Institute for Employment Research

The Research Institute of the Federal Employment Agency

IAB

IAB-Discussion Paper 21/2009

Articles on labour market issues

Improving retrospective life course data by combining modularized self-reports and event history calendars

Experiences from a large scale survey

Katrin Drasch
Britta Matthes

Improving retrospective life course data by combining modularized self-reports and event history calendars

Experiences from a large scale survey

Katrin Drasch (IAB)
Britta Matthes (IAB)

Contents

Abstract 4
Zusammenfassung 4
1 Introduction 5
2 Previous Research on Recall Differences 6
3 Theoretical Background: The Structure of Autobiographical Memory and Memory Pathways 8
3.1 Hierarchical Pathways to Stimulate Top-Down Retrieval 8
3.2 Separate, Chronological Time-Lines to Stimulate Parallel Recall 9
3.3 Sequential Pathways to Stimulate Dating of Episodes 9
3.4 Autobiographic Memory Retrieval Problems 9
4 Stimulation of Autobiographical Memory: Combining Modularized Self-Reports and Event History Calendars 10
4.1 Modularized Self-Reports as Tool for Data Collection 10
4.2 Event History Calendar as Recall Aid 12
5 Data, Measurement, and Empirical Methods 13
6 Completeness and Dating Accuracy in ALWA 16
6.1 Completeness: Added Episodes 16
6.1.1 Descriptive Results 16
6.1.2 Influencing Factors on Adding Episodes 18
6.2 Dating Accuracy: Corrected Episodes 19
6.2.1 Descriptive Results 19
6.2.2 Influencing Factors on Correcting Episodes 21
6.3 Discussion of the Major Findings 22
7 Conclusion 23
References 25
Appendix 1: Statistics 29
Appendix 2: Technical implementation 34
Recently published 36

Abstract

Event history calendars (EHC) have proven to be a useful tool to collect retrospective autobiographic life course data. One problem is that they are only standardized to some extent. This limits their applicability in large-scale surveys. However, in such surveys a modularized retrospective CATI design can be combined with EHC. This data revision module is directly integrated into the interview and used as a data revision module. Hereby insights from cognitive psychology are applied. The data revision module stimulates the respondent's memory retrieval by detecting both temporal inconsistencies, such as gaps, and overlapping or parallel events. This approach was implemented in the IAB-ALWA study (Work and Learning in a Changing World), a large-scale representative telephone survey with 10,000 respondents. By comparing the uncorrected data with the final data after revision, we investigate to what extent the application of this data revision module improves data quality or more precisely, time consistency and dating accuracy of individual reports.

Zusammenfassung

Biografische Kalender (EHC) haben sich als ein nützliches Instrument zur Erfassung retrospektiver autobiografischer Lebensverlaufsdaten erwiesen. Problematisch ist allerdings, dass biographische Kalender nur zu einem gewissen Grad standardisiert sind, was deren Einsetzbarkeit in Bevölkerungsumfragen einschränkt. Allerdings ist es möglich in solchen Umfragen das modularisierte retrospektive CATI Erhebungsdesign mit einem biographischen Kalender zu einem Korrekturmodul zu kombinieren, welches Erinnerungsfehler bereits im Verlauf des Interviews aufspüren und gemeinsam mit den Befragten korrigieren kann. Dazu werden Erkenntnisse aus der Kognitionspsychologie genutzt. Das Korrekturmodul stimuliert das Erinnerungsvermögen des Befragten, indem es sowohl auf temporäre Inkonsistenzen, wie beispielsweise Lücken als auch auf überlappende und parallele Ereignisse hinweist. Dieser Ansatz wurde in der IAB-ALWA Studie (Arbeiten und Lernen im Wandel), einer repräsentativen Telefonumfrage mit 10.000 Befragten, verfolgt. Wir untersuchen in welchem Ausmaß das Korrekturmodul die Datenqualität oder genauer gesagt die zeitliche Konsistenz und die Datierungsgenauigkeit der einzelnen berichteten Ereignisse verbessert, indem wir die ungeprüften Daten mit den korrigierten Daten nach der Datenprüfung vergleichen.

Keywords: Event history calendars, large-scale survey, retrospective life course reports, survey methods

JEL Classification: C42, C81, C83

Acknowledgements: We would like to thank Bernhard Christoph and Holger Seibert, as well as participants of the RC 33 2008 conference in Naples and ESRA 2009 conference in Warsaw for valuable comments. We also thank Maria Fleischmann for proofreading. All remaining errors are of course ours.

1 Introduction

Retrospective life course surveys collect information on the characteristics and the dating of past episodes. Thus, event history data are extremely valuable for the analysis of life courses and long-term social change (Blossfeld/Golsch/Rohwer 2007). To additionally be useful for statistical analysis, event histories must as far as possible meet two criteria: they have to be complete and consistent. Completeness demands that information on one or more episodes is available for the whole time period under investigation. Information can, for instance, lack about an individual's activities after finishing school and before starting university. This indicates incomplete data, because we do not know whether the individual worked, was unemployed, or simply went on holidays during that time. The second criterion, consistency, means that the collected information is not contradictory, neither regarding the timing of events nor with respect to substantial matters. A temporal, as well as substantial contradiction is e.g. if an individual reported to be in military service and to be employed full-time simultaneously. In sum, incomplete or inconsistent life course reports are considered visible signs of individual autobiographical memory failure.

In recent years, applied researchers have used increasingly aided recall techniques to overcome recall errors. Various methods, such as the use of landmark events (e.g. a persons marriage date), the reconstruction of autobiographic information with event history calendars (EHC), or the collection of data in a domain specific context with modularized reports can be applied to stimulate autobiographical memory. In several experimental studies the effectiveness of recall methods has been demonstrated (e.g. Belli et al. 2004; Belli/Shay/Stafford 2001; Van der Vaart 2004). Applying those methods has an influence on recall accuracy in terms of completeness, dating accuracy and characteristics of an episode (for an overview see Glasner/van der Vaart 2009).

In the IAB-ALWA study we use a combination of aided recall techniques throughout the interview. ALWA (Work and Learning in a Changing World), a project located at the Institute for Employment Research (IAB) adopts this hybrid design. For an overview of this survey we refer to Kleinert, Matthes, and Jacob (2008). We combine modularized self-reports and event history calendars to improve consistency and completeness of the autobiographic data which were collected retrospectively. The event history calendar itself is used as a data revision module. We employ the EHC to detect and resolve inconsistencies in collaboration with the respondent during the interview, and not as a plausibility check after the interview process was finalized. The interview process itself is highly standardized. The implemented survey instrument uses insights from cognitive psychology and integrates these insights in the survey design. This is done in order to stimulate autobiographic memory better than it had been achieved in previous studies. Apart from experimental studies (cf. Matthes/Reimer/Künster 2007; Reimer/Matthes 2006; van der Vaart/Glasner 2007), this is the first time that such a hybrid design is applied in a large scale survey.

In past retrospective life course surveys, e.g. the German Life History Study (GLHS) the editing process was lengthy and costly because it took place completely outside the interview process itself. So, additional to the improvement of data quality, a reason for using this instrument is to guarantee the cost efficiency of the survey. This meant that each single case had to be edited after the interview by applying standardized rules for data editing. This does enhance data quality considerably (Matthes/Reimer/Künster 2007). However, editing life-courses without the respondent is associated to with a more severe and substantial problem. It is not always possible to derive comprehensive and unambiguous rules regarding the decision which of the given information is the most valid one without contacting the respondent again (cf. Hillmert 2002; Matthes/Reimer/Künster 2007) or using rules of thumb.

To our knowledge, no study exists that investigates the effectiveness of combined recall techniques in a large scale survey. Previous studies were of experimental design, using a split-ballot or assessed test-retest reliability. This paper enriches research on data quality of retrospective reports and adds an additional category (cf. Glasner/van der Vaart 2009) on how to assess data quality: this category compares individual reports in a large scale survey with and without the second stage of data collection. We thus oppose the data before the data revision module was entered with the one after completion of the revision module. By applying such an approach, survey design can be used to overcome memory problems at least to some extent. Moreover, this paper sheds light on the question whether certain groups of individuals benefit more from implying such recall techniques. Thus, the aim of this paper is twofold: experiences with this hybrid design in a large scale survey will be reported and more detailed insights on the improvements of data quality will be provided. In particular, we focus on two common recall errors: the omission and the misdating of episodes.

The outline of this paper is as follows. We start with some selected empirical results on recall problems. Next, we provide some insights into the structure of autobiographical memory and how this structure can be used to design a standardized survey instrument. We proceed with empirical results on the effectiveness of such an instrument applied as a data revision module in a large scale survey. Finally, a summary and some preliminary practical conclusions and suggestions for further research will be given.

2 Previous Research on Recall Differences

Intuitively, individual differences as well as characteristics of the episode itself matter for recalling an episode (Klein/Fischer-Kerli 2000). Both attributes have been examined in several experimental and quasi-experimental studies. Dating accuracy has been studied less frequently at least in survey research. A concise selection of the results is presented in this paragraph.

Individual characteristics such as age and gender have frequently been examined in empirical research. With regard to the education of the respondents, however, so far

little research exists. De Graaf and Wegener (1989) find that a higher level of education leads to more inconsistent answers. Regarding age Elias (1997) states that older age groups (over 60 years) show less agreement when reporting unemployment compared to younger age groups. Paull (2002) also uncovers that recall errors occur more frequently for older people. Gender differences in the autobiographical recall of events are frequently under study (cf. Auriat 1991; Paull 2002; Reimer/ Matthes 2006). Skowronski and Thompson (1990) conclude that women are better in both remembering and dating events. Klijzing and Prophet (1998) show that the failures to remember month or year of an event are more pronounced for men than for women. In addition they state that the recall lapse increases with age.

