be skipped easily if the respondent did not want to answer them.

Secondly, while the effects of landmark events on data quality were negligible, the effects of the visual feedback tool on data completeness were mostly positive. It appeared to enhance the completeness of the retrospectively collected life course data by increasing the numbers of reported episodes for some, but not all types of events. In our study, visual feedback increased the numbers of reported unemployment episodes and family leave episodes. This result is consistent with Belli et al.'s (2007) finding that an Event History Calendar enhanced reports of employment and unemployment episodes in a life course survey. We found some indication that the higher numbers of reported unemployment episodes in our study may have been due to enhanced retrieval of temporally distant unemployment episodes specifically, as the effect of the visual feedback was only significant for older, but not for more recent episodes.

Landmark events, the other major component of calendar instruments, were not found to increase the number of any type of reported events, nor did they increase the amount of time accounted for. Neither landmarks nor visual feedback increased the total numbers of reported major relationship, accommodation episodes, and employment episodes. It seems likely that the completeness of reported job histories could have been very high for major episodes regardless of the data collection instrument. In sum, our results point towards the possibility that calendar instruments are especially helpful for remember-
ing non-standard, shorter, or more ‘easy-to-forget’ episodes within certain thematic domains, while having little to no effect on the retrieval of other events.

Thirdly, the effects of the calendar recall aid on data consistency were modest. Visual feedback enhanced the consistency between reported and inferred durations in the accommodation section of the questionnaire. It did not, however, decrease error rates with regard to consistency of start and end dates and chronological consistency of reports (‘first things first’). Again, landmark events had no significant effect on data quality.

Fourthly, as expected, interview durations were longer for respondents who had to report landmark events at the beginning of the interview. Offering respondents visual feedback did not affect interview durations significantly.

Fifthly, respondents in the visual feedback conditions edited their answers slightly more often than respondents in the conditions without visual feedback. Landmark events had no effect on data editing. It is possible that this finding offers a partial explanation for the (moderate) positive effects of the visual feedback on data quality in our study.

Finally, the majority of subjects who were presented with a recall aid evaluated it positively in comparison to regular online questionnaires. This was true for all three experimental conditions, although it seemed that the visual feedback was evaluated somewhat more favorably than the landmark events.

The wider impact of the results presented in this section will be discussed and put into context in the conclusion chapter.

Limitations of the study

The most serious limitation of the present study into the effects of calendar recall aids on the quality of retrospective life course data was the lack of control information. We were not able to obtain ‘gold standard’ data such as administrative records or complete life course survey data, against which we could judge the accuracy of our respondents’ answers. As a consequence, it was impossible for us to determine the correctness of event dates as well as event frequencies.

Another limitation of our study was that the design of the questionnaire’s landmark section was suboptimal in terms of routing and user-friendliness. Respondents in the landmark conditions, who did not want to report personal landmark events, still had to an-
swer all six landmark questions, six date questions, and click through the associated automatic checks (i.e. “You did not provide a response to the previous question. Is that correct?”) in order to get through to the main questionnaire. It is not unlikely that there was a causal relationship between this design flaw and the high break-off rates in the landmark conditions. However, it has to be noted that this limitation could only affect break off rates if respondents were trying to skip the landmark section. Respondents who entered six landmark events were not confronted with any extra questions.

In future methodological studies into the effects of web-based calendar recall aids these issues should be addressed.
Chapter 6:
Conclusion and Discussion

6.1. Summary of findings

The study set out to answer the questions about the effects of calendar instruments in social surveys that were raised in the introduction:

1. What do we know about the effects of calendar instruments on data quality?
2. Which cognitive processes are involved in answering retrospective questions? Do these processes differ between interviews with calendar instruments and interviews with regular retrospective questionnaires?
3. Which components of calendar instruments are relevant to improving the quality of retrospective data in web-based life course surveys?
4. How can this knowledge be used to further improve the design of calendar techniques for web-based life course surveys?

First of all, we reviewed the literature on the effects of calendar instruments on data quality in social surveys (research question 1). Based on the findings of a limited number of earlier methodological studies, the picture emerged that calendar instruments tended to have moderate positive effects on several measures of data quality. Those measures include completeness and logical consistency of retrospective reports, as well as some aspects of dating accuracy. With regard to data completeness, one of the most important effects of the calendar method was the reduction of gaps in the respondents’ retrospective accounts. Different calendar instruments improved time accounted for with regard to fertility histories (Goldman et al., 1989) and employment histories (Engel, Keifer, & Zahm, 2001). In some – but not all – studies, calendar instruments led to higher numbers of reported events, especially for target events that were relatively difficult to recall, due to low saliency, high frequency, and temporal remoteness of the event(s).
Several studies also found that calendar instruments enhanced the logical consistency of retrospectively reported life history data. Empirical results were mixed with regard to dating accuracy. Often, calendar instruments were found to reduce absolute error, but there was usually little effect on the direction of the error.

Interestingly, the mode in which the interview was conducted did not seem to have great consequences for the sort of effects that calendar instruments had on data quality. Largely similar outcomes were found in a face-to-face survey with a relatively simple paper-and-pencil calendar recall aid (Van der Vaart, 2004), CATI interviews in which only the respondent could see the (again, relatively simple) calendar recall aid (Van der Vaart & Glasner, 2007a), and CATI interviews in which a complex Event History Calendar (paper-and-pencil or electronic) was only visible to the interviewer (Belli et al, 2001 and Belli et al, 2007 respectively). It should be noted that in the two latter studies conducted by Belli and his colleagues, interviewers in the calendar condition but not in the control condition (in which a regular standardized question-list was used) used a ‘conversationally flexible’ interviewing method, which in itself could have had positive effects on data quality (Schober & Conrad, 1997).

The fact that calendar instruments improved data quality in telephone interviews, in which the respondent could not make use of the visual properties of the instrument, suggests that the instrument is not only a recall aid for the respondent, but also a tool that can be used by the interviewer in order to perform consistency checks during data collection (Reimer & Matthes, 2007). This assumption was corroborated by Dijkstra, Smit, and Ongena (2009), who reported that the interviewers in their CAPI survey ‘heavily used cross-checking’ during the EHC interviews.

The literature review revealed that, generally speaking, the results of earlier methodological studies of calendar instruments were quite promising with regard to the method’s potential positive effects on data quality. However, those effects appeared to be dependent on the subject of the survey and the difficulty of the recall task. As Belli (2007; p. 618) noted with respect to interviewer-administered Event History Calendars, the positive effects of the calendar method may be especially pronounced for retrospective questions that “require the interweaving of several elements simultaneously”, such as complex employment histories, whereas regular question-lists can be perfectly adequate for collect-
ing straightforward retrospective information that is well-represented in autobiographical memory (i.e. ‘time-tagged’ events such as marriages).

The results of the literature review suggest that aided recall techniques should be attuned to the subject matter of the specific survey, and their expected net benefits should be judged on a case-to-case basis. Differences in target information (e.g. complex educational histories versus marriage histories) and survey populations (young versus elderly respondents) affect the difficulty of the retrieval task, and the applicability of aided recall techniques.

Also, especially in light of the finding that the use of calendar instruments tended to increase operational costs in comparison to regular standardized questionnaires, the review encouraged us to look for more systematic approaches to the development and optimization of calendar methods, in order to increase the relevance and (cost-) effectiveness of the instrument.

The pilot study reported in the empirical part of the dissertation compared the cognitive processes involved in answering retrospective questions between interviews with regular retrospective questionnaires and interviews in which a calendar recall aid was used (research question 2). From the cognitive interviews we conducted, we found no clear pattern of effects of the calendar recall aid on the types of retrieval processes that survey respondents used when answering retrospective questions. Given the small sample size of that pilot, however, it was difficult to draw any definitive conclusions to that regard.

Nonetheless, the cognitive pilot study provided some useful insights into the way in which respondents apply public and personal landmark events. The numbers of reported landmarks varied considerably between respondents. We found that older respondents reported fewer landmarks, and that respondents who reported more landmarks also reported more transitions in the reference period. The number of reported landmark events was related to the number of reported direct date retrievals, but it was completely unrelated to the number of times that respondents actually used those landmark events as retrieval cues. Moreover, although respondents in the calendar condition tended to report the same type of events as landmarks as they would subsequently use as retrieval cues, the types of retrieval cues used by respondents in the regular question-list condition were very similar. Our conclusion from this cognitive pilot study was that cognitive processes in retrospective life
course interviews might not depend as much on the method of data collection as they depend on characteristics of the respondent and on the retrieval task.

In order to answer the question if and how the main components of calendar instruments, landmarks and visual feedback can improve the quality of retrospective data in web-based life course surveys (research question 3), we then conducted a large-scale methodological evaluation. In a split-ballot experimental design, three different versions of a calendar recall aid were compared to a regular web-based life course questionnaire. The evaluation study found that, for the lifetime reference period, web-based calendar instruments which included visual feedback (i.e. the instruments used in the ‘question-list plus visual feedback’ and in the ‘question-list plus visual feedback and landmarks’ conditions), enhanced the quality of retrospective reports in a way that was largely consistent with earlier findings presented in the methodological literature. Providing respondents with visual feedback led to higher numbers of reported family leave episodes and to higher numbers of reported unemployment episodes. This is in line with findings by Belli and his colleagues (2007) that Event History Calendars improved reports of unemployment episodes in life course surveys. The effect on the number of reported unemployment episodes was larger for temporally remote episodes than for recent episodes, corroborating the results of various earlier evaluation studies (e.g. Van der Vaart, 2004; Yoshihama et al., 2005). In all these cases, the effect of the recall aid was modest, but significant. Providing visual feedback also seemed to reduce the number of errors that respondents made when calculating the length of shorter accommodation episodes. However, it did not reduce error rates for other types of data consistency, such as the consistency of episode start and end dates. In the current study, visual feedback also did not increase the number of reported major employment, accommodation, and relationships episodes. Lack of effects on the reported numbers of this type of (supposedly) highly salient, major episodes could be explained by the high baseline accuracy of this type of reports and the relatively low total numbers of those major episodes that respondents reported (see also Van der Vaart, 1996).
6.2. Effects of landmark events and visual feedback

Our evaluation study was designed to disentangle the effects of landmark events and visual feedback in the context of web-based retrospective life course surveys in order to make recommendations for the optimal design of a web-based calendar recall aid (research questions 3 and 4). Indeed, we found differential effects of the two components (landmarks and visual feedback) on various quality measures. The main finding in this regard was that the positive effects described in the previous section seemed to be due to the inclusion of visual feedback of the respondents’ answers, rather than to asking respondents for a number of landmark events that they could use as retrieval cues during the survey. Respondents in the experimental condition in which the visual feedback calendar without landmark events was used, were also the most likely to evaluate the recall aid in a positive way.

The most surprising finding of this experimental evaluation was the strong effect that the inclusion of personal landmark events had on break-off rates. It is likely that the large increase in break-off rates was partially caused by suboptimal design of the calendar software’s landmark section in terms of routing and instructions. Additional research is warranted to see if the problem persists if those technical issues are resolved. Nonetheless, we would like to stress again that more than 70% of the drop-out in the landmark conditions occurred before respondents even answered the first landmark question. This suggests that the sheer inclusion of landmark questions, whose purpose may not have been clear to all respondents, lead to increased break-off rates. For now, the finding illustrates the importance of user-friendliness of web-based surveys. With no interviewer present to build rapport and to assist the respondent in completing the questionnaire, a large part of the respondent’s motivation to finish the interview depends on the design of the survey. If answering the questions is perceived as an unclear, cumbersome, and time-costly task, respondents’ motivation can falter, causing them to break off the survey. More generally, the high break-off rates in the landmark conditions serve as a reminder that recall cues must not only help
the respondent overcome cognitive difficulties, but at the same time should not be too time-consuming or redundant, as we need to maintain motivational levels.

