methods, data, analyses | Vol. 11(1), 2017, pp. 45-66 DOI: 10.12758/mda.2016.013

Collecting Event History Data with a Panel Survey: Combining an Electronic Event History Calendar and Dependent Interviewing

Josef Brüderl[1], *Laura Castiglioni*[1], *Volker Ludwig*[2], *Klaus Pforr*[3] & *Claudia Schmiedeberg*[1]

1 *LMU Munich*
2 *University of Kaiserslautern*
3 *GESIS*

Abstract

Many panel surveys collect event history data on events occurring between two waves. This is usually done by asking lists of questions on the various changes that took place between interviews (Q-Lists). Recently, some panel surveys introduced a different data collection method: the Event History Calendar (EHC), credited for collecting more accurate data. However, even the use of an EHC cannot prevent the issue that events tend to be reported spuriously at the seam of consecutive waves (seam effect). On the other hand, research has shown that dependent interviewing (DI) can help reduce this seam effect. Thus, the combination of EHC and DI (DI-EHC) promises to provide more accurate event history data that are not plagued by a seam effect. The German Family Panel pairfam was one of the first panel studies to use DI-EHC. In this article we first report on the practical aspects and the pros and cons of DI-EHC. Further, we report the results of an experiment in which we test whether DI-EHC reduces the seam effect. In sum our practical experiences and the results of our experiment indicate that the instrument is less burdensome than traditional Q-Lists and produces more accurate data. In particular, DI-EHC reduces the seam effect significantly.

Keywords: Event History Calendar; Dependent Interviewing; Seam Effect; Panel Survey; Questionnaire Design

1 Introduction

Panel surveys ask prospective questions about respondents' life situation and status at the time of the interview. When done repeatedly over several waves, this process produces panel data. In addition, many panel surveys also collect event history data by asking respondents retrospective questions regarding status changes such as transitions and events that occurred in the time since the last interview. Compared with classic panel data, such event history data allow for a more precise modelling of the timing of certain events (e.g., survival analysis). Traditionally, event history data have been collected by means of question lists (Q-Lists), looping over the statuses that have been reported by respondents and asking about the beginning and end time of each episode. These loops can move forward from the status at the last interview or backward from the current status.

However, retrospective reports of episodes can be biased by recall mistakes. A number of recall errors have been reported in the literature (Eisenhower, Mathiowetz, & Morganstein, 1991; Sudman & Bradburn, 1974). For instance, sometimes respondents do not report events or episodes altogether, leading to omission or underreporting of events. In other cases, timing errors such as telescoping or time expansion occur, i.e. reporting an event as having been more or less frequent than it actually was. These mistakes are a potential source of bias in event history data.

1.1 Event History Calendars

To improve the quality of retrospective data, calendar-based techniques – initially in form of paper-and-pencil calendars – have been suggested since the late 1960s as an alternative to Q-Lists (Balan, Browning, Jelin, & Litzler, 1969; Freedman, Thornton, Camburn, Alwin, & Young-DeMarco, 1988). Calendar instruments typically consist of a two-dimensional grid with the X-axis representing the timeline

Acknowledgments

This paper uses data from the German Family Panel pairfam, coordinated by Josef Brüderl, Karsten Hank, Johannes Huinink, Bernhard Nauck, Franz Neyer, and Sabine Walper. pairfam is funded as long-term project by the German Research Foundation (DFG). We are grateful to an anonymous reviewer, whose comments improved this paper very much. Further, we thank Madison Garret for language editing and proofreading.

Direct correspondence to

Josef Brüderl, Ludwig-Maximilians-Universität München, Institut für Soziologie, Konradstr. 6, 80801 München, Germany
E-Mail: bruederl@lmu.de

(e.g. with months or years being the columns), and the Y-axis life domains such as employment or residences (with the respective statuses in place of the rows)[1]. Using this grid, respondents receive visual cues about the period on the timeline and can easily indicate for which cells of the grid an event or episode should be recorded. Landmark events such as birthdays or holidays can be included in the calendar to facilitate the timing of events. For retrospective surveys, calendar-based methods have become rather common (see the literature review provided in Glasner, 2011, p. 45). Since the late 1990s calendar instruments have also been introduced in electronic form in large panel surveys. In the panel context it has become common to term such instruments "Event History Calendars", or EHC (Belli, Stafford, & Alwin, 2009).

Many survey researchers argue that calendar instruments facilitate recall accuracy by means of a graphical presentation of timelines with visual cues that better fit respondents' idiosyncratic autobiographic memory structures (Belli, 1998). Furthermore, the conversational style of the interview improves respondents' recall (Belli, 2000; Caspi et al., 1996). Based on the graphical timeline, respondents are able to relate events to each other and detect gaps and inconsistencies in records (van der Vaart, 2004). For instance, landmark events can be used as temporal anchor points to which respondents can relate other events (e.g., "We moved to X the week before Christmas"). Similarly, multiple-domain calendars can help to link events across life domains (e.g., "We moved in together just before I graduated"). Accordingly, evaluations of calendar-based techniques have shown that calendar instruments improve data quality regarding completeness and consistency compared to data collection by means of question lists (Belli & Callegaro, 2009; Glasner & van der Vaart, 2009). Although the beneficial effects of calendars were found to be more important for recall of less recent events (Glasner & van der Vaart, 2009), they may be just as helpful for accurate reports of the relatively short periods between panel waves.