Furthermore characteristics of the episode itself have been examined, e.g. the episode's type and the complexity of life histories. Reimer and Matthes (2006) find differences in recall depending on the type of episode and the length of the episode. Respondents are more likely to forget short episodes and those that are regarded unimportant (Barclay 1986). Belli et al. (2007), confirming earlier research by Mathiowetz and Duncan (1988), state that especially unemployment spells are weak in data quality. In contrast, employment episodes are usually remembered well. Paull (1997) also shows that unemployment spells are especially error prone. Dex (1991) notices that short periods of unemployment are omitted frequently. The author also points out that among other factors the saliency of events and the temporal reference of events has an effect on human recall. Also similar events over time are difficult to remember as the time lapse increases (Groves 2000). Dating errors are influenced by the ability to remember the event at all. Often misdating an event by some systematic calendar related factor occurs, e.g. by one year (Huttenlocher/ Hedges/Bradburn 1990). Additionally those errors heavily depend on the time-lag of the episode and the time of potential retrieval (Rubin/Baddeley 1989). The omission of episodes is considered to be a function of the temporal distance to the event (Paull 2002). Peters (1988) finds that recall errors are reduced by 3 to 4 percent per year closer to the interview date. For providing these results, she assessed the respondent's family status with a maximum memory period of 15 years and an average retrospective interval of five years. With increasing complexity of life histories recall error might get worse, but little indication of this increasing error will probably become apparent in the data. The reason is that visible errors do not augment with complexity of lives, because respondents tend to smooth their life courses (Reimer 2001).

The results of previous experimental research lead to some general expectations about the effectiveness of the data revision module in the IAB-ALWA survey. On the one hand, we expect to find individual differences. First, we assume that men and women differ in remembering and dating of events. Second, we generally presume an age effect. Whether this effect will already be visible for the age group under study (18-52) has to be examined. Third, an educational effect meaning that higher educated people make less use of corrections is anticipated. However, we think that

this effect might be levelled out by the increasing complexity of life-courses of better educated individuals.

On the other hand we expect effects related to characteristics of the episode. First, we assume that the episode's type matters for remembering and dating episodes. Second, we assume that the longer an episode is, the more likely it will be remembered before entering the data revision module. Third, we hypothesize that the number of episodes already reported reduces the chance of recalling a further episode. Finally, we hypothesize that it is for both remembering and dating events important to evaluate how much time has elapsed since the episode took place. This denotes the time-lag to the interview date.

3 Theoretical Background: The Structure of Autobiographical Memory and Memory Pathways

In the following paragraph we give a brief overview of the most important attributes of autobiographical memory.[1] These attributes serve as the theoretical background for the design of the IAB-ALWA instrument. Cognitive psychology found that autobiographical human memory is organized as follows (Barsalou 1988; Conway 1996; Conway/Pleydell-Pearce 2000): Retrospections that are related to certain events in an individual's life course are organized in the long-time memory as a network of mental representations which are connected by different pathways. Hereby, three types of pathways are distinguished (cf. Reimer/Matthes 2006): Hierarchical pathways start with general memory and run to more specific memory. Sequential pathways flow along a sequence of episodes which are connected temporarily or causally. Parallel pathways shift between various levels of abstraction and make use of temporal parallelisms within or across life domains. Each one of those types of pathways can be utilized in a specific manner in order to stimulate the retrieval of autobiographic memory.

3.1 Hierarchical Pathways to Stimulate Top-Down Retrieval

Conway's model (1996) of autobiographic memory considers that this kind of memory is hierarchically organized in packages, which contain thematic and temporal information. Episodes are distinguished according to their duration and/or relationship with the respondent in 'lifetime periods' (long-term extended episodes), 'themes' (specific domains), and 'general episodes' (Conway 1996).[2] This means that the life course of an individual is grouped in several domains, which can in turn be organized in several general episodes. This allows top-down retrieval and thus the recalling or dating of specific events in the context of more general events. For example the lifetime period 'working life' can be organized in several 'themes' such

[1] For more detailed insights of cognitive processes in self-reports we refer to Tourangeau (2000) for remembering events and Bradburn (2000) for dating events. For an overview of recall problems in retrospective surveys we refer to Reimer and Matthes (2006).

[2] A summary of Conway's model is given in (Tourangeau/Rips/Rasiski 2000: 75 f.)

as training, employment or unemployment episodes. Once this general context is set, detailed information about 'general episodes', e.g. periods with the same employee can be easier retrieved. Higher order structures can be used to stimulate the memory of more specific information (Barsalou 1988).

3.2 Separate, Chronological Time-Lines to Stimulate Parallel Recall

Human memory is stored on an extended-episode time-line and episode types (Barsalou 1988). This means that episodes are organized in separate hierarchically and chronologically structured partonomies for various life domains. Furthermore, Barsalou (1988) proposes that different time-lines e.g. a time-line for school, work or relationships exist and that specific episodes that are associated with these domains are nested within each of those time-lines. For example, memory related to vocational training episodes is nested in the time-line linked to the job where the training took place. In turn, memory associated to this specific job is nested in general memory on work episodes. Furthermore, memory on these time-lines is organized chronologically meaning that contiguous episodes are closely stored together.

3.3 Sequential Pathways to Stimulate Dating of Episodes

The theoretical notions of Barsalou (1988) indicate that memory, which is related to episodes, is incorporated in generic as well as specific episode-related knowledge. Individuals for example know that after finishing university education one usually starts working. This is part of the generic knowledge about the educational system. Where and when one started working is stored in the specific episode-related memory. Essentially, this means that episodes are usually not time-tagged (Wagenaar 1986). Cognitive psychology calls this type of memory structure calendar representation (Huttenlocher/Hedges/Bradburn 1990). Hence, a specific episode is not directly connected with a date, because the memory about this episode and its dating are not stored together. Only very important or salient events are time-tagged (Glasner/van der Vaart 2008). Therefore, the dating of episodes has to be reconstructed by the respondent, often instantaneously during the interview process and thus can be regarded as especially error prone. In particular memory related to the episode's year, except for events taking place every year like anniversaries or birthdays, is easily confused (Glasner/van der Vaart 2008).

3.4 Autobiographic Memory Retrieval Problems

Parallel, sequential, and hierarchical pathways hardly ever give a true image of reality (Neisser 1986), but they often make "sense of the past in the light of the present" (Reimer/Matthes 2006: 714). This means that individuals idealize or adjust life histories in order to meet the typical or desired order of events; this can result in omitting or distorting episodes (Conway/Pleydell-Pearce 2000). Research calls this "smoothing over" one's life-course (Reimer/Matthes 2006: 718). For these reasons, we suggest that individuals may leave out short employment episodes, sometimes do not report minor changes like promotions or omit distressing or unpleasant time periods. Because we assume that this strategy is used by all individuals in a similar way,

hierarchical, parallel, and sequential pathways can nevertheless be used to stimulate human recall without biasing the results.

4 Stimulation of Autobiographical Memory: Combining Modularized Self-Reports and Event History Calendars

Reflecting these findings, in an effective interview human memory should be allowed to proceed as naturally as possible by making use of its structure. Thus, interviewing techniques should contextualize recall and encourage the usage of several retrieval strategies (cf. Reimer/Matthes 2006). The approach which was implemented in the IAB-ALWA study combines modularized self-reports and event history calendars in order to optimize the use of several retrieval strategies but differently emphasize them at certain stages of the interview. Another aim was to keep a high level of standardization which is necessary in a large-scale survey. Modularized self-reports are used to collect initial data, while a simplified event history calendar is used for revision of temporal inconsistencies.

4.1 Modularized Self-Reports as Tool for Data Collection

Barsalous (1988) suggests not to reconstruct the entire life history by using one single time-line, but to assess different types of episodes on various extended-episode time-lines. Thus, we organize an individual's life course history in several thematic domains, such as schooling, employment or partnership. This is what we call *modularization. Modularization* reflects the finding that personal events are closely linked to the context in which they take place. Therefore, they can be split up in chapters representing those contexts (Tourangeau/Rips/Rasiski 2000: 70). The idea has already been used in several waves of the German Life History Study (GLHS) (cf. Hillmert et al. 2004). Within each module, the interviewer assesses episodes that have been spent in a certain state (e.g. attending the same school type) in chronological order thus, taking advantage of the fact that episodes of the same type are often contiguous. So, *sequencing* as a chronological order strategy is used additionally to collect information. We begin with the earliest episode of a particular type and proceed in chronological order. This approach is in line with experimental research that showed that respondents tend to report long foregone episodes first (Glasner/van der Vaart 2008). It also reflects experiences from other large-scale European surveys that use a similar design, as for example the English Longitudinal Study of Ageing (ELSA) (cf. Hacker et al. 2007) or the Survey of Health, Ageing and Retirement in Europe (SHARELIFE) (cf. Schröder/Börsch-Supan 2008). However, those studies standardize their instruments only to some extent and often entrust the respondent with the decision which specific episodes to remember.

In the IAB-ALWA study we gather detailed information about specific events within extended-episode time-lines. The recall process is embedded in a hierarchical structure with a thematic domain as superordinate extended episode type. In this way, we attempt to collect temporal information on the dating of episodes on the one hand, and information that specifies the event on the other hand in a standardized

framework. Essentially, this means that every respondent moves through his life several times during the interview. By giving the possibility to report more than one type per time unit, this approach avoids that parallel or overlapping sequences are left out. Thus, it should prevent respondents from "smoothing over" their life-courses. Moreover, it accounts for the fact that individuals can be involved in two types of events in different domains at the same time, e.g. they can study and work part-time simultaneously. Hence, *parallel retrieval* is encouraged by this design.

Within some of the more complex life domains such as employment or vocational education, *top-down retrieval* (cf. Conway 1996) is additionally stimulated. By starting with the most frequent and salient episodes, e.g. main employment episodes, an alternative question wording within the modules stimulates less frequent or less salient episodes, e.g. secondary jobs or marginal employment. Thereafter, the interviewer asks for more specific episode types such as internships, traineeships, vicariates, job creation schemes and other fixed-term types of employment. In principle, however, it is left to the respondent in which sub-module he she reports the episode as long as he remains in the same domain. Each of these sub-module leads to a type-specific sequence of questions that is, however, very similar for all sequences. Additional information given in previous parts of the interview, such as answers naming the job title of the last employment, is used to stimulate the respondent's memory. This strategy is called *dependent interviewing* (Jäckle/Lynn 2004).

However, there are several problems linked to using modularized self-reports on extended-episode time-lines. First, memory outside a specific domain is not stimulated, which limits the efficacy of *parallel recall processes* described above. More specifically, this indicates that the sequential retrieval of episodes, which are not on the same extended-episode time-line, is restricted or even inhibited. Only *sequential* or *parallel pathways* within one domain are made accessible. Hence, completeness and consistency across domains cannot be guaranteed by *modularized reports*, because the overarching biographical context is blanked out. Time periods either filled with no information (gaps) or with inconsistent information existing across domains will not be visible at first sight, neither to the interviewer nor to the respondent. To overcome this problem, we included the possibility for *parallel retrieval* by giving the interviewers the possibility to re-access answers that were given before. However, *parallel retrieval* is difficult to standardize, because the underlying processes are rather individual. Therefore, we are not sure whether this strategy was successful and will produce the desired effect. Second, since episodes are usually not time-tagged, the dating of episodes has to be reconstructed by the interviewer often instantaneously during the interview process. The dating of episodes is therefore regarded as especially error prone. To overcome these problems, an event history calendar is used, serving not only as a recall aid, but also as a second stage in the data collection process.