In the following sections, potential explanations for the unexpected lack of positive effects of landmark events on data quality in our study will be discussed together with some relevant results of the cognitive pilot study described in chapter 4. First of all, the most important effect of landmarks on data quality could be their influence on the accuracy of retrieved event dates. In the studies described in this dissertation, we could not detect that effect, since neither the pilot study nor the experimental evaluation included comparisons of retrieved event dates to (more or less) objective control data.

Secondly, the cognitive pilot study as well as the experimental evaluation had very long reference periods (20 and 30+ years respectively) and the open instructions to report landmarks as well as the broad topic of the life course surveys made it impossible for respondents to anticipate which landmarks would be useful for dating events during the main part of the interview. Asking respondents to report memorable events from such a long period made it unlikely that reported landmark events were temporally close to one of the target events purely by coincidence. It could be argued that improving the landmark instructions and making them more specific would lead to better efficiency of the calendar method. Also, the effect of landmark events on event dating might depend on the length of the reference period. If the reference period is shorter (i.e. one or two years, instead of 30+ years), the ‘density’ of potentially relevant temporal anchors will be higher. This means that even if landmark instructions are kept non-directive, the odds of reporting a landmark event that happened in clear temporal proximity to the target event are much higher in surveys with short reference periods. Furthermore, recurrent, or ‘cyclic’, landmark events such as public holidays, birthdays, or school terms will only be useful as retrieval cues if the reference period does not comprise more than a few cycles/years. All in all, we may have underestimated the effect of landmark events, because of the very long reference period of our survey.

Thirdly, it is possible that the effect of landmark events may be more pronounced in a more interactive survey mode, i.e. in ‘flexible’ calendar interviews, in which the interviewer can ask respondents to reconstruct dates of target events by placing them in relation to other autobiographical events. Since web-based surveys are usually self-administered,
positive effects of calendar instruments that depend on interviewer behavior cannot be estab-
lished.

Given the clear limitations of our research, we cannot conclude from our methodo-
logical experiment (see chapter 5) that the benefits of including landmark events in web-
based life course surveys with long reference periods will offset the cost of doing so in
terms of response rates and data quality. Even though it is still possible that landmark
events improve event dating in those surveys, we saw in the cognitive interviews that re-
spondents in the calendar condition did not use more landmark retrieval cues than respon-
dents who completed a regular question-list without a recall aid. The results of the cogni-
tive pilot study point towards the possibility that important, life time events in autobi-
ographical memory (such as weddings, graduations, or the birth of children) might be con-
stantly accessible and can be used as temporal anchors during retrieval tasks with or with-
out a calendar recall aid.

6.3. Recommendations for the use of visual feedback
and landmark events in web-based surveys

In our experimental evaluation of a web-based calendar instrument, we found mainly posi-
tive - if rather modest - effects of visual answer feedback on several measures of data qual-
ity. While offering respondents visual feedback did not lead to higher reported numbers of
major episodes, it did increase the number of reported employment gaps, specifically those
gaps that were temporally remote. Apart from the relatively high development costs, we
found no indication that the inclusion of visual feedback had negative effects on the survey
process. Dropout rates in the visual feedback condition of the experiment were not signifi-
cantly higher than in the control group, nor did the feedback increase interview durations to
a significant degree. According to the results of the studies described in this dissertation,
visual feedback can be a useful addition to web-based life course surveys.

As was mentioned in previous sections, the usefulness of landmark events in cal-
endar instruments may depend on a number of factors. To begin with, they could be more
useful for studies with relatively short reference periods, possibly with the exemption of
situations in which a researcher knows in advance which types of personal landmarks can be useful as retrieval cues for the target events. For example: in psychological or epidemiological life course studies, the researcher might be aware of lifetime events that can trigger or cause certain illnesses or conditions (e.g. the loss of a spouse as a precursor of depressive episodes). In that case, asking respondents for those personal events could lead to improved accuracy of the target information. Furthermore, landmark events may be more effective in interviewer-administered surveys. With regard to self-completed web-questionnaires however, our preliminary recommendation would be to err on the side of caution. Collecting personal landmark events at the beginning of the survey is time-costly and may give respondents the impression that participation in the survey is going to take a lot of time and effort, leading to higher drop-out rates. Given the present lack of evidence for benefits of landmark events in web-based surveys, it might therefore be a better idea to not include landmark questions until clear positive effects of this design element have been found in other empirical studies.

Both design features of calendar instruments, the inclusion of landmark events and providing respondents with visual feedback, will increase the operational costs of web-based surveys. Calendars that contain visual feedback but no landmark events will mainly lead to increased costs for programming and testing the survey, as long as there is no standard software for designing web-based calendar instruments. If landmarks are used, they might also lead to increased cost in terms of interview duration and possibly break-off rates.

6.4. Calendar instruments and the survey response process

Apart from interpreting the results found in this dissertation in the context of earlier theoretical descriptions of calendar instruments, the pattern of findings allows for an alternative explanation for some of the positive effects of this method on data quality.

In their book The Psychology of Survey Response (2000), Tourangeau, Rips and Rasinski introduced a revised theoretical model of the response process. The authors de-
scribed the cognitive processes involved in answering survey questions in terms of four steps: comprehension, retrieval, judgment, and response. When answering a retrospective question such as ‘When was the last time you visited a doctor?’ the respondent will first have to understand what is being asked of her (comprehension). Secondly, she has to search through her memory for relevant pieces of information such as specific doctor visits and their dates (retrieval). Thirdly, the quality and suitability of the retrieved content has to be determined (judgment). In this stage, the respondent might, for instance, have to decide which of the doctor’s visits she retrieved was the last one, or whether or not a visit to the dentist would also count as a doctor’s visit. Finally, after making this judgment, the respondent will have to map her answer to the appropriate format, i.e. a duration, date, or response category.

The theoretical rationale of calendar instruments has mainly focused on the potential effects of calendar instruments on the retrieval of autobiographical events. However, our findings suggest that calendar instruments could also be particularly useful in the final stages of the response process, in which the respondent actually reports and sometimes edits his answers to retrospective questions. Furthermore, the impulse to edit the response can be given by the interviewer who can use the calendar instrument to detect inconsistent and incomplete event sequences in the data. It is possible that - in addition to the method’s aided recall properties - at least part of the impact of calendar instruments is connected to the opportunities they offer for response checking and editing.

### 6.5. Final remarks and directions for future research

In previous chapters, we already discussed the limitations of our study, and the greatest limitation appeared to be the lack of control data for dates and numbers of events. This means that the study has not led to many new insights concerning the effects of calendar instruments on the accuracy of retrieved event dates or behavioral frequencies. It is therefore considered necessary that future methodological studies include such control data if they address topics such as the optimal length of the reference period, effectiveness of different kinds of landmark events, or the effectiveness of the calendar together with different modes of data collection. The issue of non-response also remains very important. Apart
from that, further research is needed regarding the roles of interviewer and respondent in calendar interviews. How much of the effect of calendar instruments can be ascribed to the active and/or steering role of the interviewer? To what degree should the interviewer use probes and/or help the respondent fill out the calendar? Given the recent popularity of calendar instruments it seems obvious that many scientific fields might benefit from such methodological studies.
References


Appendices
Appendix 4.1: Retrospective questionnaire used in cognitive interviews

Section 1: Accommodation

1.1. Nu willen wij graag weten waar u de afgelopen 20 jaar heeft gewoond.

Waar woonde u op 1 januari 1987?

Straat: ___________________________________________
Plaats: ___________________________________________

1.2. Tot wanneer heeft u op dat adres gewoond?

Jaar: ________________
Maand: ________________

Ik woon er nog steeds -> Verder met blok 2

1.2.1. Hoe zeker bent u van dit antwoord?
(antwoordmogelijkheden voorlezen

Weet jaar en maand absoluut zeker
Zeker (+/- 3 maanden)
Redelijk zeker (+/- 6 maanden)
Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
Weet alleen de maand, maar niet het jaar
Weet alleen het jaar
Moet gokken
1.3. Waar heeft u daarna gewoond?

Straat: ___________________________________________
Plaats: __________________________________________

1.4. Tot wanneer heeft u op dat adres gewoond?

Jaar: ________________
Maand: ________________
Ik woon er nog steeds -> Verder met blok 2

1.4.1. Hoe zeker bent u van dit antwoord?

Weet jaar en maand absoluut zeker
Zeker (+/- 3 maanden)
Redelijk zeker (+/- 6 maanden)
Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
Weet alleen de maand, maar niet het jaar
Weet alleen het jaar
Moet gokken

1._. Waar heeft u daarna gewoond?

Straat: ___________________________________________
Plaats: __________________________________________

1._. Tot wanneer heeft u op dat adres gewoond?

Jaar: ________________
Maand: ________________
Ik woon er nog steeds -> Verder met blok 2
1._1. Hoe zeker bent u van dit antwoord?

Weet jaar en maand absoluut zeker
Zeker (+/- 3 maanden)
Redelijk zeker (+/- 6 maanden)
Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
Weet alleen de maand, maar niet het jaar
Weet alleen het jaar
Moet gokken

Section 2: Employment

Nu volgen enkele vragen over uw werksituatie sinds januari 1987.

Als betaald werk wordt beschouwd:
- werk voor eigen rekening of risico;
- werk als gezinslid in gezins- of familiebedrijf
- werk in loondienst
- werk in sociale werkplaats
- in opleiding bij een bedrijf of instelling, maar met loon of salaris
- stagiair met loon of salaris

2.1. Waartoe rekende u zich op 1 januari 1987 in de eerste plaats?
  o verrichtte betaald werk
  o was werkzoekend (verder met 2.5.)
  o was scholier, student, stagiair met alleen onkostenvergoeding (verder met 2.5.)
  o deed het huishouden (verder met 2.5.)
  o was gepensioneerd of met vervroegd pensioen (verder met 2.5.)
  o was geheel of gedeeltelijk arbeidsongeschikt (verder met 2.5.)
  o deed vrijwilligerswerk of ander onbetaald werk (verder met 2.5.)
  o iets anders, namelijk _____________________ (verder met 2.5.)
2.2. Was u werknemer in vaste of tijdelijke dienst, oproepkracht, uitzendkracht of was u
een zelfstandige/freelancer of vrij beroepsbeoefenaar?
  o werknemer in vaste dienst
  o werknemer in tijdelijke dienst
  o oproepkracht
  o uitzendkracht
  o zelfstandige, meewerkend in gezinsbedrijf, freelancer, vrij beroep

2.3. Welke functie had u in deze baan?
Functieomschrijving: ____________________________

2.4. Hoeveel uren werkte u toen gewoonlijk per week?
Aantal uren: ________

2.5. Is uw werksituatie sindsdien veranderd?
  o Ja
  o Nee -> verder met blok 3

2.6. Wanneer is uw werksituatie na 1 januari 1987 voor het eerst veranderd?
Jaar: ____________
Maand: ____________

2.6.1. Hoe zeker bent u van dit antwoord?
  Weet jaar en maand absoluut zeker
  Zeker (+/- 3 maanden)
  Redelijk zeker (+/- 6 maanden)
  Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
  Weet alleen de maand, maar niet het jaar
  Weet alleen het jaar
  Moet gokken
2.7. Waartoe rekende u zich na deze verandering in de eerste plaats?
  o verrichtte betaald werk
  o was werkzoekend (verder met 2.11.)
  o was scholier, student, stagiair met alleen onkostenvergoeding (verder met 2.11.)
  o deed het huishouden (verder met 2.11.)
  o was gepensioneerd of met vervroegd pensioen (verder met 2.11.)
  o was geheel of gedeeltelijk arbeidsongeschikt (verder met 2.11.)
  o deed vrijwilligerswerk of ander onbetaald werk (verder met 2.11.)
  o iets anders, namelijk _____________________ (verder met 2.11.)