Data quality is increased further thanks to calendar instruments as they improve the interviewing process. The graphical representation of the information already recorded in the calendar renders detection of gaps and inconsistencies very easy for the interviewers, who receive cues to probe accordingly. For this reason, EHCs are also implemented in telephone surveys such as the PSID, where solely the interviewer, not the respondent, can see the calendar. A typical feature of calendar-based data collection is the greater degree of flexibility allowed to interviewers: they may deviate from the given question order and wording to help the respondent more accurately recall a series of episodes (Belli & Callegaro, 2009). Indeed, research has shown that interviewer variance is slightly increased by the use of EHC methods in a CATI survey (Sayles, Belli, & Serrano, 2010), which can be interpreted

1 For a detailed description of characteristics of calendar-based data collection see Callegaro (2007).

as a sign of greater flexibility provided by this method. As a consequence, a conversational interaction is possible which may lead to higher motivation and reduce satisficing (Belli & Callegaro, 2009; Belli, Lee, Stafford, & Chou, 2004; Krosnick, Narayan, & Smith, 1996). Calendar instruments are in fact reported to be preferred by both respondents and interviewers over question lists (Freedman et al., 1988). A field test of the newly developed EHC in the re-engineered SIPP revealed that respondents perceived the calendar-based instrument as more interesting than the traditional interview (Chan, 2009). In the experimental comparison between EHC and Q-Lists conducted in the PSID 1998 Calendar Methods Study (Belli, Shay, & Stafford, 2001), interviewers reported to have enjoyed the EHC interviews more than traditional question lists.

1.2 Seam Effect

In the context of panel surveys, recall errors may produce a specific methodological problem: the so-called "seam effect". A seam effect means that we observe a higher rate of change at the seam between two consecutive panel waves than within the period a respondent reports on during the interview (Burkhead & Coder, 1985; for a review see Callegaro, 2008). Seam effects are the product of both the underreporting of transitions within a wave ("constant wave reporting") and spurious changes between waves (Jäckle, 2008; Rips, Conrad, & Fricker, 2003). In particular, a spurious change can occur if the respondent classifies the same status differently in two consecutive waves (misclassification). Another mechanism is omission: in this case, the last months of an ongoing episode from the previous calendar are "forgotten" in the next wave. Finally, due to backward telescoping transitions are often dated back to the seam. Thus, paradoxically, when collecting event history data via panel surveys we might minimize retrospective recall bias on the one hand; however, on the other we introduce artificially high transition rates at the seams.

Data collection using an EHC can help to decrease seam effects (Callegaro, 2007). Research has shown that calendar-based data collection methods are often superior to question lists with regard to underreporting or time error (Belli, Shay, & Stafford, 2001; Belli, Smith, Andreski, & Agrawal, 2007). Thus, as calendar instruments facilitate recall – e.g. due to the use of landmark events, visual cues regarding the temporal order of episodes, and the visibility of inconsistencies in entries – the accuracy of event history data will be improved and inconsistencies between waves will be less likely (Callegaro, 2007; Rips et al., 2003).

As a more specific method to tackle the seam problem, dependent interviewing (DI) has been introduced by a number of panel studies since the 1990s. Information from previous waves is preloaded to tailor the wording of questions (proactive DI), or for automatic consistency checks (reactive DI) (Callegaro, 2008; Jäckle, 2009). For instance, in proactive DI, instead of recording the employment status at the

beginning of the reference period, the interviewer asks if the respondent has maintained the same employment status recorded in the previous interview ("according to my records, last year you …, is this still the case?"). In reactive DI, automatic consistency checks may highlight if the status reported for the same point in time in the previous wave differs from that reported at the current interview. In this case, interviewer and respondent can revise the data together to solve the inconsistency.

DI has been proven an effective method to reduce seam effects (Jäckle & Lynn, 2007; Moore, Bates, Pascale, & Okon, 2009) as preloads reduce both the chance of misclassification and omission (Lynn, Jäckle, Jenkins, & Sala, 2012; Lynn & Sala, 2006). Further, the problem of backward telescoping should be minimized: transitions cannot be dated back to the seam as the preloaded status must first come to an end.

1.3 DI-EHC

Building upon this knowledge, it seems promising to combine EHC and DI (DI-EHC) as a means to increase recall accuracy and reduce seam effects (Callegaro, 2008). While DI reduces spurious change between waves due to misclassification, omission, or backward telescoping, EHC may help reduce constant wave reporting and underreporting of short or seemingly irrelevant episodes. One of the aspects of DI which could be potentially problematic is that DI can trigger cognitive satisficing (see Krosnick, 1991): respondents might feel that their interview is easier and shorter if they confirm the data prompted by the preload. However, by implementing DI in a calendar setting the pairfam questionnaire does not offer any strong incentives for confirming the preload throughout the reference period as the interviewers need the same amount of time whether they check off one category or another[2]. Hence, also in this respect, the combination of EHC and DI might trigger positive synergies between the two methods.

The German Family Panel pairfam has introduced a DI-EHC for collecting data on partnerships, residences, education, and employment. The aim was to improve recall accuracy and to reduce the seam effect. Due to the lack of validation data, we cannot investigate whether the accuracy of the data increased. However, we can investigate whether the seam effect increased among a randomly chosen subgroup of the respondents for whom we experimentally excluded preload data in the education and employment calendar compared to the majority of respondents for whom all preloads were included. Therefore, the main purpose of this article is

2 Hoogendoorn (2004) found that the issue of acquiescence in connection with proactive DI can be solved by certifying that confirming the preloads would not translate to a sizable shortening of the questionnaire. Also Eggs and Jäckle (2015) and Jäckle and Eckman (2016) have found no support for the hypothesis that proactive DI leads respondents to satisfice.

to report the results of our randomized methods experiment on the effectiveness of DI-EHC for reducing the overall seam effect.