4.2 Event History Calendar as Recall Aid

Traditionally, event history calendars assist to collect information about episodes and their dating on a grid with the horizontal axis as the time-line and the vertical axis split up in several life domains (Freedman et al. 1988). An event history calendar contains two features: It covers limited but specific and important life domains and utilizes a flexible interviewing style to stimulate memory retrieval. Belli (1998) first linked insights from cognitive psychology with the rationale of constructing such event history calendars (EHC) and use the for data collection. He considers it a powerful tool to collect life histories, because it constitutes a reference point for interviewer and respondent in order to both cover important life domains and connect them with each other. However, little attention has been paid to confirming this notion empirically. Van der Vaart (2004) found that event history calendars are an effective tool for the collection of long foregone, unimportant, and less frequent episodes. However, the usage of event history calendars in large-scale surveys that are administered with computer-assisted telephone interviews is associated with several shortcomings. One is that they hardly can be used as standardized instruments in telephone surveys, because the event history calendar is not visible for the respondent. Thus, implementing an event history calendar also requires experienced and skilled interviewers.

Due to these shortcomings, event history calendars are not utilized in the first stage of the data collection tool implemented in our study. However, in line with suggestions of previous research, they are used as a recall aid and a data revision module in the second stage of data collection (cf. Glasner/van der Vaart 2009). All episodes that are collected throughout the interview are merged and displayed together on a list abandoning the structure of several separate time-lines. This idea has been developed and tested by Matthes/Reimer/Künster (2005). Memory cues in the form of specific survey questions and contextual information become available to stimulate the respondents' retrospective memory. However, inconsistencies are only resolved when they are visible to the interviewer (Freedman et al. 1988). Therefore, algorithms are programmed to detect whether there are any temporal gaps, overlaps or parallel episodes that are not considered possible by the survey designers. In particular, gaps can occur because temporal information was only collected for a limited number of life domains. Also, previously hidden gaps and inconsistencies that exist across domains become apparent in the data revision module. These gaps and inconsistencies can be viewed as a signal of potential recall problems and are communicated and resolved in collaboration with the respondent. Hence, *parallel recall* is stimulated in the event history calendar where all chronological information is used simultaneously.

As a next step, data on these gaps and inconsistencies is collected by giving information about temporal connections (i.e. episodes directly located before or after the inconsistency) to the respondent. In order close these gaps, the questionnaire leads back to the modularized self-reports and the adequate episode. Data on other gap episodes that could not be previously reported is directly collected in the data revi-

sion module, e.g. information on longer periods of holidays or times as homemaker without being on parental leave. For overlapping or parallel episodes the inconsistency can either be confirmed by the respondents or the start and/or end dates of episodes can be adjusted until they adjoin each other. This accounts for the finding that episodes are often not time-tagged, and both start and end dates are remembered better in conjunction with other episodes.

The chronological order of imputing gaps and resolving inconsistencies is not left to the interviewer. Rather, the problem solving procedure is embedded in a pre-defined routine diminishing the interviewers' cognitive effort and equalizing differences in interview skills. To ensure standardized reporting of gaps and inconsistencies, scripted questions, prescribing exact question wordings, guide interviewer and respondent through the interview process (cf. Reimer/Matthes 2006). A narrative style of interviewing is not possible in such a large scale survey design but the scripted questions were pre-formulated, to sound as naturally as possible. Moreover*, dependent interviewing* strategies, which adapt the wording of the questions to previous answers of the respondents, were implemented. Interviewers were encouraged to give additional information when something was unclear and to use probing strategies, such as *landmark events* to stimulate memory. However, we only included personal landmarks (e.g. moving dates) in our tool because public landmarks (e.g. 9/11) were found to be of minor importance (Hacker et al. 2007; Van der Vaart/Glasner 2008). The data revision module is recursive, which means that all changes except for open text comments are implemented instantaneously. This method can result in new gaps and overlaps that have to be resolved. We expect, however, that this procedure does not lead to more gaps or inconsistencies.[3]

In sum, the hybrid modularized self-report/event history calendar design adopts advantages of insights from cognitive psychology, because it uses available idiosyncratic structures in the autobiographical memory of a respondent. The given chronological order and parallel retrieval is not used in the first stage but emphasised in the data revision module in the second stage. In the next paragraph we will assess empirically who benefits from this module and investigate whether the data revision module really improves completeness and dating accuracy in our study.

5 Data, Measurement, and Empirical Methods

The data revision module is implemented in the IAB-ALWA study which collected life history data of 10,177 respondents, belonging to birth cohorts from 1956 to 1988. The data collection was accomplished with CATI techniques and is based on methods developed for the German Life History Study (GLHS) which was conducted at the Max Planck-Institute for Human Development in Berlin in the 1990s and early 2000s (cf. Kleinert/Matthes/Jacob 2008). Not including gaps 98,149 episodes were collected in the following life domains: schooling, vocational education, employment,

[3] The technical implementation of the data revision module is described in Appendix 2.

unemployment, military service (including civil service and voluntary social service) and parental leave. In addition, the residential history was assessed. Other episode types could be added in the data revision module as different types of gap episodes. The number of gap episodes amounted to 9,339. Those episodes are for example longer periods of holidays, times as homemaker, sickness leave or other unspecified time periods.

Insights from cognitive psychology and experimental research on recall (cf. Glasner/van der Vaart 2008; Huttenlocher/Hedges/Bradburn 1990; Matthes/Reimer/Künster 2007) reveal that autobiographical memory works differently with respect to the collection of information about episodes and their dating. Therefore, we base our empirical analyses on two dependent variables indicating a change made in the data revision module. Doing so, we claim to interpret the differences between data before and after the data revision module as corrected errors. In our data it is possible to differentiate between added episodes that refer to completeness and changed episodes that refer to dating accuracy. Thus, the first dependent variable measures added episodes that have been omitted in the modularized questionnaire but were reported in the data revision module. It was possible to add episode types that were forgotten in the modularized questionnaire. Those amount to 3,996 episodes. This also means that we do not take the gap episodes into consideration because they could not be reported in the modularized questionnaire before. In sum, we arrive at a 12.4 per cent increase in episodes (3.7 per cent added episodes; 8.7 per cent gap episodes) compared to the data before the data revision module, indicating an overall improvement of completeness.

The second dependent variable refers to dating corrections in episodes. It was possible to modify all episode types with respect to begin and end of the episode (month and year reported). We applied a restrictive measure of changed episodes and excluded all date corrections which were implemented in the programming routine of the data revision module, e.g. seasonal date corrections. Seasonal date corrections are estimates that were used when respondents could not remember the exact month in which the event took place, but were able to remember the season. The reason for exclusion is that the distinction between date corrections induced by the respondent and those generated by the program itself is not always clear cut. In sum, 4.4 per cent of the episodes were corrected. More precisely, in 1.9 per cent of the episodes the starting month was corrected, in 2.6 per cent the final month, in 1.7 per cent the starting year and in 1.9 per cent the final year. Often more than one component of the date was corrected.[4] The structure of these episodes can be obtained from Table 1.

[4] Results are provided upon request.

Table 1
Summary statistics first and second stage of data collection by episode type

	First stage: episodes collected (N)	Second stage: episodes added (N)	increase in episodes (in per cent)	Second stage: episodes corrected (N)
schooling	24,270	330	1.35	827
vocational training	19,302	529	2.74	1,076
employment	33,667	1,949	5.79	1,926
unemployment	9,410	1,128	11.99	529
military service	3,509	60	1.71	212
parental leave	4,967			111
vocational preparation schemes	3,024			85
gap episode		9,339		131
Total excluding gaps		**3,996**		**4,766**
Total including gaps		**13,335**		

Source: ALWA, own calculations. In total 98,149 (excluding gaps) or 107,488 (including gaps) episodes were collected

Independent variables are individual as well as event specific characteristics. On the individual level we include a dichotomous variable indicating the gender of the respondent, a variable denoting the age, centered on the mean, and the highest educational degree of the respondent at the moment the interview took place. Education was coded as a series of dummy variables representing different levels of the ISCED-97 classification (International Standard Classification of Education 1997) (cf. OECD 1999). To measure event specific characteristics, we use a further set of dummy variables indicating the event type. We distinguish between schooling, vocational training, employment and unemployment. In addition, we included the duration of the episode in months, centered on the mean, the time-lag between the episode's end and the interview date in months, standardized by the mean age of the respondents, and the total number of episodes a respondent reported before he or she entered the data revision module. Summary statistics of those measures can be obtained from Table A in the appendix.

Descriptive results reveal that there are a number of factors, influencing the likelihood of adding or correcting episodes.[5] To account for this, we conducted multivariate analyses, using both the likelihood to add and the likelihood to correct an episode as dependent variables. Thus, the dependent variables are dichotomous variables with a value of 1 indicating a change induced by the data revision module and a value of 0 indicating no change with respect to the data revision module.

The dependent variables are examined in separate logistic regressions which take the nested structure of the date into account. This means that one respondent can

[5] Figures are provided upon request.

deliver several episodes. We calculate robust standard errors to account for these clusters (cf. Huber 1964). In our models only around 4,000 events for the logistic regression on the likelihood to add an event and around 5,000 events for the likelihood to correct an event can be analysed. We thus consider the rare event structure of the data with a small number of events but a multitude of non-events (almost 100,000). As suggested by King and Zeng (2001), we calculate logit models in rare events data. They make use of the statistical finding that in the case of a rare event structure, the event itself is more statistically informative than a non-event. Technically this means that consistent ML-(Maximum likelihood) estimators are calculated that give a lower mean square error for a coefficient in the presence of rare event data.[6] Since previous research indicated that gender differences are especially pronounced in autobiographic memory (cf. Auriat 1991; Matthes/Reimer/Künster 2007; Paull 2002), we decided to include separate models for men and women. A chow test also confirmed that the coefficients jointly vary between men and women.[7] Furthermore, we examine separate models for different types of episodes, because it can be assumed that they are remembered and retrieved not in the same manner. The following paragraph presents several descriptive as well as multivariate results on the completeness and dating accuracy in the IAB-ALWA study.