2.7.1. MITS MEN ZOWEL VOOR ALS NA DE VERANDERING BETAALD WERK VERRICHTTE

Welk(e) aspect(en) van uw werksituatie veranderde(n) er toen?
  o Soort dienstverband
  o Functie
  o Gewerkte uren
  o Anders, namelijk _____________________

U kunt meerdere mogelijkheden aanvinken.

ALLEN, MITS ASPECT IS VERANDERD:

2.8. Was u na deze verandering werknemer in vaste of tijdelijke dienst, oproepkracht, uitzendkracht of was u een zelfstandige/freelancer of vrij beroepsbeoefenaar?
  o werknemer in vaste dienst
  o werknemer in tijdelijke dienst
  o oproepkracht
  o uitzendkracht
  o zelfstandige, meewerkend in gezinsbedrijf, freelancer, vrij beroep

2.9. Wat was uw nieuwe functie?
Functieomschrijving: _____________________
2.10. Hoeveel uren werkte u toen gewoonlijk per week?
Aantal uren: _________

2.11. Is uw werksituatie sindsdien veranderd?
  o Ja
  o Nee -> verder met blok 3

2._. Wanneer is uw werksituatie daarna voor het eerst veranderd?
Jaar:_______________
Maand:______________

2._.1. Hoe zeker bent u van dit antwoord?
  Weet jaar en maand absoluut zeker
  Zeker (+/- 3 maanden)
  Redelijk zeker (+/- 6 maanden)
  Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
  Weet alleen de maand, maar niet het jaar
  Weet alleen het jaar
  Moet gokken

2._. Waartoe rekende u zich na deze verandering in de eerste plaats?
  o verrichte betaald werk
  o was werkzoekend
  o was scholier, student, stagiair met alleen onkostenvergoeding
  o deed het huishouden
  o was gepensioneerd of met vervroegd pensioen
  o was geheel of gedeeltelijk arbeidsongeschikt
  o deed vrijwilligerswerk of ander onbetaald werk
  o iets anders, namelijk ________________________
2._.1. Mits men zowel voor als na de verandering betaald werk verrichtte

Welk(e) aspect(en) van uw werksituatie veranderde(n) er toen?

- Soort dienstverband
- Functie
- Gewerkte uren
- Anders, namelijk ____________________

U kunt meerdere mogelijkheden aanvinken.

2._. Was u na deze verandering werknemer in vaste of tijdelijke dienst, oproepkracht, uitzendkracht of was u een zelfstandige/freelancer of vrij beroepsbeoefenaar?

- werknemer in vaste dienst
- werknemer in tijdelijke dienst
- oproepkracht
- uitzendkracht
- zelfstandige, meewerkend in gezinsbedrijf, freelancer, vrij beroep

2._. Wat was uw nieuwe functie?
Functieomschrijving: ______________________

2._. Hoeveel uren werkte u toen gewoonlijk per week?
Aantal uren: _________

et cetera
Section 3: Education

De volgende vragen gaan over opleidingen. Het gaat om opleidingen die niet uitsluitend voor een hobby bedoeld zijn en die minstens 3 maanden hebben geduurd.

3.1. Heeft u vanaf 1 januari 1987 nog een opleiding gevolgd?
   o Ja
   o Nee -> verder met blok 4

3.2. Volgde u op 1 januari 1987 een opleiding?
   o Ja
   o Nee -> verder met 3.5.

3.3. Wat voor type opleiding was dat?
   (bijv. MTS, secretaresseopleiding, HEAO et cetera)
   Type opleiding: ____________________

3.4. Wanneer heeft u deze opleiding beëindigd?
   Jaar: _________
   Maand:_______
   o Nog niet beëindigd

3.4.1. Hoe zeker bent u van dit antwoord?
   Weet jaar en maand absoluut zeker
   Zeker (+/- 3 maanden)
   Redelijk zeker (+/- 6 maanden)
   Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
   Weet alleen de maand, maar niet het jaar
   Weet alleen het jaar
   Moet gokken

3.5. Wanneer bent u na 1 januari 1987 voor het eerst weer een opleiding gaan volgen?
   Jaar: _________
   Maand:_______
3.5.1. Hoe zeker bent u van dit antwoord?
   - Weet jaar en maand absoluut zeker
   - Zeker (+/- 3 maanden)
   - Redelijk zeker (+/- 6 maanden)
   - Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
   - Weet alleen de maand, maar niet het jaar
   - Weet alleen het jaar
   - Moet gokken

3.6. Wat voor type opleiding was dat?(bijv. MTS, secretaresseopleiding, HEAO et cetera)
Type opleiding: ____________________

3.7. Wanneer heeft u deze opleiding beëindigd?
   Jaar: _________
   Maand:_______
   - Nog niet beëindigd

3.7.1. Hoe zeker bent u van dit antwoord?
   - Weet jaar en maand absoluut zeker
   - Zeker (+/- 3 maanden)
   - Redelijk zeker (+/- 6 maanden)
   - Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
   - Weet alleen de maand, maar niet het jaar
   - Weet alleen het jaar
   - Moet gokken

3.8. Heeft u sindsdien nog meer opleidingen gevolgd?
   - Ja
   - Nee -> verder met blok 4

3.9. Wanneer bent u aan de eerstvolgende opleiding begonnen?
   Jaar: _________
   Maand:_______
3._1. Hoe zeker bent u van dit antwoord?
   - Weet jaar en maand absoluut zeker
   - Zeker (+/- 3 maanden)
   - Redelijk zeker (+/- 6 maanden)
   - Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
   - Weet alleen de maand, maar niet het jaar
   - Weet alleen het jaar
   - Moet gokken

3._. Wat voor type opleiding was dat?
Type opleiding: ____________________

3._. Wanneer heeft u die opleiding beëindigd?
   Jaar: _________
   Maand:_______
   - Nog niet beëindigd

3._.1. Hoe zeker bent u van dit antwoord?
   - Weet jaar en maand absoluut zeker
   - Zeker (+/- 3 maanden)
   - Redelijk zeker (+/- 6 maanden)
   - Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
   - Weet alleen de maand, maar niet het jaar
   - Weet alleen het jaar
   - Moet gokken

et cetera
Section 4: General Health

4.1. Als u terugdenkt aan januari 1987, hoe zou u uw gezondheid in die periode over het algemeen noemen?
   - slecht
   - matig
   - goed
   - zeer goed
   - uitstekend

4.2. Is uw gezondheidstoestand sindsdien voor een periode van drie maanden of langer veranderd?
   - Ja
   - Nee -> verder met blok 5

4.3. Wanneer is uw gezondheidstoestand veranderd?
   Jaar: _________
   Maand:_______

4.3.1. Hoe zeker bent u van dit antwoord?
   - Weet jaar en maand absoluut zeker
   - Zeker (+/- 3 maanden)
   - Redelijk zeker (+/- 6 maanden)
   - Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
   - Weet alleen de maand, maar niet het jaar
   - Weet alleen het jaar
   - Moet gokken

4.4. Hoe zou u uw gezondheid vanaf deze verandering over het algemeen noemen?
   - slecht
   - matig
   - goed
   - zeer goed
   - uitstekend
4.5. Is uw gezondheidstoestand sindsdien nog eens voor een periode van drie maanden of langer veranderd?
   o Ja
   o Nee -> verder met blok 5

4._. Wanneer is uw gezondheidstoestand veranderd?
Jaar: _________
Maand:_______

4._.1. Hoe zeker bent u van dit antwoord?
   Weet jaar en maand absoluut zeker
   Zeker (+/- 3 maanden)
   Redelijk zeker (+/- 6 maanden)
   Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
   Weet alleen de maand, maar niet het jaar
   Weet alleen het jaar
   Moet gokken

4._. Hoe zou u uw gezondheid vanaf die verandering over het algemeen noemen?
   o slecht
   o matig
   o goed
   o zeer goed
   o uitstekend
Section 5: Hospitalizations

5.1. Bent u sinds 1 januari 1987 wel eens voor 24 uur of langer opgenomen geweest in een ziekenhuis?
   ○ Ja
   ○ Nee -> einde vragenlijst

5.2. Wanneer was dat?
   Jaar: _________
   Maand: _______

5.2.1. Hoe zeker bent u van dit antwoord?
   - Weet jaar en maand absoluut zeker
   - Zeker (+/- 3 maanden)
   - Redelijk zeker (+/- 6 maanden)
   - Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
   - Weet alleen de maand, maar niet het jaar
   - Weet alleen het jaar
   - Moet gokken

5._. Bent u sindsdien nog een keer voor 24 uur of langer opgenomen geweest in een ziekenhuis?
   ○ Ja
   ○ Nee -> einde vragenlijst

5._. Wanneer was dat?
   Jaar: _________
   Maand: _______
5._1. Hoe zeker bent u van dit antwoord?

   Weet jaar en maand absoluut zeker
   Zeker (+/- 3 maanden)
   Redelijk zeker (+/- 6 maanden)
   Niet erg zeker (zit er mogelijk meer dan 6 maanden naast)
   Weet alleen de maand, maar niet het jaar
   Weet alleen het jaar
   Moet gokken

et cetera
### Appendix 4.2:

**Calendar used in ‘question-list plus calendar’ condition (truncated)**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PERSONAL EVENTS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOUSING</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of contract</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hours/week</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job description</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDUCATION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEALTH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOSPITAL STAYS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 5.1:

Life course questionnaire used in web interviews

**RAge** age
Hoe oud bent u?
Range: 0..120

**RGender** Gender
Wat is uw geslacht?
1 Man
2 Vrouw

**calenderType**
1 Gewone vragenlijst, zonder kalender
2 Vragenlijst met kalender en met landmarks
3 Vragenlijst met kalender maar zonder landmarks
4 Vragenlijst zonder kalender maar met landmarks

**RC001_start** START OF RETROSPECTIVE CHILDREN SECTION
Deze vragenlijst gaat over uw levensloop. We stellen vragen over de kinderen die u hebt (gehad), uw relaties, uw banen, verhuizingen, uw gezondheid en een paar extra vragen over periodes in uw leven. We beginnen met de kinderen die u mogelijk hebt of hebt gehad.