1.4 Contents of the Paper

This paper is structured as follows: First, we will give an overview of EHC modules in existing panel studies. Then, we will describe the structure of the DI-EHC in pairfam and practical aspects of its implementation. As Glasner and van der Vaart (2007) point out, in recent years calendar instruments were developed without taking advantage of experiences made in other studies. With our overview and the practical guide to the pairfam DI-EHC we hope that other studies might learn from our own experiences. The results of our experiment will follow. Finally, we conclude and discuss lessons learned and give recommendations for future developments.

2 EHC in Other Panel Studies

The Survey of Income and Program Participation (SIPP) was among the first panel studies to use calendar techniques. In fact, the first considerations when introducing calendar-based data collection in the SIPP already aimed at eliminating the seam effect and included DI techniques (Kominski, 1990). Interviewers filled out a graphical paper-and-pencil calendar after the first interview and handed it over to the respondent at the beginning of the second interview. After the second wave interview, the interviewer updated the calendar and gave it to the respondent again at the beginning of the third interview. Although the interviews were conducted using conventional question lists, respondents could use information displayed in the calendar from the previous waves as well as visual cues when answering the retrospective questions. This early EHC was implemented to aid respondents rather than as a data collection instrument itself: data were still collected by standard question lists, and the paper-based calendar distributed to respondents was used only to illustrate data entries from previous waves as a mere recall aid. A further step was taken in 2007 when a computer-assisted EHC was designed as an integral part of the survey. The reason for this development was the decision to change from the former design of three interviews per year to an annual survey (Fields & Callegaro, 2007). This shift raised concerns about respondents' ability to accurately report over this longer period. After field tests in 2008 (Chan, 2009; Pascale, 2009) and 2010 (Moore, 2012), the re-engineered SIPP including the computerized EHC was finally fielded in the 2014 SIPP Panel.

The Panel Study of Income Dynamics (PSID) also implemented an EHC when the interview cycle was changed from annual to biennial interviews (Beaulé, Das-

cola, & Liu, 2009). The "1998 PSID Calendar Methods Study" was conducted to compare the quality of data collected using the EHC versus standardized question list methods (Belli et al., 2001), but it was not until 2003 that the PSID employment module was reprogrammed as an EHC (Belli et al., 2007). As the PSID is a telephone survey, the EHC was only designed to help the interviewer detect inconsistencies such as gaps in employment history and overlaps in employment spells (Beaulé et al., 2009). The calendar spans a 2-year period and is rather detailed, with a third-of-a-month as the smallest unit (Belli et al., 2007). It contains the five following life domains: landmark events, residence, employment, not working, and time away. All domains were visible on one screen with separate summary timelines for each to facilitate parallel retrieval. Programmed consistency checks helped the interviewer detect potential inconsistencies. The experiences of the PSID team showed that by using the calendar method, post-processing time could be reduced (Beaulé et al., 2009).

A similar approach was taken by the adult cohort of the National Education Panel Survey (NEPS) in Germany. As it is also a telephone survey, event history data on education and employment are collected via Q-Lists using DI. This Q-List module is followed by a calendar-based data-revision module: the survey software automatically reorganizes all entries into calendar form in order to support the interviewer in correcting inconsistencies and in detecting biography gaps (Drasch, Kleinert, Matthes, & Ruland, 2016; Trahms, Matthes, & Ruland, 2016).

Other studies use simple calendars for single domains, for instance the Household, Income and Labour Dynamics in Australia Survey (HILDA, (Watson, 2009)) and the German Socio-Economic Panel (SOEP). We will not describe these calendars in detail here.

To summarize, several large scale panel studies have been combining DI and Q-Lists. Several studies have also used EHCs, albeit mainly for the purpose of data editing. However, to our knowledge thus far only one large scale panel study –pairfam – has implemented a combination of DI and EHC.

3 DI-EHC in the German Family Panel (pairfam)

The German Family Panel pairfam (Huinink et al., 2011) is a multi-disciplinary, longitudinal study on partnership and family dynamics in Germany based on a nationwide random sample of initially more than 12,000 persons of the three birth cohorts: 1971-73, 1981-83, 1991-93. Starting in 2008 the panel study collects data in annual waves via computer-assisted personal interviews (CAPI) administered by professional interviewers.

The purpose of the study is to collect comprehensive data on respondents' intimate relationships and family life, as well as social and economic circumstances. Research topics include partnership formation, institutionalization of intimate relationships, family formation and parenthood, and separation and divorce. For such research questions, accurate data regarding the temporal ordering of events including the start of a relationship, moving in together, marriage, or separation are crucial. Therefore, one of the core features in pairfam is an EHC on intimate relationships, places of residence, and occupations (i.e., school enrollment and labor force participation) spanning the period between the previous and current interview.

The EHC in pairfam has several unique features (presented in more detail below) to ensure high quality data. When developing the calendar we considered both theoretical findings on memory structure (Belli, 1998) as well as existing instruments (e.g. Belli et al., 2007; Reimer & Matthes, 2007), adapting them to the specific interests and needs of the pairfam study. In particular, the pairfam EHC incorporates DI techniques and implements an individually adjusted calendar span: the starting month is that of the last interview, and the maximal duration is set to 32 months to accommodate for one wave non-participants.