6 Completeness and Dating Accuracy in ALWA

6.1 Completeness: Added Episodes

6.1.1 Descriptive Results

In this paragraph we examine which episodes have a high likelihood to be omitted in the modularized questionnaire and to be added in the data revision module.[8] Furthermore, we focus on the most important episode types and exclude parental leave and vocational preparations schemes from the analyses. We also focus on differences between episode types because research found that there are variations in recall depending on the type of episode (Reimer/Matthes 2006).

Depending on the type, Figure 1 depicts the absolute probability of an episode to be added. This is in line with previous research results (Matthes/Reimer/Künster 2007). Almost half of the episodes are employment episodes, but also unemployment episodes are added frequently. In total, over two third of the episodes are either employment or unemployment episodes. The other episodes, like school, vocational training and military service are less likely to be forgotten. This result is not surpris-

---

6 Details can be found in Tomz/King/Zeng (2003), King/Zeng (1999a), and King/Zeng (1999b). The ReLogit Stata ado (cf. Tomz/King/Zeng 1999 [earlier version]) file can be downloaded from http://www.jstatsoft.org/v08/i02. Note that model fit statistics, e.g. Pseudo-R^2 cannot be calculated. Regression diagnostics and model fit statistics to check for specification errors and to get an impression of model quality were calculated using standard logistic regression models.

7 Results are provided upon request.

8 In the figures the data revision module is called x-module due to limited space availability.

ing, because unemployment and employment episodes are expected to occur more often than vocational training, schooling or military service.

Figure 1
Added Episodes by Type

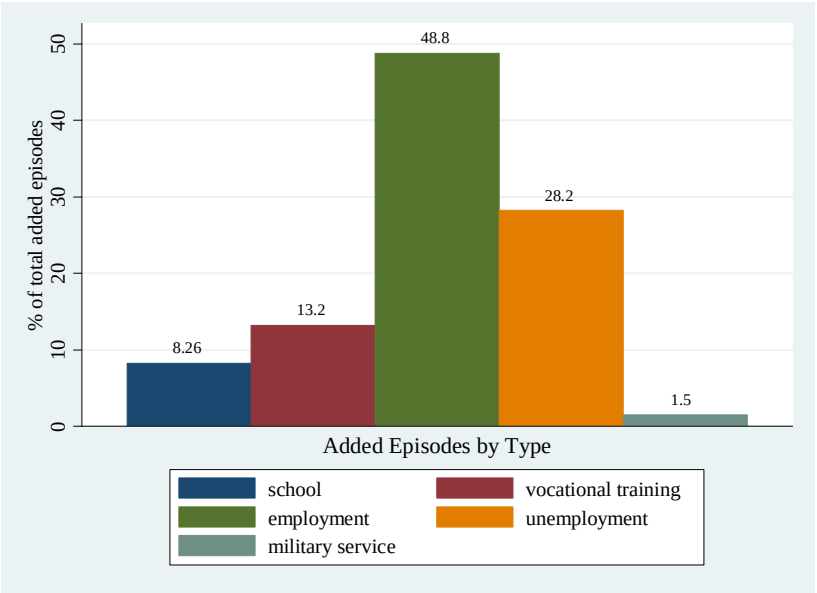
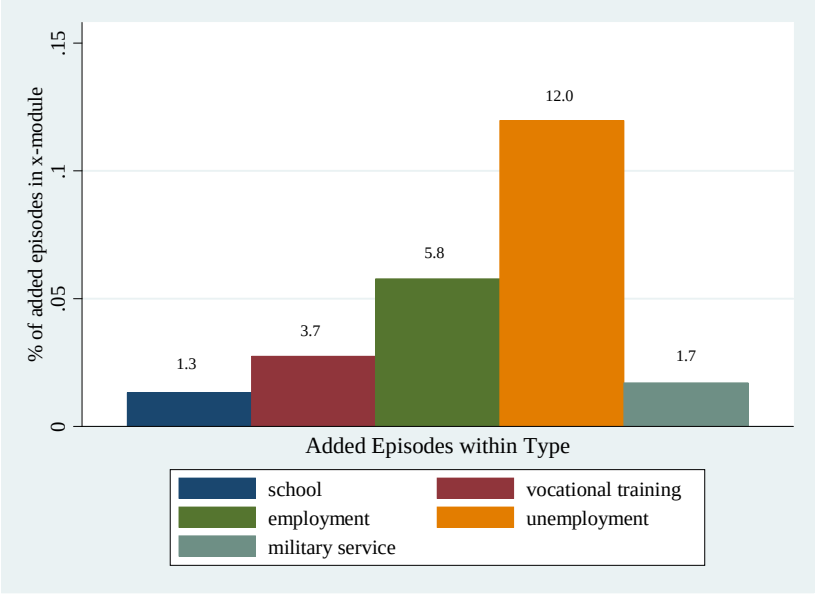


Figure 2
Added Episodes within Type



It is even more interesting, to change the focus to the analysis of added episodes within the episode type. Figure 2 reflects the relative probability of an episode to be added. It becomes clear that unemployment episodes have by far the largest relative probability to be omitted in the modularized questionnaire and to be added later on. In total, 12 per cent of the unemployment episodes were added in the data revision module. This indicates that unemployment episodes are especially error prone. For employment episodes, the relative probability to make additions is with around

six percent also quite large, although it amounts only to half of the likelihood of unemployment episodes. For the other types of episodes, the relative probabilities are a lot smaller and range between 1.3 and 2.7 per cent. This indicates that the data revision module seems to be more important for episode types that occur more often, are less bound to institutions, or are less salient to the respondent. Revising data seems to have the highest importance for the collection of unemployment histories, because they are considerably more complete now than they would be without making use of the data revision module.

6.1.2 Influencing Factors on Adding Episodes

To account for the fact that different influence factors work simultaneously, we conducted multivariate analyses on the likelihood of adding an episode.[9] First, we estimate a model including both men and women. Women have a 19 per cent lower likelihood compared to men to make an addition. As can be seen from Table B (in the appendix), women are less likely to make additions. Age also has an influence on episodes. Every additional year above the mean age of 39, increases the likelihood of adding an episode. Obviously, for individuals at the end of their 30s it becomes increasingly challenging to remember their complete life history. This effect is again highly significant. In contrast, the level of education does not matter. There is hardly any evidence that less educated people are more likely to omit episodes except for low educated men.

Turning to the characteristics of the type of episode, we find that, compared to employment episodes, all other types of episodes, except for those reporting on unemployment, were added with a significantly lower likelihood. This finding means that employment as well as unemployment episodes are especially error prone and confirms our descriptive results. Furthermore, characteristics of the episode itself matter for the probability of omitting the episode first and adding it later on. Each additional month of duration, starting from the mean duration of around three years, decreases the likelihood to add an episode later on. This indicates that longer episodes are less likely to be omitted. The likelihood to add an episode decreases with every additional observation above the average number of episodes (around 10 episodes). The time-lag that exists between the start date of the episode and the interview date, beginning with the average start date of around 14 years before the interview took place, has a positive impact on adding episodes. Thus, the more time elapsed between the episode and the interview date, the more likely it is that the episode will be omitted first and added in the data revision module. In sum, all variables relating to the characteristics of the episode are highly significant.

The separate models for men and women reveal similar effects as the model including both men and women. Only small differences caused by the level of education and military service episodes are reflected. Again, a positive age effect for men and

[9] Corresponding univariate analyses can be provided upon request.

women and almost no effect of the level of education is found. The influence of the episode type also resembles that of the full model. Once more, a negative duration effect and a negative effect of the number of episodes are reported. One striking difference, however, has to be mentioned: the time-lag between the interview date and the end of the episode has a positive and highly significant effect for men but not for women. For women it does not matter how much time has elapsed since the event started but for men this time-lag significantly matters.

We also modelled each of the episode types separately (Table C in the appendix). Concerning individual characteristics, a negative effect for females can only be found for employment episodes. Thus, gender does not equally affect omitting or adding an episode for all types of episodes. A positive age effect can only be obtained in the models for schooling, employment and unemployment episodes. An impact of the level of education is noted for vocational training episodes, and to a lesser extent in unemployment histories. For vocational training episodes less educated individuals have a significantly lower likelihood to omit and add an episode in the data revision module. One possible interpretation is that less educated persons have less complex educational histories to remember. For military service, no significant impact can be found for the variables included in the model. The effects related to characteristics of the event itself are quite similar for the duration and the number of episodes previously reported. However, the negative effect of the time-lag between the end of an episode and the interview date, in the case of schooling events is reversed compared to the full model. This indicates that this type of episode is remembered differently.

6.2 Dating Accuracy: Corrected Episodes

6.2.1 Descriptive Results

In this paragraph we examine the likelihood of episodes to be revised in the data revision module in terms of dating corrections. Doing this, additional episodes are included in the models, because all episodes could be changed with respect to the starting and/or final date (month and year). Thus, corrected episodes serve as a proxy for dating accuracy. For the results we decided to focus again on the most important episode types. In some respect, the results resemble those presented earlier in the results section for added episodes. First, the absolute probability of an episode to be corrected is reported. This probability is displayed in Figure 3.

42 per cent of the corrected episodes are employment episodes, which is by far the greatest number across all episode types. However, school and vocational training episodes are also corrected frequently. Bearing in mind that these episode are usually concentrated in the early life course of an individual, this result is not surprising. Fewer corrections occur in unemployment episodes and also in the dating of military or civil service.