**RC022** EVER HAD CHILDREN
Hebt u ooit een kind van uzelf gehad (waarvan u de biologische ouder bent)? Telt u alstublieft ook kinderen mee die slechts korte tijd geleefd hebben.
1 ja
5 nee

IF EVER HAD CHILDREN = ja THEN

| **RC023** NUMBER OF CHILDREN
| Hoeveel biologische kinderen hebt u (gekregen)? Telt u alstublieft ook kinderen mee die elders leven of die overleden zijn.
| Range: 0..20

| LOOP FROM 1 TO [NUMBER OF CHILDREN] DO

| \ **RC024** YEAR OF BIRTH CHILD
| [We beginnen met het oudste kind.] In welk jaar is uw [eerste/tweede/derde vierde/5e/6e/7e/8e/9e/10e/11e/12e/13e/14e/15e/16e/17e/18e/19e/20e] kind
geboren?
Range: TYear

RC025 NAME CHILD
Hoe heet uw kind?
String

RC026 GENDER CHILD
Is [NAME CHILD] een jongen (man) of een meisje (vrouw)?
1 man
2 vrouw

RC027 CHILD STILL ALIVE?
Is [NAME CHILD] nog in leven?
1 ja
5 nee

IF RC027[biochild] = nee THEN

RC028 YEAR OF DEATH CHILD
In welk jaar is [NAME CHILD] overleden?
Range: TYear

ENDIF

RC029 LEFT JOB BECAUSE OF CHILD
Bent u tijdelijk of voorgoed met werken gestopt toen [NAME CHILD] geboren was?
1 ja, tijdelijk gestopt met werken
2 ja, gestopt en nooit meer gewerkt
5 nee, geen onderbreking
9 niet van toepassing: ik werkte toen niet

IF RC029[biochild] = ja, tijdelijk gestopt met werken THEN

RC030 HOW LONG WAS MATERNITY INTERRUPTION
Hoe lang bent u toen gestopt met werken?
1 minder dan een maand
2 tussen 1 en 3 maanden
3 tussen 3 en 6 maanden
4 tussen 6 maanden en 1 jaar
5 tussen 1 jaar en 3 jaar
6 meer dan 3 jaar, maar ik ben later weer gaan werken

RC030a WHEN STARTED WORKING AGAIN
In welk jaar bent u weer gaan werken?
Range: TYear
RC031 SOURCES OF INCOME MATERNITY LEAVE
Welke bronnen van inkomsten had u toen [NAME CHILD] geboren werd? Meerdere antwoorden mogelijk.
1 inkomen uit werk (ook eigen zaak)
2 inkomen van echtgeno(o)t(e) of partner
3 zwangerschapsuitkering
4 kinderbijslag
5 steun van familie (niet uw echtgeno(o)t(e) of partner) en vrienden
6 interen op eigen vermogen of spaargeld
97 anders

IF RC031[biochild] = anders THEN
RC031a OTHER SOURCES OF INCOME MATERNITY LEAVE
Welke andere bronnen van inkomsten bedoelt u?
Memo
ENDIF
ENDDO
ENDIF

RC038 ADOPTED CHILDREN
Hebt u ooit een kind geadopteerd?
1 ja
5 nee

IF ADOPTED CHILDREN=ja THEN
RC039 NUMBER OF ADOPTED
Hoeveel kinderen hebt u geadopteerd?
Range: 0..20
LOOP FROM 1 TO [NUMBER OF ADOPTED] DO
RC040 ADOPTED CHILD NAME
[We beginnen met het eerste kind dat u hebt geadopteerd.//] Hoe heet uw [eerste/tweede/derde/vierde/5e/6e/7e/8e/9e/10e/11e/12e/13e/14e/15e/16e/17e/18e/19e/20e] geadopteerde kind?
String
RC041  CHILD YEAR OF ADOPTION
In welk jaar hebt u [[We beginnen met het eerste kind dat u hebt geadopteerd.///] Hoe heet uw [eerste/tweede/derde/vierde/5e/6e/7e/8e/9e/10e 11e/12e/13e/14e/15e/16e/17e/18e/19e/20e] geadopteerde kind??///] geadopteerd?
Range: TYear

RC042  ADOPTED CHILD GENDER
Is [[We beginnen met het eerste kind dat u hebt geadopteerd.///] Hoe heet uw [eerste/tweede/derde/vierde/5e/6e/7e/8e/9e/10e/11e/12e/13e/14e/15e/16e/17e/18e/19e/20e] geadopteerde kind???/\] een jongen (man) of een meisje (vrouw)?
1 man
2 vrouw

RC043  ADOPTED CHILD YEAR OF BIRTH
In welk jaar is [[We beginnen met het eerste kind dat u hebt geadopteerd.///] Hoe heet uw [eerste/tweede/derde/vierde/5e/6e/7e/8e/9e/10e/11e/12e/13e/14e/15e/16e/17e/18e/19e/20e] geadopteerde kind???/\] geboren?
Range: TYear

RC044  ADOPTED CHILD STILL ALIVE
Is [[We beginnen met het eerste kind dat u hebt geadopteerd.///] Hoe heet uw [eerste/tweede/derde/vierde/5e/6e/7e/8e/9e/10e/11e/12e/13e/14e/15e/16e/17e/18e/19e/20e] geadopteerde kind???/\] nog in leven?
1 ja
5 nee

IF RC044[adoptchild] = nee THEN

RC045  ADOPTED CHILD YEAR OF DEATH
In welk jaar is [[We beginnen met het eerste kind dat u hebt geadopteerd.///] Hoe heet uw [eerste/tweede/derde/vierde/5e/6e/7e/8e/9e/10e/11e/12e/13e/14e/15e/16e/17e/18e/19e/20e] geadopteerde kind???/\] overleden?
Range: TYear

ENDIF

RC046  LEFT JOB BECAUSE OF CHILD
Bent u tijdelijk of voorgoed met werken gestopt toen u [[We beginnen met het eerste kind dat u hebt geadopteerd.///] Hoe heet uw [eerste/tweede/derde/vierde/5e/6e/7e/8e/9e/10e/11e/12e/13e/14e/15e/16e/17e/18e/19e/20e] geadopteerde kind???/\] adopteerde?
1 ja, tijdelijk gestopt met werken
2 ja, gestopt en nooit meer gewerkt
5 nee, geen onderbreking
9 niet van toepassing: ik werkte toen niet
IF RC046[adoptchild] = ja, tijdelijk gestopt met werken THEN

RC047 HOW LONG WAS MATERNITY INTERRUPTION
Hoe lang bent u toen gestopt met werken?
1 minder dan een maand
2 tussen 1 en 3 maanden
3 tussen 3 en 6 maanden
4 tussen 6 maanden en 1 jaar
5 tussen 1 jaar en 3 jaar
6 meer dan 3 jaar, maar ik ben later weer gaan werken

ENDIF

RC048 SOURCES OF INCOME MATERNITY LEAVE
Welke bronnen van inkomsten had u toen [[We beginnen met het eerste kind dat u hebt geadopteerd.//] Hoe heet uw [eerste/tweede/derde/vierde/5e/6e/7e/8e 9e/10e/11e/12e/13e/14e/15e/16e/17e/18e/19e/20e] geadopteerde kind?///] werd geadopteerd?
1 inkomen uit werk (ook eigen zaak)
2 inkomen van echtgeno(o)t(e) of partner
3 zwangerschapsuitkering
4 kinderbijslag
5 steun van familie (niet uw echtgeno(o)t(e) of partner) en vrienden
6 interen op eigen vermogen of spaargeld
97 anders

IF RC048[adoptchild] = anders THEN

RC048a OTHER SOURCES OF INCOME MATERNITY LEAVE
Welke andere bronnen van inkomsten bedoelt u?
Memo

ENDIF

ENDDO

ENDIF

RC060_end Dit is het einde van het gedeelte 'kinderen'.

RP001_start START OF THE PARTNER SECTION
De volgende vragen gaan over relaties.
RP002  EVER HAD PARTNER
Hebt u een langdurige relatie (gehad) die belangrijk voor u is of was?
Denkt u hierbij aan relaties waarbij u bent getrouwd, waarbij u ongehuwd hebt samenge-
woond of waarbij uw partner op een ander adres woont (woonde) dan u.
1 ja
5 nee

IF EVER HAD PARTNER = ja THEN

| RP018  NAME OF PARTNER
| Wat is of was de voornaam van uw partner?
| Vult u hier alstublieft de naam van uw partner in.
| String

| RP018a  GENDER PARTNER
| Is [NAME OF PARTNER] een man of een vrouw?
| 1 Man
| 2 Vrouw

| RP016  MARRIED TO PARTNER
| Bent u met [NAME OF PARTNER] getrouwd of getrouwd geweest?
| 1 ja
| 5 nee

| RP016a  LIVED WITH PARTNER
| Woonde u voor ten minste een gedeelte van uw relatie in hetzelfde huis met [NAME OF
| PARTNER]?
| 1 ja
| 5 nee

| RP017  START PARTNERSHIP
| In welk jaar begon deze relatie?
| Range: 1900..2009

| RP019  STILL IN A RELATIONSHIP WITH PARTNER
| Zijn [NAME OF PARTNER] en u nog steeds samen?
| 1 ja
| 5 nee

| IF RP019[partnerCnt] = nee THEN

| | RP010  REASONS FOR NOT LIVING WITH PARTNER
| | Waarom wonen u en [NAME OF PARTNER] niet meer samen?
| | 1 De relatie is verbroken (ook echtscheiding)
| | 2 Mijn partner is overleden
| | 3 Mijn partner is naar een verpleeg- of verzorgingshuis verhuisd
| | 97 Andere reden
174
RP020  END PARTNERSHIP
| In welk jaar [zijn NAAM PARTNER en u uit elkaar gegaan/is NAAM PARTNER
| overleden?
| Range: 1900..2009
|
| ENDIF
|
RP015  ANY OTHER PARTNERS
| Hebt u (afgezien van de hiervoor genoemde relatie(s) ^FLAllPartners) nog een andere
| belangrijke langdurige relatie gehad?
| Denkt u hierbij aan relaties waarbij u bent getrouwd, waarbij u ongehuwd hebt samenge-
| woond of waarbij uw partner op een ander adres woont (woonde)
| dan u.
| 1 ja
| 5 nee
| ENDIF

RP022_end  END OF THE PARTNER SECTION
Dit is het einde van het onderdeel 'relaties'.

AC001_start  START OF THE ACCOMODATION SECTION
Het volgende onderdeel gaat over de verschillende plekken waar u tijdens uw
leven gewoond hebt.