The EHC is a stand-alone Java application that is fully integrated into the pairfam interview. This was done for the higher flexibility regarding graphic interfaces offered by Java compared to the available CAPI software[3]. The EHC starts as a pop-up window after some "warm-up" questions. It displays some information we feed forward from the previous wave(s), the so-called "preloads". At the end of the EHC module the collected data are stored in the main dataset of the CAPI-software so that all entries are available in the following part of the interview for routing[4].

Interviewers allow the respondent to observe the screen while completing the EHC to ensure that both the interviewer and the respondent profit from the graphic representation of the entries. For each calendar a scripted introductory text is given while further probing in the case of gaps and inconsistencies is not scripted. In addition, questions are scripted for each line in the calendar ("In which months were you together with [name of the partner]?").

For illustration purposes, we present here the introduction question in the case of the partnership calendar:

> *"We would now like to know in which months you and your partner were in a relationship, if and when you were living together, or were married. You can see here a calendar with one column for each month between the last interview and today. Also, your current partnership status is entered in the column*

3 The pairfam questionnaire was initially programmed in IN2Quest until wave 3. From wave 4 onwards, the questionnaire runs on a NIPO CAPI software. Both programs offer limited support for tailoring the graphic of the questionnaire interface.

4 The EHC is designed to be fully functional also without preloads. This is to facilitate the implementation of a refreshment sample.

*labelled 'now'. For each partner there are three lines, one for having a relationship, one for living together, and one for marriage or civil union. We will proceed as follows: You will look at the screen and tell me what has happened since the last interview [*first-time respondents: *"EHC time period"], and I will enter the data. After we are finished, you can check if the information I have entered is correct".*

The EHC covers three life domains: intimate relationships (including cohabitation and marriage), places of residence, and education and/or employment episodes. Furthermore, two "synoptic calendars" conclude the EHC in order to crosscheck entries before the EHC is closed and the normal interview continues. A detailed description of the EHC including specific wording of questions and consistency checks in pop-up windows can be found in the pairfam codebooks (e.g. pairfam, 2015).

All calendars enable monthly entries and cover the time span since the last interview. Therefore, as the period between two interviews varies between respondents the length of each calendar is adjusted for each respondent individually. The last available information (column "now") from the previous interview is used to produce the preloads for the next wave.

In addition to the monthly entries the pairfam EHC includes one further column for the current situation (column "now") in order to take into account the most recent changes in respondents' lives. This is particularly important as information from the EHC is subsequently used for routing purposes in the remainder of the interview. For instance, let us assume that the interview takes place in January and the respondent reports that their relationship ended in January, too. Some respondents might then say that January is the last month in a relationship (for instance, if they split up towards the end of the month) and hence should be marked as such in the calendar. However, for the following section of the interview we will want to save the information that the respondent is currently single. For this reason, the interviewer will enter January as the last month in a relationship and make sure that the cell corresponding to 'now' of the respective row is unmarked.

Once all data are entered, a box at the end of each row in the calendar must be ticked in order to show that the row has been completed. The check mark disappears again if the interviewers alter any entry in this row. This feature was implemented to ensure that interviewers notice unintentional changes of the record.

The first life domain covered in the EHC is intimate relationships. Other calendars (e.g. SIPP) often begin with respondents' residential mobility, since moving is a rather seldom event and moving dates tend to be easy to recall. In the pairfam study we decided to start with a life domain that is more central to our study as we thought that this might help to keep the attention high. The first screen requires that respondents list the names of the partners with whom they were together since the last interview (partner list). If a respondent had a partner in the last interview his/

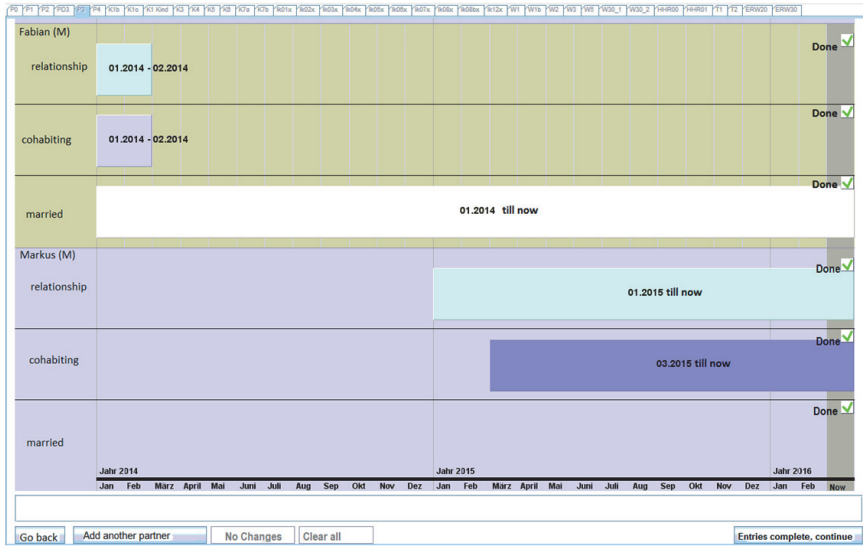


Figure 1 Partnership calendar containing two relationships: the respondent was married and living together with her partner at the time of the previous interview, is now still legally married but lives together with a new partner.

her name, gender, and date of birth is already included in that list as a preload. This introductory screen was not introduced for a methodological but rather for a technical reason: knowing the number of partners prior to opening the calendar helps optimizing the height of the rows in the calendar, as the calendar always contains only as many rows as necessary for the number of partners mentioned.