Figure 3
Corrected Episodes by Type

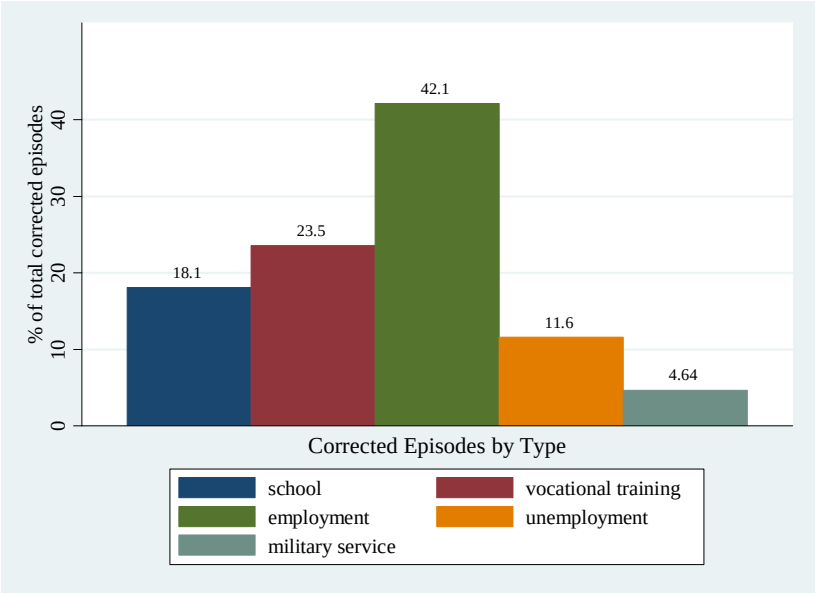


Figure 4
Corrected Episodes within Type

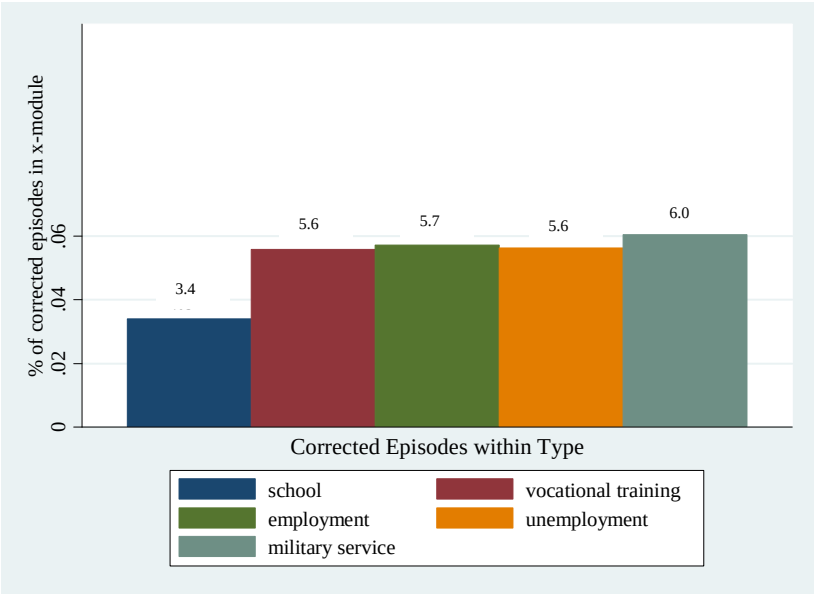


Figure 4 displays the relative probability of an episode to be corrected. When looking at corrected episodes within the type, it becomes quite clear that corrections in the dating are obviously not bound to the episode type. The percentage of corrections is quite similar across types. Around 6 per cent of the episodes of one type are corrected. The only exception is schooling, where the percentage appears to be lower compared to the other types, because only episodes ending after the respondent's 14th birthday were included in the data revision procedure. However, these findings already indicate that adding and correcting episodes are driven by different mechanisms.

6.2.2 Influencing Factors on Correcting Episodes

For corrected measures a different picture emerges than it did for added episodes (Table D in the appendix): Out of the individual characteristics, only for gender a significant effect is reported, which is, again negative. Female respondents are less likely to make a dating correction. In contrast to the results of adding an episode, neither an age nor an education effect is found. Individual characteristics are obviously not relevant for correcting episodes but only for adding episodes. Turning to the effects of the characteristics of the episode, vocational training, unemployment, and military service episodes are more likely to be corrected with respect to the dating compared to employment episodes. This result contradicts the one of added episodes. However, compared to employment episodes, those of schooling are less likely to be corrected. This is not surprising, because schooling episodes are only checked from the 14th birthday of a respondent. Furthermore, they are located in the early life course of an individual and are non-recurrent unique events. In contrast, employment episodes take place over a longer time span in the life course of an individual. There is a highly significant, small, and positive effect of the episode's duration and a positive effect of the time-lag. The last result is not surprising: the more time has elapsed since the episode took place, the more likely it will be corrected. The first result is somewhat surprising. The longer an episode, starting from the average duration of around three years, the more likely it is that it will be corrected. An explanation could be that for long episodes it becomes more difficult to determine the exact dating with respect to month and year.

Examining men and women separately, individual characteristics do not matter at all. There is no significant effect of age or level of education. Episode characteristics, however, matter for both men and women. Vocational training episodes have a higher likelihood to be corrected only by women, while military service, including civil service and voluntary social service episodes, is more likely to be corrected by men only. The last result is not surprising because most of the military service episodes in Germany are of men.[10] The positive duration effect is not gender-specific, a finding which contradicts the model for added episodes with respect to the direction and no gender difference. There is no effect of the number of episodes reported in none of the models reflecting that data corrections are not bound to the frequency of reported events.

Assessing the influence of different characteristics for different types of episodes separately, one receives ambiguous findings concerning individual characteristics (Table E in the appendix). No gender effect can be obtained for corrections. The age effect is obviously different for different types of episodes, leaving a positive age effect of schooling episodes and a negative effect for vocational training, employ-

[10] Germany installed a compulsory military service for men but not for women. However, it is possible for women to make a social, ecological or European year as a voluntary service. These episodes are also included in the military service episodes because they, and the civilian service, can also serve as a replacement for compulsory military service.

ment and unemployment episodes. A highly significant positive effect of holding a lower secondary degree or having reached the first stage of tertiary education in the academic field can only be found for employment episodes. With regard to the duration, the results are quite similar for almost all types of episodes with a positive effect, except for military service episodes, where a negative effect can be found. Except for unemployment, no effect of the number of episodes that are reported can be found. Again, the positive effect of the time-lag between episode and interview date for all other of episodes except for military service episodes is reversed and negative for schooling episodes.

6.3 Discussion of the Major Findings

First, we analyzed which episodes have a greater likelihood to be forgotten in the modularized collection of data, but can be retrieved in a different setting in the data revision module. In sum, for the probability of adding an episode, both individual (gender, age and education) and event specific characteristics (type, number of episodes, time-lag and duration) matter. However, the analyses showed that event specific characteristics are more important. Overall, the variables which were taken into consideration in the models seem highly relevant. Especially pronounced are gender differences, indicating that women tend to forget fewer episodes. The results can be interpreted in two ways: First, women initially tend to forget fewer episodes including the long forgone ones. This means that they have less to add later on. Second, it is possible that women benefit less from the data revision module. We are confident that the first option is more likely because it would resemble previous research on gender differences (cf. Auriat 1991).

Separate models for men and women reveal that the time-lag between interview date and end of the episode matters only for men but not for women. For women in contrast to men it does not matter how much time has elapsed since the event took place in order to remember the event. This is in line with previous experimental research of several authors (Auriat 1991; Paull 2002; Reimer/Matthes 2006). Separate models for different types of episodes reveal that especially individual characteristics have diverse influences on different event settings. These findings can at least partly be substantiated by the fact that different types of episodes, like schooling episodes, occur earlier in the life-course of an individual. Furthermore, different people could inconsistently be affected by those events, e.g. military service is in most cases taken by men because it is compulsory for them. For schooling episodes we found a result which is rather surprising: the longer the time-lag to the interview date, the less likely it is that the episode will be remembered not until the data revision module. Bearing in mind that schooling episodes take place chronologically early in life and can also be viewed as unique events which are usually remembered well, it becomes clear that schooling episodes, no matter how much time has elapsed, are not likely to be forgotten.

Second, we examined which episode types are more error prone in terms of dating accuracy. A different picture is drawn here than for adding an episode. Event spe-

cific characteristics are again more important than individual characteristics. It is not surprising that the number of observations does not matter, because our model refers to dating accuracy and this should intuitively have nothing to do with the number of episodes previously remembered. Furthermore, it indicates that episodes are indeed not time-tagged (Glasner/van der Vaart 2008). When assessing separate models for different types of episodes it becomes clear that especially individual characteristics age are of varying importance for different type of episodes. The dating of schooling episodes is more likely to be corrected with increasing age, which can again be explained by the peculiarities of schooling episodes – their uniqueness and their chronological positioning especially for people in their 40ties. Vocational training and employment episodes are in turn less likely to be corrected by older people. In sum, for dating accuracy it can be assumed that there are more relevant characteristics besides those examined in this paper. However, other memory features are probably important s but those are generally difficult to trace.

7 Conclusion

An important contribution of this study was to conduct multivariate analyses on the effectiveness of the data revision module. This strategy seems absolutely necessary when bearing in mind that omitting episodes or dating them wrongly is, as suggested by cognitive psychology, a multifaceted concept. Furthermore, this paper brings together theoretical notions from cognitive psychology with the design of standardized, large-scale surveys. We followed several guidelines how to stimulate human recall with an appropriate survey design given by for example Groves et al. (2000: 215) or Tourangeau (2000: 40). Among other implications, they suggest to use life event calendars, multiple cues, longer and decomposition techniques. Overall, the data revision module that was incorporated as a recall aid in the IAB-ALWA survey improves completeness of the data and dating accuracy. Individual and event specific characteristics matter for adding an episode. This picture is different for correcting an episode. Here, only event specific characteristics are crucial.

A special emphasis has to be put on the results regarding the recall of one particular type of episode: unemployment episodes are obviously especially error prone when assessing them in a standardized interview setting, such as modularized self-reports. This is in line with previous experimental research results of Paull (2002, 1997), Belli et al. (2007), and Dex and McCulloch (1998). In sum, we conclude that unemployment episodes have been probably underestimated in previous life-course studies. From a psychological point of view, unemployment carries a “social stigma” (Reimer 2004) and stable periods of unemployment are rarely found, because unemployed people are usually constantly looking for work. In our view, unemployment periods can indeed be seen as non-events (cf. Elias 1997) where recall is only to a minor extent bound to pathways that can be stimulated because nothing special happens or special characteristics cannot be remembered. The data revision module, however, stimulates the recall of unemployment episodes and is therefore especially useful for this type of episode.