AC003  WHEN ESTABLISHED HOME
In welk jaar bent u voor het eerst op uzelf gaan wonen of hebt u uw eigen
huishouden gevormd? Als u nooit op uzelf bent gaan wonen kunt u 9997
invullen.
Range: TYear9997

IF WHEN ESTABLISHED HOME != 9997 THEN
| LOOP FROM 1 TO 15 DO
| |
| IF accomodation counter = ja OR AC023[accomodation counter-ja] = ja THEN
| |
| IF (accomodation counter = 2 OR AC023[accomodation counter-ja] = ja) AND
| AC023[accomodation counter-ja] != ja THEN
| |
| AC007  START LIVING AT RESIDENCE
| In welk jaar bent u naar uw volgende woning verhuisd? Als u het niet
| weet, probeert u dan een zo goed mogelijke schatting te geven.
| Range: 1900..2008
| |
| ENDIF
AC010 CITY RESIDENCE
In welke stad of welk dorp woonde u toen?
String

AC011 WHERE LIVED
Hoe wilt u deze woning noemen? Deze naam is bedoeld om bij de volgende
vragen aan te geven om welke woning het gaat. Bijvoorbeeld de Molenstraat,
de flat of de Keereweer.
String

AC008 TYPE OF RESIDENCE
[Is/Was] [WHERE LIVED] een privé woning (Een privé woning is een
woning die u bezit of huurt)?
1 ja
5 nee

AC013 WAS RESIDENCE IN CURRENT COUNTRY
[Is/Was] [WHERE LIVED] gelegen in Nederland?
1 ja
5 nee

IF AC013[accomodationCnt] = nee THEN

AC014 COUNTRY OF RESIDENCE (NOT CURRENT)
In welk land [is/was] [WHERE LIVED] gelegen?
1 Oostenrijk
2 België
3 Tsjecho-Slowakije
4 Denemarken
5 Finland
6 Frankrijk
7 Duitsland
8 Griekenland
9 Hongarije
10 Ierland
11 Italië
12 Engeland
13 Noorwegen
14 Polen
15 Portugal
16 Slowakije
17 Spanje
18 Zweden
19 Zwitserland
21 Rusland
22 Amerika
23 Ander Europees land
AC015 REGION OF RESIDENCE (NOT CURRENT)
In welke regio [is/was] [WHERE LIVED] gelegen?
1 Noord
2 Oost
3 Zuid
4 West
5 Midden
ENDIF

AC017 AREA OF RESIDENCE
Hoe zou u de omgeving van [WHERE LIVED] omschrijven?
1 Een grote stad
2 Een kleine stad
3 Een groot dorp
4 Een klein dorp
5 Buitenaf op het platteland

AC009 TYPE OF PRIVATE RESIDENCE
[Is/Was] [WHERE LIVED] uw bezit?
1 ja
5 nee
IF AC009[acomodationCnt] = ja THEN

AC018 HOW ACQUIRED PROPERTY
Hoe hebt u deze woning verkregen?
1 met eigen middelen
2 met behulp van een hypotheek
3 met geld van familie
4 gekregen uit een erfenis
5 gekregen als cadeau
6 op een andere manier

AC019 PRICE OF OWNED PROPERTY
Voor welke prijs hebt u [WHERE LIVED] gekocht? We willen graag de marktprijs van de woning weten ten tijde van de aankoop. In de volgende vraag kunt u aangeven of het bedrag in euro of gulden is gegeven.
String
AC020 CURRENCY OF OWNED PROPERTY
Is het bedrag dat u hebt genoemd in gulden, euro of in een andere valuta?
1 gulden
2 euro
3 andere valuta

IF AC020[accomodationCnt] = andere valuta THEN

AC020a OTHER CURRENCY OF OWNED PROPERTY
Welke valuta bedoelt u?
String
ENDIF

ENDIF

AC023 LIVED ANYWHERE ELSE
Hebt u nog ergens anders gewoond?
1 ja
5 nee

IF AC023[accomodationCnt] = ja THEN

AC021 STOPPED LIVING AT RESIDENCE
In welk jaar bent u vertrokken uit [WHERE LIVED]?
Range: TYear

AC021a SHORT TERM LIVING
Bent u hetzelfde jaar naar uw volgende woning verhuisd?
1 ja
2 nee

IF AC0nee1a[accomodationCnt] = nee THEN

AC021b LIVING AT RESIDENCE
In welk jaar bent u naar uw volgende woning verhuisd?
Range: TYear

ENDIF
| | | IF AC009[accomodationCnt] = ja THEN |
| | | AC022 WHAT DONE WITH PROPERTY |
| | | Wat hebt u met de oude woning gedaan toen u verhuisd bent? |
| | | 1 verkocht |
| | | 2 gehouden |
| | | 3 aan iemand gegeven |
| | | 96 anders |
| | | ENDIF |
| | | ENDIF |
| | | ENDIF |
| | | ENDIF |
| | | ENDDO |
| | | ENDF |

AC025_end END OF THE ACCOMMODATION SECTION
Dit is het einde van de vragen over uw woonsituaties.

RE001_start START OF THE WORK HISTORY SECTION
Nu volgen enkele vragen over uw arbeidsverleden.

RE002 AGE FINISHED FULLTIME EDUCATION
In welk jaar bent u gestopt met voltijds onderwijs? Voer alstublieft 9997 in als u nooit naar school bent geweest of nog steeds voltijds onderwijs volgt.
Range: TYear9997

RE004 INTRODUCTION TO WORK HISTORY
Nu volgen wat vragen over elke baan die u hebt gehad voor meer dan 6 maanden. Korte baantjes voor verschillende werkgevers tellen mee als één baan als u in principe hetzelfde deed voor deze werkgevers.

RE005 EVER DONE PAID WORK
Hebt u ooit betaald werk verricht dat langer dan een periode van 6 maanden heeft geduurd?
1 ja
5 nee
IF EVER DONE PAID WORK = ja THEN

RE006  START FIRST PAID JOB
Bent u direct gaan werken na uw laatste opleiding?
1 Met eerste baan begonnen direct na het verlaten van opleiding
2 Gat van 6 maanden of langer tussen mijn eerste baan en het verlaten van opleiding
3 Met eerste baan begonnen voordat opleiding werd verlaten

IF START FIRST PAID JOB = Gat van 6 maanden of langer tussen mijn eerste baan en het verlaten van opleiding THEN

RE007  SITUATION IN GAP AFTER EDUCATION
Wat beschrijft het beste de situatie waarin u zich bevond meteen nadat u uw opleiding had verlaten?
1 Werknemer of zelfstandige
2 Werkloos en op zoek naar werk
3 Werkloos en niet op zoek naar werk
4 Kortdurende baan (voor minder dan 6 maanden)
5 Ziek of arbeidsongeschikt
6 Zorg voor thuis of familie
7 Vrije tijd, reizen, niets doen
8 Gepensioneerd
9 Training/cursus
10 Voltijd vervolgonderwijs
11 Militaire dienst, krijgsgevangene of gelijksoortig
12 Uw vermogen beheren
13 Vrijwilligers- of gemeenschapswerk
14 Dwangarbeid of in gevangenis
15 In ballingschap of verbannen
16 Werkkamp
17 Concentratiekamp
97 Anders

ENDIF

LOOP FROM 1 TO 15 DO

IF job counter = ja OR RE038[job counter] = ja THEN

IF job counter = 1 OR GAP AFTER LEAVING THIS JOB=Er was een periode van meer dan 6 maanden voor de nieuwe baan begon OR GAP AFTER LEAVING THIS JOB=Begonnen met nieuwe baan voor de oude afgelopen was THEN

RE011  YEAR STARTED JOB
In welk jaar bent u met uw [eerste/volgende] betaalde baan (als werknemer of zelfstandige) begonnen, die langer dan 6 maanden duurde?

Range: TYear
Wat was dat voor baan? Geeft u alstublieft de exacte benaming of titel.

In wat voor een soort bedrijfstak werkte u als [TITLE OF JOB ]?

1 Landbouw of delfstofwinning
2 Industrie
3 Energie- of waterwinning
4 Bouwnijverheid
5 Handel
6 Horeca
7 Vervoer
8 Financiële instellingen
9 Zakelijke dienstverlening
10 Openbaar bestuur
11 Onderwijs
12 Gezondheid en welzijnszorg
13 Cultuur, sport en recreatie
14 Ideële en belangenorganisaties
15 Iets anders

In deze baan als [TITLE OF JOB ], was u in loondienst, ambtenaar of zelfstandige?

1 Loondienst
2 Ambtenaar
3 Zelfstandige

Werkte u in deze baan als [TITLE OF JOB ], full-time, part-time of een combinatie van beide?

1 Altijd full-time
2 Altijd part-time
3 1 keer veranderd van full-time naar part-time
4 1 keer veranderd van part-time naar full-time
5 Meerdere keren veranderd
IF RE016[jobCnt] = 1 keer veranderd van part-time naar full-time THEN

**RE017 WHY WORKED PART-TIME**
Wat was de hoofdreden dat u in het begin part-time werkte?
1 Verzorging van kinderen of kleinkinderen
2 Verzorging van ouders
3 Verzorging van andere familieleden
4 Vanwege eigen slechte gezondheid
5 Onderwijs/training/cursus
6 Een voltijds baan was niet beschikbaar
97 Anders

IF RE016[jobCnt] = 1 keer veranderd van full-time naar part-time THEN

**RE018 WHEN CHANGED TO FULL-TIME**
In welk jaar bent u veranderd van part-time naar full-time in uw baan als [TITLE OF JOB ]?
Range: TYear

ENDIF

IF RE016[jobCnt] = 1 keer veranderd van full-time naar part-time THEN

**RE019 REASONS CHANGING TO PART-TIME**
Wat was de hoofdreden dat u veranderd bent van full-time naar part-time?
1 Verzorging van kinderen of kleinkinderen
2 Verzorging van ouders
3 Verzorging van andere familieleden
4 Vanwege eigen slechte gezondheid
5 Onderwijs/training/cursus
6 Een voltijds baan was niet beschikbaar
97 Anders

**RE020 WHEN CHANGED TO PART-TIME**
In welk jaar bent u veranderd van full-time naar part-time in uw baan als [TITLE OF JOB ]?
Range: TYear

ENDIF

**RE038 PAID JOB AFTER RETIREMENT**
Na deze baan als [TITLE OF JOB ], hebt u nog ander werk gedaan voor langer dan 6 maanden (zelfs na uw pensioen)?
1 ja
5 nee

IF RE038[jobCnt] = ja THEN
**RE026** YEAR STOPPED IN THIS JOB
In welk jaar bent u gestopt met werken als [TITLE OF JOB]?  
Range: TYear

**RE031** REASONS LEFT JOB
Waarom bent u gestopt met dit werk?  
1 Ontslag genomen  
2 Ontslagen  
3 In overeenkomst (tussen mij en werkgever)  
4 Fabriek, kantoor of bedrijf werd gesloten  
5 Tijdelijke baan was afgelopen  
6 Ging met pensioen  
97 Andere reden

**RE032** GAP AFTER LEAVING THIS JOB
Bent u met uw volgende baan begonnen direct nadat uw oude baan als [TITLE OF JOB] was afgelopen of was er een periode van meer dan 6 maanden voor uw met uw volgende baan begon?  
1 Direct begonnen met nieuwe baan  
2 Er was een periode van meer dan 6 maanden voor de nieuwe baan begon  
3 Begonnen met nieuwe baan voor de oude afgelopen was  
4 Dit was mijn laatste betaalde baan (ook eigen baas)

**RE033** DONE IN GAP AFTER LEAVING THIS JOB
Wat is de beste omschrijving van de situatie waarin u zich bevond in de tijd vóórdat u met uw volgende baan begon?  
1 Werknemer of zelfstandige  
2 Werkloos en op zoek naar werk  
3 Werkloos en niet op zoek naar werk  
4 Kortdurende baan (voor minder dan 6 maanden)  
5 Ziek of arbeidsongeschikt  
6 Zorg voor thuis of familie  
7 Vrije tijd, reizen, niets doen  
8 Gepensioneerd  
9 Training/cursus  
10 Voltijd vervolgonderwijs  
11 Militaire dienst, krijgsgevangene of gelijksoortig  
12 Uw vermogen beheren  
13 Vrijwilligers- of gemeenschapswerk  
14 Dwangarbeid of in gevangenis  
15 In ballingschap of verbannen  
16 Werkkamp  
17 Concentratiekamp
97 Anders

**RE034** INCOME DURING GAP AFTER LEAVING THIS JOB

Welke bronnen van inkomsten had u?