After the partner list is completed, the partnership calendar is shown (Figure 1). All names entered in the list appear automatically in the calendar view. First the partner from the last interview is listed, then, where necessary, new partners. For each of the partners there are three rows: the first one reports in which months the relationship existed, the second one in which months the respondent and their partner lived together, and the last one is for reporting marriage duration. For the preloaded partner the cells of the interview month of the last interview are marked according to the information given in the column "now" in the last interview. Thus, we preload information on the status of the partnership also.

To avoid incorrect entries, the calendar includes a number of consistency checks which are run as soon as the interviewer declares the data entry to be complete for this life domain. For instance, if parallel marriage episodes (i.e. with two partners) are entered, a pop up window will indicate an inconsistency that requires

a correction. In other cases such as parallel cohabitation episodes, the consistency check only triggers a pop-up window but correction is not required, as this is a rare but possible arrangement. Additionally, if respondents indicate that a relationship has ended, a pop-up window appears with a question as to whether the relationship ended due to separation or death of the partner. Similarly, in case of a new marriage a pop-up window asks if the wedding ceremony was religious or civil.

Like the partnership calendar, the following residence calendar is preceded by one question recording all the places of residence in which the respondent has lived since the past interview (residence list). For each place of residence the municipality and the federal state are entered. The place of residence at the time of the last interview is already included at the top of the list as a preload variable. The design of the residence calendar is similar to the partnership calendar: each place of residence listed in the introduction question is assigned to a row of the calendar table. The preloaded place of residence is displayed in the upper row and the month of the past interview is already marked (preload). For each place of residence the interviewer marks the months in which the respondent lived there (see Figure 2). Gaps are not acceptable: the respondents must indicate a place of residence for each month. Overlaps of two places of residence of not more than one month are allowed in order to account for moves within a month. If respondents wish to add a further place of residence during calendar completion, a button adds a further row without turning back to the residence list. In addition, as respondents might have difficul-

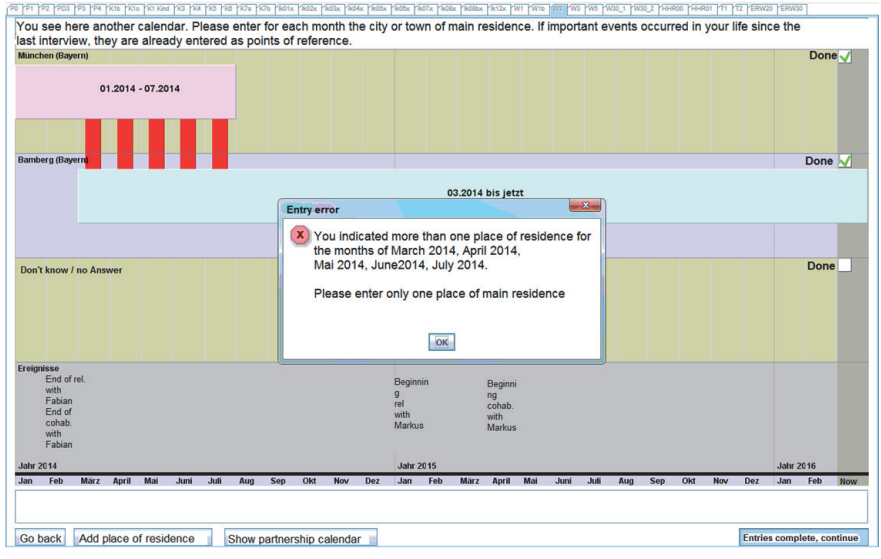


Figure 2 Residence calendar. The respondent entered parallel residence spells for five months instead of marking only the main place of residence

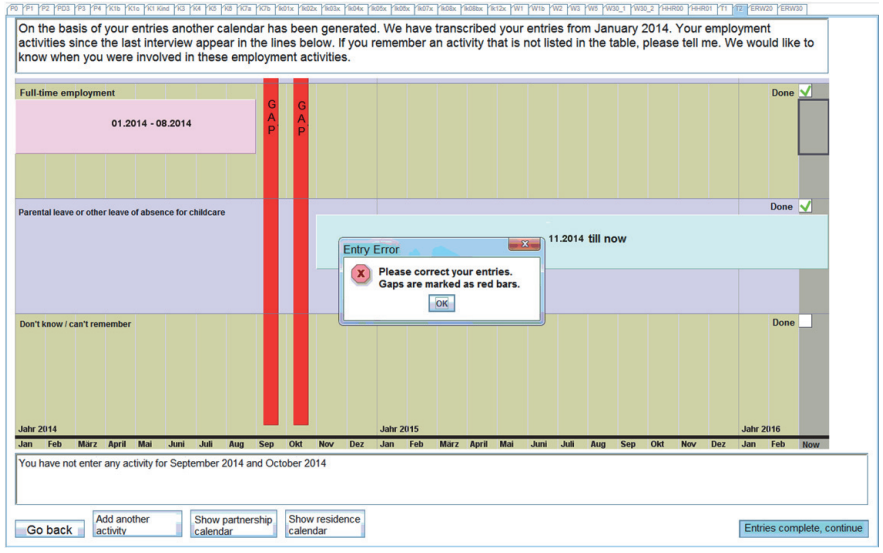


Figure 3 Activity calendar. The respondent entered no activity for two months, which are marked red.

ties to align the timing of a move with major events indicated in the partnership calendar, beginning and end time of relationships and children's dates of birth are displayed at the bottom of the calendar.