Regarding the practicability of the data revision module, we can confirm that including such a module in a large scale survey design seems feasible. It obviously facilitates the respondents' memory retrieval. Of course, the freedom of interviewers has to be limited in favour of a standardized interview, which is in sharp contrast to the original idea of the event history calendar method (Belli 1998; Freedman et al. 1988). One advantage, however, is that interviewers can be trained within a short period of time to handle the tool, because interview training is not as intensive as for using standard EHC (cf. Glasner/van der Vaart 2007). Once interviewers are familiar with retrospective interviewing techniques and probing strategies to facilitate retrieval, the data revision module is easy to handle. Furthermore, it levels out that interviewer skills are unevenly distributed and makes data as complete and consistent as possible, before they are edited without the assistance of the respondent.

The data revision module is, however, only an intermediate step. Further editing is still necessary because one problem of human memory, the misclassification of events is not touched at all by the data revision module (cf. Reimer 2004). These kinds of errors occur because respondents report episodes in the wrong module and it becomes necessary to shift episodes from one module to another while preserving as much information as possible. Theoretically, it would be possible to include this step already in the data revision module, but we think that these errors are hard to detect during the interview process and doing so would overstrain the interviewer as well as the respondent. Thus, we decided that this procedure will take place after the interview with a standardized editing process. Nonetheless, we are confident that the data revision module reduces the editing work connected with the survey, and therefore will reduce total survey costs. However, the cost side cannot be answered unequivocally: on the one hand the data revision probably reduces editing; on the other hand costly interviewing time is extended. The most important advantage of including a data revision module directly in the interview process is in our view is that unclear and inconsistent information is either verified or corrected by the respondents themselves. Therefore, data collected with a data revision module are in our view more reliable.

A shortcoming of this study is that it only shows the improvement of the data revision module compared to the modularized self-reports with respect to remembering and dating episodes. However, no gold standard is available; the data quality still refers to relative data quality. Whether the episode list is really complete and correctly dated will remain unclear, because we have no information about reality and we can only make use of available information. Needless to say that it cannot be expected that all episodes or dates are generally retrievable from autobiographic memory. Hence, absolute data quality cannot be reached. However, we are confident that most errors appear as inconsistencies at some point in the interview process. For unemployment and employment episodes another way to assess the quality of retrospective life course data exists: they can be compared to process generated data provided by the German Federal Employment Agency. Of course these data can neither be considered perfect nor seen as a benchmark reflecting reality.

This step, however, is not yet done; but once completed, it could convince opponents of life-course data of the adequateness of this way of data collection.

This paper is only a modest step towards incorporating insights from cognitive psychology into survey methodology. There are still a number of questions unanswered. Thus, future research can be expanded in several directions. First, a comparison of the modularized interviews with standardized conventional questionnaire instruments as well as the more flexible event history technique would be desirable. As the advantages and disadvantages of these methods become visible, it would allow future survey designers to choose the optimal strategy for their purposes. Second, because there is insufficient empirical research, the sequence of modules is chosen more or less arbitraryily. This is also true for the (reverse) chronological order. So far, using a reverse chronological order gave ambiguous results (Tourangeau 2000: 40). Shedding light one those issues could again enhance data quality. Third, the peculiarities of unemployment episodes discussed above need further attention when designing new surveys. Finally, it would be interesting to assess how the flexible tool to display additional information is used in the modularized interview to stimulate recall. Thereby it would be possible to get deeper insights of the functioning of such a data revision module and further improve it.

References

Auriat, Nadia (1991): Who forgets? An analysis of memory effects in a retrospective survey on migration history. In: European Journal of Population, vol. 7, no. 4, pp. 311–342.

Barclay, Craig (1986): Schematization of autobiographical memory. In: Rubin, David C. (ed.), pp. 82–99.

Barsalou, Lawrence (1988): The content and organization of autobiographical memories. In: Neisser, Ulric; Winograd, Eugene (eds.): Remembering reconsidered: Ecological and traditional approaches to the study of memory. New York: Cambridge University Press, pp. 193–243.

Belli, Robert F. (1998): The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in surveys. In: Memory, vol. 6, no. 4, pp. 383–406.

Belli, Robert F.; Lee, Eun Ha; Stafford, Frank P.; Chou, Chia-Hung (2004): Calendar and question-list survey methods: Association between interviewer behaviors and data quality. In: Journal of Official Statistics, vol. 20, no. 2, pp. 185–218.

Belli, Robert F.; Shay, William L.; Stafford, Frank P. (2001): Event history calendars and question list surveys: A direct comparison of interviewing methods. In: Public Opinion Quarterly, vol. 65, no. 1, pp. 45–74.

Belli, Robert F.; Smith, Lynette M.; Andreski, Patricia M.; Agrawal, Sangeeta (2007): Methodological comparisons between CATI event history calendar and standardized conventional questionnaire instruments. In: Public Opinion Quarterly, vol. 71, no. 4, pp. 603–622.

Bradburn, Robert M. (2000): Temporal representation and event dating. In: Stone, Arthur A. et al., pp. 48–62.

Blossfeld, Hans-Peter; Golsch, Katrin; Rohwer, Götz (2007): Techniques of event history modeling using Stata: New approaches to causal analysis. Mahwah, NJ: Erlbaum.

Conway, Martin A. (1996): Autobiographical knowledge of autobiographical memory. In: Rubin, David C. (ed.) Remembering our past: Studies in autobiographical memory. Cambridge, MA: Cambridge University Press, pp. 67–93.

Conway, Martin A.; Pleydell-Pearce, Christopher W. (2000): The construction of autobiographical memories in the self–memory system. In: Psychological Review, vol. 107, no. 2, pp. 261–288.

De Graaf, Nan Dirk; Wegener, Bernd (1989): (Un)reliability of job career and social resource data. Berlin: Max Planck Institute for Human Developement and Education & University of Heidelberg. Unpublished manuscript.

Dex, Shirley (1991): The reliability of recall data: A literature review. In: ISER Working Paper 1991–11.

Dex, Shirley; McCulloch, Andrew (1998): The reliability of retrospective unemployment history data. In: Work, Employment & Society, vol. 12, pp. 497–509.

Elias, Peter (1997): Who forgot they were unemployed? In: ISER Working Paper 1997–19.

Freedman, Deborah; Thornton, Aland; Camburn, Donald; Alwin, Duane; Young-DeMarco, Linda (1988): The life history calendar: A technique for collecting retrospective data. In: Sociological Methodology, vol. 18, pp. 37–68.

Glasner, Tina; van der Vaart, Wander (2009): Applications of calendar instruments in social surveys: A review. In: Quality and Quantity. vol. 43, no. 3, pp. 333–349.

Glasner, Tina; van der Vaart, Wander (2008): Cognitive processes in event history calender interviews: A verbal report analysis. In: Proceedings of the Seventh International Conference on Survey Methodology, Napels, Italy, 1. – 5. September 2008.

Groves, Robert M.; Fowler, Floyd J. Jr.; Couper, Mick P.; Lepkowski, James M.; Singer, Eleanor; Tourangeau, Roger (2004): Survey Methodology. Hoboken, NJ; Wiley & Sons.

Hacker, Elizabeth; Cox Kate; Lessof, Carli; Miceli, Colin (2007): Retrospective data collection. In: National Centre for Social Research (NatCen). Unpublished manuscript.

Hillmert, Steffen (2002): Edition von Lebensverlaufsdaten: Zur Relevanz einer systematischen Einzelfallbearbeitung bei standardisierten Befragungen. In: ZUMA-Nachrichten no. 26, pp. 120–140.

Hillmert, Steffen; Künster, Ralf; Spengemann, Petra; Mayer, Karl-Ulrich (2004): Projekt Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland: Dokumentation. In: Materialien aus der Bildungsforschung no. 78. Berlin: Max-Planck-Institut für Bildungsforschung.

Huber, Peter J. (1964): Robust estimation of a local parameter. In: Annals of Mathematical Statistics, vol. 35, pp. 73–101.

Huttenlocher, Janellen; Hedges, Larry V.; Bradburn, Norman M. (1990): Reports of elapsed time: Bounding and rounding processes in estimation. In: Journal of Experimental Psychology: Learning, Memory & Cognition, vol. 34, no. 1, pp. 138–147.

Jäckle, Annette; Lynn, Peter (2004): Dependent interviewing and seam effects in work history data. In: ISER Working Papers Number 2004–24.

King, Gary; Zeng, Langche (1999a): Logistic regression in rare events data. Department of Government, Harvard University, http://GKing.Harvard.Edu.

King, Gary; Zeng, Langche (1999b): Estimating absolute, relative, and attributable risks in case-control studies. Department of Government, Harvard University, available from http://GKing.Harvard.Edu.

King, Gary; Zeng, Langche (2001): Logistic regression in rare events data. In: Political Analysis, vol. 9, no. 2, pp. 137–163.

Klein, Thomas; Fischer-Kerli, David (2000): Die Zuverlässigkeit retrospektiv erhobener Lebensverlaufsdaten: Analysen zur Partnerschaftsbiographie des Familiensurveys. In: Zeitschrift für Soziologie, vol. 29, no. 4, pp. 294–312.

Kleinert, Corinna; Matthes, Britta; Jacob, Marita (2008): Die Befragung "Arbeiten und Lernen im Wandel": theoretischer Hintergrund und Konzeption. In: IAB-Forschungsbericht 05/2008.

Klijzing, Erik; Prophet, Helen (1998): On the quality of FFS event history data. In: Globalife Working Paper Series no. 13.

Mathiowetz, Nancy; Duncan, Greg (1988): Out of work, out of mind: Response errors in retrospecitve reports of unemployment. In: Journal of Business & Economic Statistics, vol. 6, no. 2, pp. 221–229.

Matthes, Britta; Reimer, Maike; Künster, Ralf (2007): Techniken und Werkzeuge zur Unterstützung der Erinnerungsarbeit bei der computergestützten Erhebung retrospektiver Längsschnittdaten. In: Methoden–Daten–Analysen, vol. 1, no. 1, pp. 69–92.

Matthes, Britta; Reimer, Maike; Künster, Ralf (2005): TrueTales – Ein neues Instrument zur Erhebung von Längsschnittdaten. Arbeitsbericht 2 des Projektes Frühe Karrieren und Familiengründung: Lebensverläufe der Geburtskohorte 1971 in Ost- und Westdeutschland. Berlin: Max-Planck-Institut für Bildungsforschung.

Neisser, Ulric (1986): Nested structure in autobiographical memory. In: Rubin, David C. (ed.), pp. 71–82.

OECD (1999): Classifying educational programmes: Manual for ISCED–97. Implemenation in OECD countries 1999 edition.
URL: http://www.oecd.org/dataoecd/7/2/1962350.pdf.