1. Inkomen van echtgeno(o)t(e) of partner
2. Financiële steun van familie (geen echtgeno(o)t(e) of partner) en vrienden
3. Particuliere of algemene arbeidsongeschiktheidsverzekering
4. Uitkering van de overheid of verzekeringmaatschappij
5. Woning verkocht
6. Spaargeld opmaken
7. Anders

IF RE034[jobCnt] = 95 THEN

**RE034_other** OTHER INCOME DURING GAP AFTER LEAVING THIS JOB

Welke andere bronnen van inkomsten bedoelt u?

Memo

ENDIF

ENDIF

END IF

ENDDO

ENDIF

**RE045_end** END OF WORK HISTORY SECTION

Dit is het einde van het deel over uw werkverleden.
HS001_start  START OF THE HEALTH HISTORY SECTION
Nu volgen wat vragen over uw gezondheid.

HS003  HEALTH IN GENERAL
Hoe zou u uw gezondheid nu omschrijven?
1 Geweldig
2 Zeer goed
3 Goed
4 Matig
5 Slecht

HS005  CHILDHOOD HEALTH MISSED SCHOOL FOR 1 MONTH+
Bent u gedurende uw jeugd ooit langer dan een maand thuis (al dan niet in bed) gebleven om een gezondheidsreden? Met jeugd bedoelen we vanaf uw geboorte tot en met 15-jarige leeftijd.
1 ja
5 nee

HS006  CHILDHOOD HEALTH: IN HOSPITAL FOR 1 MONTH+
Hebt u gedurende uw jeugd ooit langer dan een maand in het ziekenhuis gelegen om een gezondheidsreden?
1 ja
5 nee

IF Gender = Vrouw THEN

| HS049  START OF MENSTRUAL PERIOD
| In welk jaar begon uw eerste menstruatie?
| Range: TYear

| IF START OF MENSTRUAL PERIOD = EMPTY THEN

| HS050  ESTIMATE START OF MENSTRUAL PERIOD
| Hoe oud was u ongeveer?
| 1. jonger dan 11 jaar
| 2. tussen 11 en 12 jaar
| 3. tussen 13 en 15 jaar
| 4. tussen 16 en 18 jaar
| 5. ouder dan 18 jaar

| ENDIF

| HS051  In which year did you have your last period or menstrual bleeding
| In welk jaar menstrueerde u voor het laatst? Vult u alstublieft 9997 in als u nog menstrueert.
| Range: TYear9997
EVER HAD PHYSICAL INJURY TO DISABILITY
Hebt u ooit lichamelijk letsel opgelopen dat leidde tot een permanente handicap, invaliditeit, of beperkingen met wat u in het dagelijkse leven kunt doen?
1 ja
5 nee

IF EVER HAD PHYSICAL INJURY TO DISABILITY = ja THEN

TYPE OF INJURY
Wat voor soort letsel was dit?
String

WHEN RECEIVED THIS INJURY
In welk jaar liep u dit letsel op? Als u meer dan één keer letsel hebt opgelopen, geeft u dan de eerste keer op.
Range: TYear

NUMBER PERIODS OF ILL HEALTH
Hoeveel periodes van slechte gezondheid of invaliditeit, die langer dan een jaar duurden, heeft u gehad (naast alle letsels die u mogelijk eerder hebt opgegeven)?
1 Geen
2 Een
3 Twee
4 Drie
5 Meer dan drie
6 Ik ben de meeste tijd van mijn leven ziek geweest

TYPE 1 OF ILLNESS FOR PERIODS OF ILL HEALTH
Welke van de volgende gezondheids problemen hebt u ooit gehad (als volwassene)?
1 Rugpijn
2 Artritis, inclusief osteoartritis en reuma
3 Osteoporose
4 Angina (hartkwaal) of een hartaanval (inclusief hartinfarct of or coronair trombose)
5 Andere hartaandoening
6 Diabetes of een te hoog bloedsuikergehalte
7 Beroerte
8 Astma
9 Problemen aan de luchtwegen anders dan astma (bijv. chronische bronchitis en longemfyseem)
10 Tuberculose (TBC)
11 Ernstige hoofdpijn of migraine
HS056  TYPE 2 OF ILLNESS FOR PERIODS OF ILL HEALTH
Hier is een tweede lijst met gezondheidsproblemen. Welke van de volgende
problemen zorgden voor slechte gezondheid of invaliditeit (die u als volwassene
had)?
1 Leukemie of lymfoom
2 Kanker of kwaadaadige tumor (exclusief minder ernstige vormen van huidkanker)
3 Emotionele, zenuw, of psychiatrische problemen
4 Vermoeidheid, bijv. bij ME, MS
5 Gynaecologisch (vrouwen) problem
6 Problemen met gezichtsvermogen
7 Infectieziektes (gordelroos, bof, TBC, HIV)
8 Allergie (anders dan astma, bijv. voedselintolerantie, hooikoorts)
96 Geen van deze
97 Andere problemen

IF TYPE 2 OF ILLNESS FOR PERIODS OF ILL HEALTH = Andere problemen THEN

| HS057  SPECIFY OTHER SERIOUS CONDITION |
| Welke ander gezondheidsprobleem bedoelt u? |
| Memo |
| ENDIF |

IF (cardinal( TYPE 1 OF ILLNESS FOR PERIODS OF ILL HEALTH ) > 0 AND
NOT(Geen van deze IN TYPE 1 OF ILLNESS FOR PERIODS OF ILL HEALTH )) OR (cardinal(TYPE 2
OF ILLNESS FOR PERIODS OF ILL HEALTH ) > 0 AND NOT(Geen van deze IN TYPE 2
OF ILLNESS FOR PERIODS OF ILL HEALTH )) THEN

| HS059  START OF PERIOD OF HEALTH PROBLEMS |
| In welk jaar begon deze periode van gezondheidsproblemen? |
| Range: TYear |

| HS060  END OF PERIOD HEALTH PROBLEMS |
| In welk jaar eindigde deze periode van gezondheidsproblemen? Vult u
alstublieft 9997 in als deze periode nog niet beëindigd is. |
| Range: TYear9997 |

| HS061  HELP FROM FAMILY OR FRIENDS DURING HEALTH PROBLEMS |
| Heeft uw familie of hebben vrienden u geholpen in deze tijd, bijvoorbeeld door
geld te geven of u te verplegen? |
| 1 Helemaal niet |
| 2 Een beetje
| 3 Heel veel
| 4 niet van toepassing

**HS062_intro**  WORKING DURING THIS TIME

Hebt u gewerkt (al dan niet part-time) in deze tijd?
1 ja
5 nee

**IF WORKING DURING THIS TIME = ja THEN**

**HS062** EXPERIENCED CONSEQUENCES OF ILLNESS

Hebt u ooit een van de volgende situaties meegemaakt door uw gezondheidsproblemen?
1 Geweigerde promotie
2 Projecten van mindere kwaliteit toegekend gekregen
3 Kritiek van baas of collegas
4 Salaris verlaging
9 Ik heb niet gewerkt tijdens deze periode
96 Geen van bovenstaande

**ENDIF**

**HS063** LONG TERM EFFECTS OF HEALTH PROBLEMS

Welke langere termijn effecten hebben gezondheidsproblemen gehad op uw leven?
1 Beperking in mijn mogelijkheden op het werk
2 Negatieve invloed op mijn gezinsleven
3 Positieve invloed op mijn gezinsleven
4 Moeilijkheden in mij sociale leven
5 Moeilijkheden bij het invullen van mijn vrije tijd
6 Overtuigd om het beste te maken van het leven
7 Nieuwe mogelijkheden geopend
96 Geen van bovenstaande
97 Anders

**IF LONG TERM EFFECTS OF HEALTH PROBLEMS = Anders THEN**

**HS064** OTHER LONG TERM EFFECTS

Welke andere effecten bedoelt u?
Memo

**ENDIF**

**ENDIF**

**HS065_end** END OF HEALTH SECTION.
Dit is het einde van de vragen over uw gezondheid.
INTRODUCTION TO GENERAL LIFE QUESTIONS
Tot slot zijn er nog een paar algemene vragen over uw leven.

PERIOD OF HAPPINESS
Als u terugkijkt op uw leven, was er een bepaalde periode waarin u gelukkiger was dan in de rest van uw leven?
1 ja
5 nee

IF PERIOD OF HAPPINESS = ja THEN

WHEN HAPPINESS PERIOD STARTED
Wanneer is deze periode van geluk begonnen?
Range: TYear

WHEN HAPPINESS PERIOD STOPPED
Wanneer stopte deze periode? Vult u alstublieft 9997 in wanneer de periode nog niet afgelopen is.
Range: TYear9997

ENDIF

PERIOD OF STRESS
Was er een bepaalde periode waarin u onder meer stress stond vergeleken met de rest van uw leven?
1 ja
5 nee

IF PERIOD OF STRESS = ja THEN

WHEN STRESS PERIOD STARTED
Wanneer is deze stressperiode begonnen?
Range: TYear

WHEN STRESS PERIOD STOPPED
Wanneer is deze periode gestopt? Vult u alstublieft 9997 in als de periode nog niet afgelopen is.
Range: TYear9997

ENDIF

PERIOD OF POOR HEALTH
Was er een bepaalde periode waarin uw gezondheid slecht was vergeleken met de rest van uw leven?
1 ja
5 nee
IF PERIOD OF POOR HEALTH = ja THEN

| GL032 WHEN POOR HEALTH PERIOD STARTED  |
| Wanneer is deze periode van slechte gezondheid begonnen? |
| Range: TYear |

| GL033 WHEN POOR HEALTH PERIOD STOPPED  |
| Wanneer is deze periode gestopt? Als deze periode nog niet afgelopen is, vult u dan alstublieft 9997 in. |
| Range: TYear9997 |

ENDIF

GL002 PERIOD OF FINANCIAL HARDSHIP
Was er een bepaalde periode waarin u het financieel zwaar had?
1 ja
5 nee

IF PERIOD OF FINANCIAL HARDSHIP = ja THEN

| GL003 WHEN FINANCIAL HARDSHIP PERIOD STARTED  |
| Wanneer is deze periode van financiële tegenspoed begonnen? |
| Range: TYear |

| GL004 WHEN FINANCIAL HARDSHIP PERIOD STOPPED  |
| Wanneer is deze periode gestopt? Als deze periode nog niet geeindigd is, vult u dan alstublieft 9997 in. |
| Range: TYear9997 |

ENDIF

GL040 DISCRIMINATED AGAINST
Er zijn periodes waarin mensen vervolgd of gediscrimineerd worden, bijvoorbeeld vanwege hun politieke opvattingen, religie, nationaliteit, afkomst, seksuele voorkeur, of hun achtergrond. Mensen kunnen ook vervolgd of gediscrimineerd worden vanwege de politieke opvattingen of de religie van naaste familieleden. Bent u ooit slachtoffer geweest van een dergelijke vervolging of discriminatie?
1 ja
2 nee
IF DISCRIMINATED AGAINST = ja THEN

| GL041 MAIN REASON OF PERSECUTION |
| Wat was de hoofdreden dat u werd vervolgd of gediscrimineerd? |
| 1 politieke opvattingen |
| 2 religieuze opvattingen |
| 3 afkomst of nationaliteit |
| 4 seksuele voorkeur |
| 5 achtergrond |
| 6 politieke opvattingen of religie van naaste familieleden |
| 7 andere reden |

ENDIF

GL042 INTRODUCTION TO FINAL QUESTION
Tot nu toe hebben we u gevraagd naar bepaalde aspecten van uw leven. We begrijpen dat er andere aspecten van uw leven zijn die belangrijk zijn.