Finally, the calendar for education and employment (activity calendar) is preceded by a list of 22 possible activities (activity list). In contrast to the first two calendars, the activity list is not prefilled with preloads. The activities reported in the previous interview are displayed afterwards in the calendar together with the activities ticked in the activity list. Thus, concerning the activity calendar DI is not used in the first step, when collecting last years activities, but only in the second step, when filling out the calendar. It turned out that this design decision was suboptimal (see below).

The activity calendar (Figure 3) contains a row for each of the relevant activities. If no activity is ticked in the activity list, the row "don't know" appears in the calendar (in addition to the rows of the preloaded activities). Also in this case some crosschecks are programmed to ensure that unlikely combinations of activities cannot be entered by mistake. Gaps are not acceptable: months with no status are marked red in the calendar and a pop-up window lists all months with no information.

Our experiences with the EHC are positive. Most parts of our DI-EHC work smoothly and deliver plausible data. The residence calendar was the only

one which required a few structural changes in the first years of its implementation as we decided to reduce the level of precision of our residential history by focusing only on the primary place of residence. In the first years, we required the respondents to enter also their secondary residence places, but this effort did not pay off: the data collected were often contradictive and many changes turned out to be spurious. After giving up a comprehensive data collection of secondary residences, we opted for a simple question as to whether respondents have a second residence. This information is necessary for subsequent questions regarding respondents' mobility, for example. Further minor adjustments were necessary in the partnership and children modules. We introduced additional checks to avoid preloads being deleted by mistake and now require the interviewer to enter a reason for deleting partners or children.

Interviewers are used to standardized question lists but not to the more flexible calendar-based instruments. Therefore, interviewers were made acquainted with the EHC prior to the field start of wave 2 – the first one with an EHC. Nonetheless, the pairfam team discovered that a certain number of typical coding errors had occurred in the first waves after implementation. Preload deletions occurred particularly often. In subsequent waves some of these errors could be eliminated by implementing additional pop-ups. To further improve data quality we also introduced an interview rehearsal in wave 4. Before the start of each wave, a fictive case is constructed with a large number of (more or less complicated) events and transitions during the period covered by the EHC. Interviewers receive a written description of the case and have to record this fictional interview in the EHC. From the data produced the project team can examine the errors made and specifically address those issue both during interviewer training and in the interviewer handbook. Interviewers who made too many errors receive additional training. After the introduction of this rehearsal interview the number of coding errors decreased. In the following waves, we tailored the description of the fictive case to address specific concerns and recurrent mistakes detected during data cleaning.

Producing preloads for the next wave is quite demanding: Each year at the end of May the survey agency delivers the raw data of the last wave and preloads are needed by the end of September before the fieldwork of the next wave begins in October. We feed forward more than 300 preload variables which need to be validated for plausibility and, in the case of data such as names, places, and job descriptions, must be checked for spelling errors. Preloads must be prepared carefully as mistakes can cause unpleasant incidents during the interview.

4 Does DI-EHC Reduce the Seam Effect? Results from a Randomized Experiment

In order to investigate the effectiveness of the pairfam DI-EHC in reducing the seam effect, we implemented an experiment in wave 3 of the survey: we randomly selected 1,000 (11%) of the 9,069 wave 2 respondents and deleted their preloads before fieldwork started. We decided to limit the experiment to the activity calendar (education and employment status). In wave 3, 7,383 respondents from wave 2 could be re-interviewed. For 813 of these respondents (11%) no preloads where shown in their activity calendar. The other 6,570 respondents got a complete set of preloads[5].

Calendar data collected in wave 3 were matched with those from wave 2 in order to analyze transitions in educational or employment status at the seam from wave 2 to wave 3. We use the monthly information on respondents' status (variables ehc19i$m1-ehc19i$m18). We set up a long format panel data set where each row is a person-month. The data cover all months from the wave 1 interview to the wave 3 interview[6]. For the wave 2 interview month (MonthIntW2) three pieces of information on the activity status are available: status in MonthIntW2 as collected in wave 2, current status also collected in wave 2 (ehc19i$), and status in MonthIntW2 as collected in wave 3 (ehc19i$m1). For transition analyses on a monthly basis one has to decide, which status information should represent MonthIntW2. We decided to use the information on the current activity status that is recorded in the wave 2 interview (ehc19i$). Individual panels are organized such that the seam between waves 2 and 3 is at month 0 and up to 17 preceding and 17 following months are available. Note that due to varying wave distance, the number of person-months varies across persons.

Our main outcome in the experiment is the proportion of respondents reporting any change in status between two ensuing months t and t-1. We expect that respondents report more changes at the seam (that is, between the month of the wave 2 interview and the following month) than off-seam (any other month). However, the seam effect should be smaller for respondents who do see preloaded calendar information during the interview compared to respondents whose preloads were deleted as part of the experiment.

Our analysis is based on pairfam data release 6.0 (Brüderl, Hank et al., 2015). More details on the study can be found in Huininik et al. (2011). We decided to use

5 Due to the experimental design, treatment assignment took place before fieldwork started. Hence, the cases interviewed were less than those originally selected. However, attrition rates are very similar across experimental groups (19.6% drop-out with preloads and 19.7% without).