Paull, Gillian (2002): Biases in the reporting of labour market dynamics. In: IFS Working Papers W02/10.

Paull, Gillian (1997): Dynamic labour market behaviour in the British Household Panel Survey: The effects of recall bias and panel attrition. In: Centre for Economic Performance & Institute of Economics Paper no. 10.

Peters, H. Elizabeth (1988): Retrospective versus panel data in analyzing lifecycle events. In: The Journal of Human Resources, vol. 23, no. 4, pp. 488–513.

Reimer, Maike (2004): Collecting event history data: About work careers retrospectively: Mistakes that occur and ways to prevent them. Arbeitspapier no. 1 des Projektes Kognition und Kommunikation bei der Längsschnittdatenerhebung. Berlin: Max-Planck-Institut für Bildungsforschung.

Reimer, Maike (2001): Die Zuverlässigkeit des autobiographischen Gedächtnisses und die Validität retrospektiv erhobener Lebensverlaufsdaten. Materialien aus der Bildungsforschung no. 71. Berlin: Max-Planck-Institut für Bildungsforschung.

Reimer, Maike; Matthes, Britta (2006): Collecting event histories with True Tales: Techniques to improve autobiographical recall problems in standardized interviews. In: Quality and Quantity, vol. 41, no. 5, pp. 711–735.

Rubin, David C. (ed.) Autobiographical memory. Cambridge, MA: Cambridge University Press.

Rubin, David C.; Baddeley, Alan D. (1989): Telescoping is not time compression: A model of the dating of autobiographical events. In: Memory and Cognition, vol. 17, no. 6, pp. 653–661.

Schröder, Mathis; Börsch-Supan, Axel (2008): Retrospective data collection in Europe. In: MEA Discussion Paper 172–08.

Skowronski, John J.; Thompson, Charles P. (1990): Reconstructing the dates of personal events: Gender differences in accuracy. In: Applied Cognitive Psychology, vol. 4, no. 5, pp. 371–381.

Stone, Arthur A. Turkkan, Jaylan S.; Bachrach, Christine A.; Jobe, Jared B.; Kurtzman, Howard S.; Cain, Virginia S.(eds.): The science of self-reports: Implications for research and practice. Mahwah, NJ: London: Erlbaum.

Tomz, Michael; King, Gary; Zeng, Langche (1999): RELOGIT: Rare events logistic regression, version 1.1. Cambridge, MA: Harvard University, October 1, http://gking.harvard.edu/

Tomz, Michael; King, Gary; Zeng, Langche (2003): ReLogit: Rare events logistic regression. In: Journal of Statistical Software, vol. 8, no 2: pp. 1–27.

Tourangeau, Roger. (2000): Remembering what happened: Memory errors and survey reports. In: Stone, Arthur A. et al., pp. 29–48.

Tourangeau, Roger; Rips, Lance J.; Rasinski, Kenneth (2000): The psychology of survey response. Cambridge, MA: Cambridge University Press.

Van der Vaart, Wander; Glasner, Tina (2008): Personal landmarks as recall aids in survey interviews. In: Proceedings of the Seventh International Conference on Survey Methodology, Napels, Italy, 1. – 5. September 2008.

Van der Vaart, Wander; Glasner, Tina (2007): Applying a timeline as a recall aid in a telephone survey: A record check study. In: Applied Cognitive Psychology, vol. 21, no. 2, pp. 227–238.

Van der Vaart, Wander (2004): The time-line as a device to enhance recall in standardized research interviews: A split ballot study. In: Journal of Official Statistics, vol. 20, no. 2, pp. 301–317.

Wagenaar, Willem A. (1986): My memory: A Study of autobiographical memory over six years. In: Cognitive Psychology, vol. 18, no. 2, pp.225–252.

Appendix 1: Statistics

Table A
Summary statistics of the independent variables

Variable	Obser-vations	Mean/ per cent	Std. Dev.	Min	Max
Female	107,488	0.51			
Age	107,488	39.12	9.16	19	52
Highest educational degree	107,395				
Primary and lower secondary education (ISCED 1 + 2)	7,093	6.60			
Upper secondary education (ISCED 3A + 3B)	45,253	42.14			
Post-secondary non tertiary education (ISCED 4A)	13,997	13.03			
First stage of tertiary education (vocational (ISCED 5B)	14,643	13.63			
First stage of tertiary education (academic) and second stage of tertiary education (ISCED 5A + 6)	26,409	24.59			
Number of Episodes	107,488	9.74	4.03	1	36
Duration	106,880	38.96	47.89	0	462
Time-Lag	106,953	171.90	129.64	0	527

Table B
logit models for rare event date, full model and models for men and women, with added episodes as dependent variable

	Model 1 full Coeff. (Robust SE)	Model 2 male Coeff. (Robust SE)	Model 3 female Coeff. (Robust SE)
Female	-0.196***		
	(0.043)		
Age	0.018***	0.014***	0.025***
	(0.003)	(0.004)	(0.004)
Highest educational degree *(Ref. ISCED 5A + 6)*			
ISCED 1 + 2	0.252**	0.491***	-0.014
	(0.088)	(0.096)	(0.103)
ISCED 3A + 3B	-0.133*	-0.041	-0.229***
	(0.054)	(0.061)	(0.062)
ISCED 4A	-0.036	0.100	-0.171*
	(0.069)	(0.085)	(0.080)
ISCED 5B	-0.093	-0.114	-0.039
	(0.072)	(0.076)	(0.090)
Type of episode *(Ref. Employment)*			
Schooling	-1.558***	-1.913***	-1.167***
	(0.097)	(0.106)	(0.104)
Vocational education	-1.120***	-1.328***	-0.863***
	(0.060)	(0.072)	(0.074)
Unemployment	0.350***	0.347***	0.349***
	(0.052)	(0.059)	(0.061)
Military service	-1.887***	-2.011***	-1.035*
	(0.155)	(0.144)	(0.419)
Duration	-0.020***	-0.016***	-0.026***
	(0.001)	(0.001)	(0.001)
Number of Episodes	-0.111***	-0.100***	-0.128***
	(0.009)	(0.007)	(0.008)
Time-Lag	0.640***	1.094***	0.172
	(0.137)	(0.155)	(0.149)
Constant	-2.879***	-2.950***	-3.050***
	(0.060)	(0.064)	(0.067)
N	89169	47074	42095

* p<0.05 ** p<0.01 *** p<0.001

Table C
logit models for rare event date, separate models for different types of episodes, with added events as dependent variable

	Model 1 schooling Coeff. (Robust SE)	Model 2 vocational Coeff. (Robust SE)	Model 3 employment Coeff. (Robust SE)	Model 4 unemploy-ment Coeff. (Robust SE)	Model 5 military Coeff. (Robust SE)
Female	0.023	0.195	-0.401***	-0.117	0.846
	(0.123)	(0.108)	(0.059)	(0.082)	(0.437)
Age	0.095***	-0.014	0.014**	0.015**	-0.011
	(0.009)	(0.009)	(0.005)	(0.006)	(0.034)
Highest educational degree *(Ref. ISCED 5A + 6)*					
ISCED 1 + 2	-0.087	-0.628*	-0.182	0.788***	0.200
	(0.275)	(0.244)	(0.186)	(0.146)	(0.777)
ISCED 3A + 3B	-0.231	-0.957***	-0.276***	0.316**	-0.278
	(0.188)	(0.136)	(0.074)	(0.109)	(0.362)
ISCED 4A	-0.338*	-0.442***	-0.172	0.246	0.339
	(0.212)	(0.155)	(0.097)	(0.139)	(0.443)
ISCED 5B	-0.500	-0.886***	-0.018	0.073	0.397
	(0.261)	(0.162)	(0.097)	(0.157)	(0.480)
Duration	-0.040***	-0.047***	-0.016***	-0.010***	-0.008
	(0.003)	(0.004)	(0.001)	(0.002)	(0.007)
Number of Episodes	-0.139***	-0.091***	-0.195***	-0.021*	-0.050
	(0.036)	(0.019)	(0.017)	(0.010)	(0.048)
Time-Lag	-5.411***	1.389***	1.861***	0.895***	0.181
	(0.337)	(0.376)	(0.198)	(0.255)	(1.904)
Constant	-1.164***	-4.408***	-2.995***	-2.699***	-4.320***
	(0.248)	(0.215)	(0.072)	(0.121)	(0.885)
N	24082	18917	33439	9248	3483

* p<0.05 ** p<0.01 *** p<0.001

Table D
logit models for rare event date, full model and modes for men and women, with corrected episodes as dependent variable

	Model 1 full Coeff. (Robust SE)	Model 2 male Coeff. (Robust SE)	Model 3 female Coeff. (Robust SE)
Female	-0.105**		
	(0.039)		
Age	-0.003	-0.008	0.000
	(0.003)	(0.004)	(0.004)
Highest educational degree *(Ref. ISCED 5A + 6)*			
ISCED 1 + 2	0.024	0.057	0.006
	(0.089)	(0.125)	(0.128)
ISCED 3A + 3B	0.055	0.051	0.065
	(0.049)	(0.069)	(0.071)
ISCED 4A	0.106	0.042	0.157
	(0.064)	(0.097)	(0.088)
ISCED 5B	0.109	0.105	0.114
	(0.059)	(0.075)	(0.098)
Type of episode *(Ref. Employment)*			
Schooling	-0.611***	-0.817***	-0.395***
	(0.063)	(0.089)	(0.089)
Vocational education	0.262***	0.121*	0.424***
	(0.040)	(0.054)	(0.057)
Unemployment	0.430***	0.334***	0.514***
	(0.053)	(0.076)	(0.074)
Military service	0.320***	0.228**	0.369
	(0.076)	(0.079)	(0.365)
Duration	0.005***	0.005***	0.005***
	(0.000)	(0.000)	(0.000)
Number of Episodes	0.004	0.004	0.004
	(0.005)	(0.007)	(0.008)
Time-Lag	0.881***	1.191***	0.543***
	(0.106)	(0.154)	(0.147)
Constant	-3.364***	-3.379***	-3.448***
	(0.051)	(0.067)	(0.068)
N	106471	52396	54075
	* p<0.05	** p<0.01	*** p<0.001

Table E
logit models for rare event date, separate models for different types of episodes, with corrected events as dependent variable