GL043 ANYTHING ELSE THAT HAS HAPPENED
Is er iets anders gebeurd in uw leven waarover u ons wilt vertellen?
1 ja
2 nee

IF ANYTHING ELSE THAT HAS HAPPENED = ja THEN

| GL044 DETAILS OF WHAT ELSE HAPPENED |
| Wilt u in het kort aangeven waar het om gaat? |
| Memo |

ENDIF

CQ001 bronnen
Tot slot nog wat algemene vragen.

Welke van de volgende informatiebronnen hebt u tijdens het invullen van de vragenlijst geraadpleegd? (U kunt meerdere antwoorden aanvinken)
1 persoonlijke administratie
2 dagboek of agenda
3 overleg met partner/derden
4 anders, namelijk [CQ001_other bronnen anders]
5 geen, dat was niet nodig
IF bronnen = 1/2/3/4 THEN
| CQ002 gebruik bronnen per thema
| U hebt aangegeven dat u bij het invullen van de vragenlijst een of meerdere informatiebronnen heeft geraadpleegd. Bij welke van de volgende thema's heeft u deze informatiebronnen gebruikt? (u kunt meerdere antwoorden aanvinken)
| 1 Vragen over relaties
| 2 Vragen over wonen/adressen
| 3 Vragen over uw werkverleden
| 4 Vragen over gezondheid
| 5 Anders, namelijk [CQ002_other gebruik bronnen per thema anders]
ENDIF

IF calenderType = 2 OR calenderType = 3 THEN
| CQ003 voorkeur wel of geen kalender
| U hebt zojuist een vragenlijst ingevuld waarbij uw antwoorden in een kalender werden weergegeven. Als we nog een keer een soortgelijke vragenlijst zouden afdrukken, zou u dan de voorkeur geven aan dit nieuwe formaat, of zou u de vragenlijst liever op de gebruikelijke manier op het scherm willen zien?
| 1 Liever met kalender
| 2 Liever zonder kalender
ENDIF

IF calenderType = 4 THEN
| CQ004 voorkeur wel of geen tijdsbalk
| U hebt zojuist een vragenlijst ingevuld waarbij u een tijdsbalk met persoonlijke gebeurtenissen als hulpmiddel kon gebruiken. Als wij nog een keer een soortgelijke vragenlijst zouden afdrukken, zou u dan de voorkeur geven aan dit nieuwe formaat, of zou u de vragenlijst liever op de gebruikelijke manier op het scherm willen zien?
| 1 Liever met tijdsbalk
| 2 Liever zonder tijdsbalk
ENDIF

| CQ005 intro willekeurige vragen
| We weten dat het soms lastig kan zijn om zich bepaalde data en gebeurtenissen te herinneren. Om die reden willen wij u voor een aantal willekeurige antwoorden vragen of u kunt aangeven hoe zeker of onzeker u daarvan was.
IF AC007[2] = RESPONSE THEN

   CQ006 check antwoord vraag verhuisd
   In welk jaar bent u verhuisd van de eerste woning waar u zelfstandig woonde naar uw tweede?
   Uw antwoord op deze vraag luidde [AC007[2]].
   Hoe zeker bent u van dit antwoord?
   1 1 Zeer onzeker
   2 2
   3 3
   4 4
   5 5 Zeer zeker

ENDIF

IF RE002 = RESPONSE AND RE002<9997 THEN

   CQ007 check antwoord vraag onderwijs
   In welk jaar bent u gestopt met voltijds onderwijs?
   Uw antwoord op deze vraag luidde [RE002].
   Hoe zeker bent u van dit antwoord?
   1 1 Zeer onzeker
   2 2
   3 3
   4 4
   5 5 Zeer zeker

ENDIF

IF RE026[1] = RESPONSE THEN

   CQ008 check antwoord vraag werk
   In welk jaar bent u gestopt met werken als [RE012[1]]?
   Uw antwoord op deze vraag luidde [RE026[1]].
   Hoe zeker bent u van dit antwoord?
   1 1 Zeer onzeker
   2 2
   3 3
   4 4
   5 5 Zeer zeker

ENDIF

ENDIF
Samenvatting

Bij het beantwoorden van retrospectieve vragen in survey-onderzoek treden vaak herinneringsproblemen op, die tot onvolledigheden en tot fouten bij de datering van gebeurtenissen kunnen leiden. Om deze problemen te verhelpen zijn er zogenaamde aided recall technieken ontwikkeld, die in surveys kunnen worden gebruikt. Mijn proefschrift beschrijft de toepassing en effecten van kalendermethoden (soms ook tijdsbalken genoemd) die als geheugensteun binnen, maar ook als volwaardig alternatief voor een gestandaardiseerde vragenlijst kunnen dienen (zie figuur A).

<table>
<thead>
<tr>
<th>2009</th>
<th>JANUARI</th>
<th>FEBRUARI</th>
<th>MAART</th>
<th>APRIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEEFTIJD</td>
<td>————</td>
<td>48</td>
<td>————</td>
<td>————</td>
</tr>
<tr>
<td>ADRES</td>
<td>A’dam, Parkstraat</td>
<td>————</td>
<td>A’dam, Kerkstraat</td>
<td>————</td>
</tr>
<tr>
<td>SAMENSTELLING HUISHOUDEN</td>
<td>partner en zoon</td>
<td>partner</td>
<td>————</td>
<td>————</td>
</tr>
<tr>
<td>WERK &amp; OPLEIDING</td>
<td>ING bank</td>
<td>————</td>
<td>————</td>
<td>Business English</td>
</tr>
<tr>
<td>PERSOONLIJKE GEBEURTENISSEN</td>
<td>Zoon op kamers</td>
<td>Spanje</td>
<td>Sarah jarig</td>
<td>————</td>
</tr>
</tbody>
</table>

Figuur A: Voorbeeld kalendermethode
Het ontwerp van deze kalenders is gebaseerd op de gedachte dat het voor de respondent makkelijker is om vragen over gebeurtenissen uit het verleden te beantwoorden wanneer hij of zij zich hierbij kan baseren op informatie over andere gebeurtenissen (bijv. “Nadat ik mijn baan in Deventer was kwijtgeraakt, verhuisden wij naar Utrecht”).

In de afgelopen jaren zijn kalendermethoden in verschillende grootschalige panelonderzoeken toegepast, zoals de *German Life History Study (GLHS)*, de *Survey of Health, Ageing and Retirement in Europe (SHARE)*, en de *Panel Study of Income Dynamics (PSID)*. Eerder onderzoek heeft uitgewezen dat het gebruik van kalenders tot verbeteringen in de kwaliteit van de verzamelde data kan leiden (Belli, Lee, Stafford, & Chou, 2004; Belli, Shay, & Stafford, 2001, Van der Vaart, 1996). Toch is er ondanks de stijgende belangstelling voor kalendermethoden relatief weinig bekend over hun concrete effecten. Ook wordt de ontwikkeling van dergelijke methoden soms ontoereikend theoretisch onderbouwd.

Momenteel is er een groot aantal verschillende kalendermethoden in omloop, die drie aspecten gemeen hebben. Ten eerste bevatten zij een grafische weergave van de tijdsperiode waarover gegevens worden verzameld. Afhankelijk van de totale lengte wordt deze periode onderverdeeld in kortere tijdseenheden, zoals jaren, maanden of dagen. Ten tweede heeft de kalender betrekking op een of meerdere levensdomeinen, zoals wonen, werken of gezondheid, of op concrete activiteiten, zoals doktersbezoeken, die op een of meerdere tijdslijnen in de kalender worden weergegeven. Ten derde heeft de respondent temporele ankerpunten, de zogenaamde *landmarks*, tot zijn of haar beschikking. Vaak zijn dit persoonlijke of publieke gebeurtenissen uit de referentieperiode van het onderzoek die de respondent als aanknopingspunten kan gebruiken om de data van andere gebeurtenissen te herleiden.

Veel onderzoekers gebruiken publieke *landmarks* die van tevoren in de kalender kunnen worden gezet, maar persoonlijke gebeurtenissen - zoals geboortes, huwelijken of verhuizingen - worden vaak als effectievere *landmarks* beschouwd (Means et al., 1988), wellicht omdat zij een centralere plek in het geheugen innemen. Daarnaast is er vaak sprake van causale verbanden tussen verschillende persoonlijke gebeurtenissen (bijvoorbeeld een aanpassing van het aantal gewerkte uren naar aanleiding van de geboorte van een kind) die eveneens het achterhalen van data kunnen vergemakkelijken. Om als *landmark* te kunnen
dienen, moet een persoonlijke gebeurtenis door de respondent makkelijk te dateren zijn en het liefst met een precieze datum in het geheugen zijn opgeslagen.

In mijn proefschrift beschrijf ik de relatie tussen designkenmerken van kalendermethoden, cognitieve processen die bij het herinneren van autobiografische gebeurtenissen een rol spelen en de kwaliteit van retrospectieve survey data. Het doel van het proefschrift is het versterken van de theoretische onderbouwing van kalenderinstrumenten en de ontwikkeling van een kalendermethode voor het gebruik in online surveyonderzoek.
Onderzoeksvragen

In het proefschrift komen de volgende onderzoeksvragen aan de orde:

1. Wat is er bekend over de effecten van kalendermethoden op de kwaliteit van retrospectief verzamelde survey data?
2. Welke cognitieve processen spelen een rol bij het beantwoorden van retrospectieve vragen? Is er met betrekking tot deze cognitieve processen een verschil tussen inter views met gebruik van kalendermethoden en interviews waarin geen kalender wordt gebruikt?
3. Welke onderdelen van de kalendermethode zijn het meest relevant voor het verbeteren van de datakwaliteit in online retrospectieve levenslooponderzoeken?
4. Hoe kunnen kalendermethoden verder worden verbeterd met oog op de toepassing in online levenslooponderzoek?

Om deze onderzoeksvragen te kunnen beantwoorden is eerst een uitgebreide literatuurstudie verricht. Hoofdstuk 1 van het proefschrift is een inleiding in de problematiek van retrospectieve onderzoeksdesigns. In dit hoofdstuk wordt onder andere een aantal herinneringsfouten besproken die in retrospectieve surveys kunnen optreden, zoals het weglaten van gebeurtenissen (”omissie”) en dateringsfouten. Ook komen hier verschillende technieken aan bod die men bij het afnemen van survey-interviews kan gebruiken om de respondenten bij het herinneren van relevante informatie te ondersteunen. Bij de bespreking van deze technieken wordt in het bijzonder ingegaan op de bovengenoemde kalendermethoden. In hoofdstuk 2 wordt een overzicht gegeven van recent onderzoek naar het autobiografisch geheugen. Hierbij is een bijzondere rol weggelegd voor hiërarchische geheugenmodellen zoals beschreven door o.a. Conway (1996), die reeds eerder gebruikt zijn als theoretische onderbouwing van de werking van kalendermethoden.
Literatuuronderzoek naar de toepassing en effecten van kalenderinstrumenten

In het derde hoofdstuk van het proefschrift worden de resultaten van een literatuuronderzoek naar de toepassing en effecten van verschillende kalendermethoden weergegeven. Een klein aantal in de literatuur beschreven methodologische experimenten toont aan dat kalenderinstrumenten een gematigd positief effect op de datakwaliteit in retrospectieve surveys kunnen hebben. Bij deze aspecten horen zowel de volledigheid en logische samenhang van de verzamelde data als de nauwkeurigheid van de datering van gebeurtenissen uit het verleden.