6 Employment status at the wave 1 interview month is not included in the wave 2 data. Due to a programming error, the information has not been recorded (variables ehc19i$m1 are empty in wave 2).

edited and released data (not exactly the raw data) as we are interested in the seam effect in the data actually available for research. However, the pairfam data team applied only minor changes to education and employment histories (Brüderl, Hajek et al., 2015).

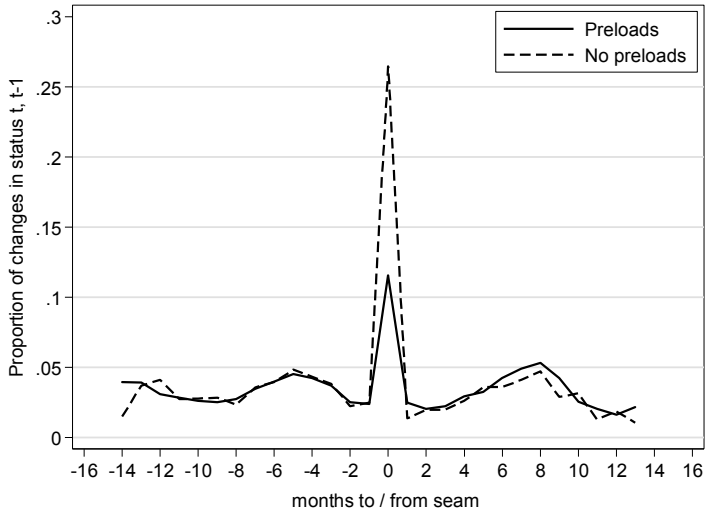
In the analyzed data there are 6,569 respondents with preloads (treatment group) and 813 respondents without preloads (control group). The two experimental groups provide 173,807 and 21,513 person-months, respectively. Note that one person belonging to the control group had to be excluded because the event history was invalid (as identified during the data cleaning process). Further, we excluded all person-months without any information on status (gaps in histories). Together, these restrictions eliminated 7,272 person-months (4.2%) from the treatment group and 912 person-months (4.2%) from the control group.

As the outcome of our analysis is defined as status changes between two months t and t-1, the final number of cases in the data set is smaller. For the earliest month of each respondent, a change is not defined. Furthermore, change is not defined for gaps in individual panels. The analysis of the proportion of changes in status is therefore limited to 159,831 (treatment group) and 19,773 person-months (control group). On average, repondents provide information on status changes for 24.3 months in both groups.

Results of our analysis are presented in graphical form in Figure 4. There is clear evidence of a seam effect in both groups. As expected, however, the seam effect is much smaller with preloads. Thus, our experiment demonstrates that using preloads substantially and significantly (see below) reduced the seam effect in the monthly education and employment histories in pairfam. Further, there seem to be no systematic differences between treatment groups off the seam. Thus, as intended, the treatment (preload information) reduces only artificial seam changes, but not "real" changes off the seam.

For taking a closer look, Table 1 shows sample proportions of monthly status changes on and off the seam for the control group (columns (1) and (2)) and for the treatment group (columns (4) and (5)). The difference in the proportion of changes on and off the seam gives us estimates of the seam effect without preloads (column (3)) and with preloads (column (6)). The difference in the seam effects between control group and treatment group then is our estimate of the treatment effect (column (7)). It tells us to what extent using preloads reduces the seam effect present in the EHC data. For the pooled sample of all three pairfam birth cohorts (first row of Table 1), the seam effect is 23.4 percentage points without preloads, but only 8.3 percentage points with preloads. Hence, providing respondents with information preloaded in the EHC substantially and significantly reduced the seam effect by 15.1 percentage points.

Table 1 also shows separate analyses for the three cohorts (born 1991-93, 1981-83 and 1971-73). For respondents from the youngest cohort, who in many cases



Note: Proportion of respondents reporting any change in activity status between months t and t-1. In each month, status can be any of 22 categories, including multiple statuses (9 categories for education, 7 categories for employment, 6 categories for non-employment; see pairfam anchor Codebook (pairfam 2015)). The figure is restricted to a maximum of 14 months before the seam and a maximum of 13 months after because due to the annual waves the number of observations outside this interval is low.

Source: pairfam release 6.0, anchor data waves 2 and 3.

Figure 4 Proportion of cases reporting a change of activity status compared to the previous month. Time line centered at the month following the wave 2 interview.

completed secondary schooling during the time observed here, the seam effect was smaller than for the two older cohorts (see Table 1). However, all results point in the same direction: for each cohort, we found a strong seam effect, which was significantly higher if preloads were deleted in the experiment.

As Figure 4 suggests, there are hardly any systematic differences in the proportions of status changes off the seam. For the pooled sample, average proportions off the seam are .031 without preloads and .033 with preloads. This difference is not significant at reasonable levels (p=.17). (There are also no significant differences when looking at birth cohorts separately.) Obviously, random assignment to the experimental groups worked well and, as desired, preloads reduced reported status changes only on the seam, but not off the seam. We further compared proportions between treatment and control group for each single month before and after the seam. We found a significant difference only in one instance, namely for the month following the seam. In this case, the proportion of changes was slightly larger for respondents with preloads than for those without preloads (.0247 com-

Table 1 Proportion of respondents reporting a change of status compared to the previous month by treatment group and estimated seam effects with and without using preloads