	Model 1 schooling Coeff. (Robust SE)	Model 2 vocational Coeff. (Robust SE)	Model 3 employment Coeff. (Robust SE)	Model 4 unemploy-ment Coeff. (Robust SE)	Model 5 military Coeff. (Robust SE)
Female	0.031	-0.021	-0.089	-0.085	-0.846
	(0.083)	(0.069)	(0.055)	(0.098)	(0.437)
Age	0.074***	-0.021***	-0.020***	-0.011	0.011
	(0.006)	(0.006)	(0.004)	(0.007)	(0.034)
Highest educational degree *(Ref. ISCED 5A + 6)*					
ISCED 1 + 2	0.030	-0.137	0.204	-0.247	0.200
	(0.181)	(0.213)	(0.140)	(0.203)	(0.777)
ISCED 3A + 3B	0.135	-0.117	0.265***	-0.231	-0.278
	(0.115)	(0.090)	(0.072)	(0.126)	(0.362)
ISCED 4A	-0.015	-0.009	0.259	0.002	0.339
	(0.141)	(0.115)	(0.093)	(0.158)	(0.443)
ISCED 5B	0.266*	-0.167	0.390***	-0.057	0.397
	(0.135)	(0.107)	(0.086)	(0.172)	(0.480)
Duration	0.004***	0.015***	0.005***	0.007***	-0.008
	(0.001)	(0.001)	(0.000)	(0.001)	(0.007)
Number of Episodes	0.013	0.000	0.005	-0.032**	-0.050
	(0.013)	(0.010)	(0.007)	(0.012)	(0.048)
Time-Lag	-4.253***	2.054***	2.467***	1.118***	0.181
	(0.229)	(0.269)	(0.166)	(0.310)	(1.904)
Constant	-1.156***	-3.361***	-3.639***	-2.720***	-4.320***
	(0.160)	(0.106)	(0.074)	(0.128)	(0.885)
N	24082	19105	33439	9248	3483
	* p<0.05	** p<0.01	*** p<0.001		

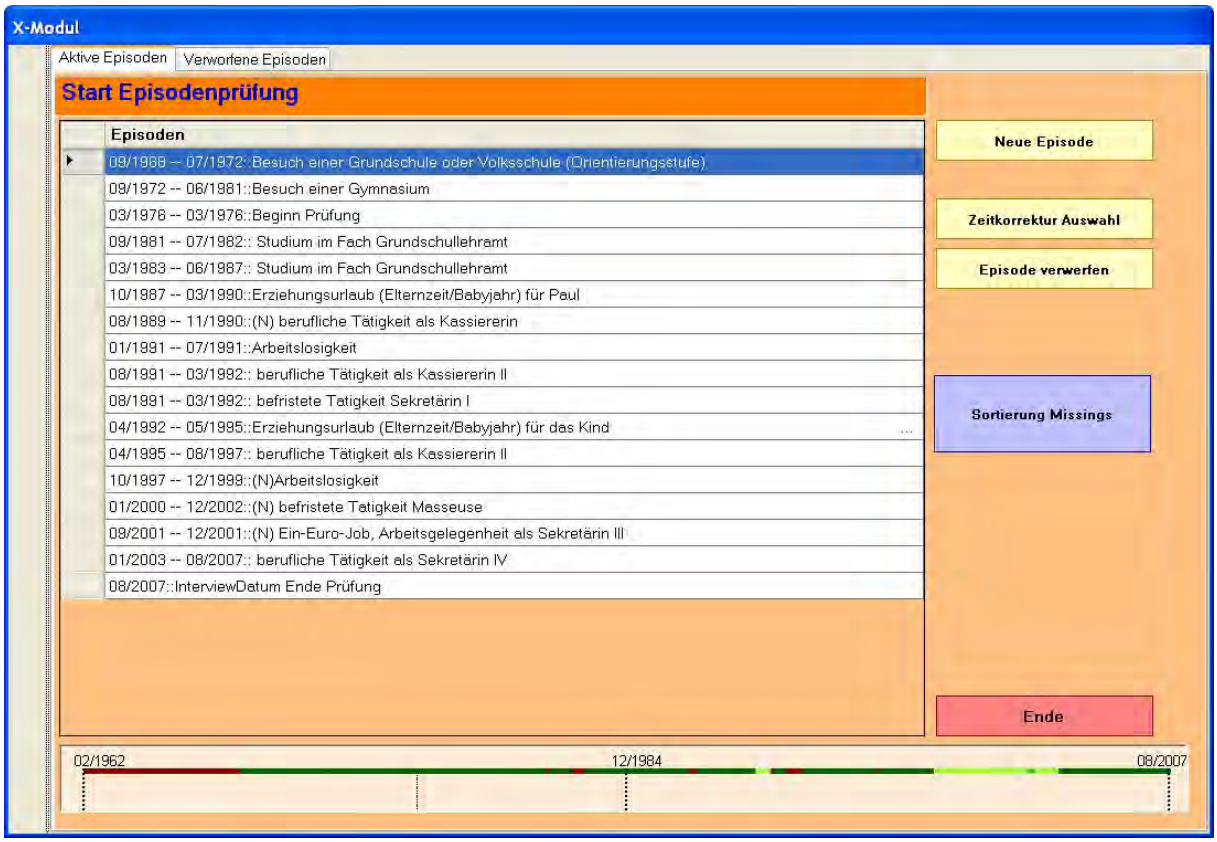
Appendix 2: Technical implementation

Design and Technical Features of the Data Revision Tool[11]

After the first stage of data collection is completed, the revision module calculates if there are any gaps or overlaps between the reported episodes by using pre-defined algorithms; it then starts the data revision routine. The technical implementation of such a data revision module is not trivial, especially when overlaps of specific events such as marginal employment episodes and schooling episodes or other possible overlaps are going to be permitted. Those have to be pre-defined when programming such a tool.

On the main screen of the data revision module all reported episodes are displayed in an episode list. This list resembles an event history calendar. Beneath, the episodes are visualized in their temporal extension as coloured episode bars in a timeline from the birth of an individual until the interview date. Gaps are shown in red and temporal inconsistencies are highlighted in light green. Figure A displays the main page.

Figure A
Layout of the main page of the data revision module



[11] A more detailed description can be found in the IAB-ALWA interviewer manual which can be provided upon request.

Once all gaps and inconsistencies are retrieved, the standardized question procedure is started. The procedure also includes previously given open text answers of the respondent (here highlighted and in blue font) such as occupational titles, using dependent interviewing techniques. This personalizes the question routine. Figure B displays an example of an automated problem solving routine of a temporal gap.

Figure B
Layout of an automated problem solving routine in the data revision module

Three different possibilities how to fill this temporal gap are provided. First it is possible to insert a new episode of any type (Code 1). Second dating corrections can be performed (Code 2). Or third a so-called side episode (e.g. a secondary job) can be declared as main activity (Code 4). The respondent can also deny to give an answer (Code 7: refused or Code 8: do not know). Gaps and inconsistencies that ended before the respondent's 14th birthday are not resolved in the data revision module due limited available interview time.

Additionally, interviewers are allowed to insert, correct or delete episodes flexibly upon their own initiative based on information given by the respondent. Therefore, different buttons (`New Episode´, `Dating Correction´ or `Delete Episode´) on the first screen on the right hand side exist that allow these procedures. Furthermore comments can be written down at any point in the data revision module if changes could not be implemented for whatever reason, e.g. when a respondent suddenly realized that he wrongly dated several episodes with the same memory error (e.g. made a mistake and wrongly dated all schooling episodes by one year).

The information that the data revision module is completed and all gaps and inconsistencies are cleared up is passed on to the interviewer by giving a visual sign. Only then or in certain emergency cases when the module is stuck for a (technical) reason the data revision module can be quitted.

Recently published

No.	Author(s)	Title	Date
6/2009	Ruppe, K.	Eingliederungszuschüsse und Betriebszugehörigkeitsdauer in Westdeutschland	4/09
7/2009	Krug, G.	Fehlende Daten beim Record Linkage von Prozess- und Befragungsdaten: Ein empirischer Vergleich ausgewählter Missing Data Techniken	5/09
8/2009	Kunz, M.	Disparities, persistence and dynamics of regional unemployment rates in Germany	5/09
9/2009	Stephan, G.	Employer wage subsidies and wages in Germany: Some evidence from individual data	6/09
10/2009	Oberschachtsiek, D.	What makes a 'Jack-of-all-Trades'?	6/09
11/2009	Kunz, M.	Unemployment dynamics in West Germany: Do districts adjust differently than larger regional units?	6/09
12/2009	Gartner, H. Merkl, Ch. Rothe, Th.	They are even larger! More (on) puzzling labor market volatilies	6/09
13/2009	Hainmueller, J. Hofmann, B. Krug, G. Wolf, K.	Do more placement officers lead to lower unemployment? Evidence from Germany	7/09
14/2009	Hornig, St. Rottmann, H. Wapler, R.	Information asymmetry, education signals and the Case of Ethnic and Native Germans	7/09
15/2009	Niebuhr, A. Granato, N. Haas, A. Hamann, S.	Does labour mobility reduce disparities between regional labour markets in Germany	8/09
16/2009	Schels, B.	Job entry and the ways out of benefit receipt of young adults in Germany	8/09
17/2009	Kopf, E.	Short-term training variety for welfare recipients: The effects of different training types	8/09
18/2009	Fuchs, M.	The determinants of local employment dynamics in Western Germany	8/09
19/2009	Kunz, M.	Sources for regional unemployment disparities in Germany: Lagged adjustment processes, exogenous shocks or both?	8/09
20/2009	Hohmeyer, K.	Effectiveness of One-Euro-Jobs: Do programme characteristics matter?	8/09

As per: 26.08.2009

For a full list, consult the IAB website
http://www.iab.de/de/publikationen/discussionpaper.aspx

Imprint

IAB-Discussion Paper 21/2009

Editorial address
Institute for Employment Research
of the Federal Employment Agency
Regensburger Str. 104
D-90478 Nuremberg

Editorial staff
Regina Stoll, Jutta Palm-Nowak

Technical completion
Jutta Sebald

Website
http://www.iab.de

Download of this Discussion Paper
http://doku.iab.de/discussionpapers/2009/dp2109.pdf

For further inquiries contact the authors:

Katrin Drasch
Phone +49.911.179 4762
E-mail katrin.drasch@iab.de

Britta Matthes
Phone +49.911.179 3074
E-mail britta.matthes@iab.de