Met betrekking tot de volledigheid van de data kan de reductie van ‘gaten’ in de tijdslijn (bijvoorbeeld door het weglaten van een of meerdere banen in een werkgeschiedenis) als een van de belangrijkste gevolgen van het gebruik van kalendermethoden worden beschouwd. In enkele van de beschreven studies leidde het gebruik van een kalendermethode ook tot grotere aantallen gerapporteerde gebeurtenissen. Dit was met name het geval wanneer het om relatief moeilijk te onthouden informatie ging, zoals zeer frequente of als redelijk onbelangrijk ervaren gebeurtenissen die al dan niet in het verre verleden hadden plaatsgevonden.

Verder blijkt uit een aantal onderzoeken dat ook de logische samenhang van de door de respondent verstrekte gegevens soms verhoogd wordt door het toepassen van een kalendermethode. Dit houdt in dat respondenten in mindere mate tegenstrijdige gebeurtenissen rapporteren, zoals overlap tussen een betaalde voltijdbaan en een periode van werkloosheid. Met betrekking tot de reductie van dateringsfouten zijn de resultaten van eerder onderzoek niet eenduidig. Vaak wordt wel een reductie in de grootte van de dateringsfout bereikt, maar geen verschil in richting van de fout, d.w.z. dat data nog steeds systematisch te vroeg of te laat worden ingeschat.

Een ander interessant gegeven dat uit het literatuuronderzoek naar voren komt, is dat de manier waarop de vragenlijst wordt afgenomen, d.w.z. telefonisch, schriftelijk of persoonlijk, geen verstrekende gevolgen had met betrekking tot het soort effecten van de methode op de datakwaliteit. Het feit dat kalendermethoden ook de datakwaliteit in telefonische interviews verbeteren waarbij de respondent de kalender niet kan zien, wijst erop dat de methode niet alleen een nuttige geheugensteun is voor de respondent. Kalenders onder-
steunen ook de interviewer bij het controleren van de gerapporteerde gegevens op overlap en andere ongerijmdheden (Reimer & Matthes, 2007) en zorgen voor deze manier voor een verbetering van de datakwaliteit.

Uit het literatuuronderzoek kan worden geconcludeerd dat de resultaten van eerder onderzoek over kalendermethoden een gematigd positief effect laten zien. Dit effect lijkt echter afhankelijk te zijn van het onderwerp van de survey en van de moeilijkheidsgraad van de herinneringstaak. Gebaseerd op zijn eigen onderzoek stelt Belli (2007) dat de positieve werking van kalendermethoden wellicht grotendeels beperkt blijft tot zeer complexe vragen naar bijvoorbeeld gedetailleerde werkgeschiedenissen. Andere informatie die makkelijk uit het geheugen opgehaald kan worden (bijv. het geboortejaar van de eigen kinderen), zou net zo goed met een reguliere vragenlijst zonder geheugensteun kunnen worden verzameld.
Cognitieve interviews

Hoofdstuk 4 beschrijft een pilotonderzoek naar de geheugenprocessen in interviews met en zonder kalenderhulpmiddelen. In dit onderzoek is gebruik gemaakt van cognitieve interviews, een methode waarbij aan proefpersonen wordt gevraagd om tijdens het beantwoorden van surveyvragen hardop na te denken. Op deze manier hoopt de onderzoeker inzicht te krijgen in de manier waarop survey-respondenten vragen begrijpen en antwoorden herleiden. De door ons verrichte cognitieve interviews laten geen eenduidig effect zien van kalendermethoden op de manier waarop men de data van autobiografische gebeurtenissen uit het geheugen ophaalt. Gezien het kleine aantal proefpersonen dat aan dit onderzoek meedeed, is het echter niet mogelijk om hier definitieve conclusies uit te trekken.

Desalniettemin levert het cognitieve pilotonderzoek een aantal nuttige inzichten op met betrekking tot de manier waarop respondenten publieke en persoonlijke landmarks gebruiken om andere gebeurtenissen te dateren. Het aantal landmarks dat voorafgaand aan het onderzoek door de proefpersonen werd genoemd, varieerde sterk en was afhankelijk van de leeftijd van de persoon. Oudere deelnemers noemden een kleiner aantal landmarks dan jongere deelnemers. Ook bleek het aantal genoemde landmarks samen te hangen met het gemak waarmee men zich de data van gebeurtenissen precies herinnerde (de zogenaamde direct date retrievals). Er is echter geen enkel verband gevonden tussen het aantal genoemde landmarks en het aantal keren dat de proefpersoon daadwerkelijk een landmark gebruikte om een andere gebeurtenis te dateren. In het onderzoek werd eveneens geen verschil gevonden tussen interviews met en zonder kalender, niet alleen wat betreft het aantal keren dat men persoonlijke gebeurtenissen als ankerpunt gebruikte maar ook in het type landmark dat hiervoor werd gebruikt. Hieruit trek ik de conclusie dat cognitieve processen in retrospectieve surveys mogelijk in mindere mate afhankelijk zijn van de gebruikte methode van dataverzameling dan van kenmerken van de respondent en van de herinneringstaak.
Experimentele evaluatie van een online kalendermethode

Om de vraag te beantwoorden welke onderdelen van de kalendermethode het meest relevant zijn voor het bevorderen van de datakwaliteit in online retrospectieve levenslooponderzoeken (onderzoeksvraag 3) is een grootschalig methodologisch evaluatieonderzoek opgezet. Hoofdstuk 5 beschrijft dit 2x2 online experiment, waarin drie verschillende versies van een kalendermethode werden vergeleken met een reguliere retrospectieve levensloopvragenlijst. Respondenten in de controleconditie vulden een reguliere online vragenlijst in waarbij geen gebruik werd gemaakt van aided recall technieken. Respondenten in de experimentelecondities beantwoordden dezelfde vragenlijst maar beschikten daarnaast over een in de vragenlijst geïntegreerde kalender met visuele feedback van hun antwoorden (conditie 2), persoonlijke landmark events (conditie 3) of beide componenten (conditie 4).

Uit het onderzoek komt voort dat de effecten van de kalendermethode in de twee condities waarin de respondenten visuele feedback kregen (d.w.z. de ‘vragenlijst plus visuele feedback’- en de ‘vragenlijst plus visuele feedback en landmarks’-conditie) in belangrijke mate overeen kwamen met eerdere in de methodologische literatuur beschreven bevindingen. De respondenten in deze condities rapporteerden significant meer onderbrekingen in hun werkgeschiedenis door zorgverlof en door werkloosheid dan respondenten in de condities zonder visuele feedback. Dit is in overeenstemming met door Belli et al (2007) gevonden positieve effecten van Event History Calendars op het aantal in een levenslooponderzoek gerapporteerde periodes van werkloosheid. Het gevonden effect was groter voor periodes van werkloosheid die verder in het verleden lagen dan voor recentere werkloosheid, een effect dat op basis van de resultaten van diverse eerdere studies kon worden verwacht (bijv. Van der Vaart, 2004; Yoshihama et al., 2005). Over het algemeen waren deze effecten van de kalendermethode klein maar significant.

Het visuele feedback leek ook het aantal fouten dat men bij het berekenen van de duur van kortere periodes in de levensloop maakte licht te verlagen. Er werd echter geen ander positief effect van de methode op de logische samenhang van gebeurtenissen, zoals de coherentie van begin- en einddata gevonden. Ook verhoogde het visuele feedback niet het aantal belangrijke banen, leefadressen en relaties dat de respondent noemde. Het gebrek aan effecten met betrekking tot dit type gebeurtenissen zou gedeeltelijk verklaard kunnen
worden door het wellicht geringe aantal fouten dat men ook zonder geheugensteun zou maken (zie ook Van der Vaart, 1996). Respondenten in de condities met visuele feedback verbeterden hun antwoorden vaker dan respondenten die geen visuele feedback kregen. Het is mogelijk dat dit gegeven een (gedeeltelijke) verklaring geeft voor de in het onderzoek gevonden positieve effecten van visuele feedback op de datakwaliteit.

Terwijl het aanbieden van visuele feedback een grotendeels positief effect op de datakwaliteit had, kon er geen verbetering worden gevonden die duidelijk te wijten was aan het opnemen van persoonlijke landmarks in de vragenlijst. Door het toevoegen van de vragen over *landmark events* werd de duur van het interview significant verlengd ten opzichte van de controleconditie en de conditie met enkel visuele feedback. Ook was er in de landmark condities een verhoogde kans dat de respondent het invullen van de vragenlijst voor- tijdig, vaak nog voordat alle *landmarks* waren ingevuld, afbrak.
Conclusie en discussie

Hoofdstuk 6 geeft een samenvatting van de belangrijkste bevindingen uit de eerdere hoofdstukken. Ten aanzien van de in het begin van het proefschrift genoemde onderzoeksvragen wordt geconcludeerd dat de tot op heden in de literatuur vermelde effecten van kalendermethoden voornamelijk positief zijn, maar afhankelijk zijn van het onderwerp van het onderzoek en de moeilijkheidsgraad van de herinneringstaak (onderzoeksvraag 1). In het cognitieve pilotonderzoek konden geen noemenswaardige verschillen worden aangetoond tussen geheugenprocessen in interviews met en zonder een kalendermethode (onderzoeksvraag 2). Het in hoofdstuk 5 beschreven evaluatieonderzoek werd uitgevoerd om een onderscheid te kunnen maken tussen de effecten van *landmark events* en de effecten van visuele feedback in de context van online retrospectief levenslooponderzoek. Er werden in dit hoofdstuk inderdaad uiteenlopende effecten van deze twee designkenmerken (*landmarks* en visuele feedback) op verschillende aspecten van datakwaliteit gevonden. De belangrijkste bevinding van dit hoofdstuk is dat de positieve effecten op de volledigheid en logische samenhang van de verzamelde data te wijten waren aan het visuele feedback dat de respondent te zien kreeg. De in het begin van het interview verzamelde *landmarks* bleken hier geen positief effect op te hebben. Een verrassend resultaat betreft de hoge uitval in de condities waarin respondenten *landmarks* moesten noemen. Het is aannemelijk dat een deel van de uitval met een gebruiksvriendelijker design van de landmarkvragen voorkomen had kunnen worden. Desalniettemin moet worden opgemerkt dat het grootste gedeelte van de uitval al in het begin van het interview plaatsvond, nog voordat de respondent de eerste landmarkvraag had beantwoord. Dit wijst erop dat alleen al het stellen van landmarkvragen waarvan de bedoeling niet meteen voor de respondent duidelijk wordt tot een hogere uitval in een online survey kan leiden (onderzoeksvraag 3). Behalve de relatief hoge ontwikkelingskosten voor de kalendermethode zijn er geen aanwijzingen dat het aanbieden van visuele feedback negatieve gevolgen zou hebben voor het onderzoeksproces. Op basis van deze bevindingen wordt de aanbeveling gedaan om kalendermethoden te gebruiken waarin visueel feedback wordt gegeven. Voorlopig zijn er geen resultaten gevonden die tot de aanbeveling zouden leiden om in het begin van retrospectieve online surveys ook *landmark events* te verzamelen (onderzoeksvraag 4).