	No preloads			Preloads			
	Off seam	On seam	Seam effect	Off seam	On seam	Seam effect	Treatment effect
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
All 3 cohorts	.031	.265	.234*** (14.98)	.033	.116	.083*** (20.85)	-.151*** (-9.35)
Cohort 1991-93	.032	.205	.173*** (7.59)	.035	.096	.061*** (10.55)	-.112*** (-4.77)
Cohort 1981-83	.040	.308	.269*** (8.73)	.040	.135	.094*** (11.71)	-.174*** (-5.41)
Cohort 1971-73	.021	.299	.277*** (9.76)	.023	.125	.102*** (13.93)	-.175*** (-5.96)
Persons	813	805	813	6,569	6,529	6,569	7,382
Person-months	18,968	805	19,773	153,302	6,529	159,831	179,604

Notes: Seam effects are calculated as the difference in proportions On seam – Off seam; No preloads: (3) = (2) - (1); Preloads: (6) = (5) - (4). Treatment effect is the difference-in-differences estimator; (7) = (6) – (3). Two-sided tests for significant differences in proportions, adjusted for clustering of persons; z-values in parentheses. *** p<.001.

Source: pairfam release 6.0, anchor data waves 2 and 3.

pared to .0136). The difference of 1.1 percentage points is significant at the 5 percent level (z= 1.96; p=.05). This finding is consistent with backward telescoping where respondents without a preload date the event back to the first month (seam) of the calendar.

Given the results from the experiment, the question arises, why there still is a seam effect despite using preloads? In the pairfam case, the reason might be that preloads were fed into the calendar only, but not into the introductory question list (as we reported in the descpription of the employment and education calendar, see Section 3). So, when the calendar was first shown to the respondents, they had already given information without having seen the preloads. This design feature is suboptimal, because without preloads respondents might misclassify their activity status in the month of the last interview (a common example is “part-time employment” classified as “marginal employment”, or the other way around). Respondents then see in the calendar the status that they reported in the last interview (preload) and the (misclassified) status that they reported just before in the activity

list. Surely not all respondents will then delete the misclassified status and continue with the preloaded status (then we would observe no spurious change at the seam). Instead, quite a few respondents probably ignore the preload and continue the calendar with the misclassified status. Such respondent behavior will produce an artificial change on the seam. A close inspection of the data produced by the DI-EHC showed that this "misclassification mechanism" is indeed a source of the seam effect in the pairfam activity calendar (results not shown, but available upon request).

5 Discussion and Conclusion

In sum, the results of the experiment show that pairfam successfully reduced the seam effect by using DI-EHC. Nevertheless, a sizable seam effect still remains even with DI. These findings are consistent with earlier research by Jäckle and Lynn (2007) which showed that proactive DI substantially reduced seam effects in monthly work histories, but did not eliminate them completely. In the pairfam case, most likely a seam effect remains due to the fact that preloads were not used by design when collecting last year's activities on a first screen.

This gives some hints how future research could improve on our results. Basically the design used in the pairfam activity calendar is only "partial DI-EHC". By showing preloads only on the second screen (the activity calendar) two mechanism producing the seam effect could be alleviated: omission and backward telescoping. However, the third mechanism – misclassification – still operated, because preloads were not used on the first screen, when a list of activities was shown. Therefore, we speculate that most of the remaining seam effect is due to misclassification. This could be investigated by designing an improved experiment with a third experimental group added that gets a "full DI-EHC" (preloads used on both the activity list and the activity calendar).

In addition, there are some more practical aspects when using DI-EHC. First, using preloads might also reduce interview duration. In pairfam the duration of each section of the questionnaire was recorded. In particular, the duration of each section of the EHC was tracked: The mean duration of the activity calendar was 1.36 minutes with preloads and 1.49 minutes without preloads. The difference is statistically significant ($p<0.01$). This result is particularly welcome as previous literature reports calendar interviews as such to be longer than Q-lists (Glasner & van der Vaart, 2009, p. 63).

Second, feedback from interviewers suggests that this instrument is less tiresome than question lists both for interviewers and respondents. In particular this is, because DI-EHC avoids boring repetitions.

Finally, there are a few things to bear in mind in order to achieve high-quality results with this method. Firstly, the effort necessary to have an effective and appealing instrument should not be underestimated. The pairfam team outsourced the programming to the field agency, and the actual software development started almost one year before the beginning of fieldwork. Even so, this proved to be a tight schedule. Our objectives in terms of flexibility and appeal would have required more specific programming and human interface design skills. Some of our aims (e.g., the parallel visualization of all three calendars) could not be achieved with the resources at our disposal. Remodeling the calendar is also quite resource intensive and in panels not desirable in order to ensure comparability across waves. Hence, it is mandatory to invest enough time in the conception phase.

One big advantage of an EHC is the possibility to implement rather complex consistency checks during data collection. Possible mistakes can be defined quite easily upfront by survey managers (e.g., pop-ups) and can be immediately communicated and corrected during the interview. In our experience, adding additional pop-ups to cross check improbable entries is rather simple, and facilitates avoiding accidental data entries/deletions.

Before fielding a survey with an EHC module, interviewers must be trained extensively to properly use this instrument. Using a partially scripted questionnaire is challenging, especially if they have never done it before (for a vivid illustration on what interviewers (and respondents) mess up when using such complex instruments see Uhrig and Sala, 2011). Furthermore, we found that not all interviewers were comfortable with the graphic interface. It is advisable to gain a good grasp of common mistakes and make sure that interviewers learn how to properly navigate the calendar (rehearsal interviews are a very effective method).

Finally, an EHC produces a large amount of information. For each life domain the status for every month is recorded. These are sequence data on the interval since the last interview. These "pieces" of the life-course must be consolidated in some kind of biographical data set. Often this is also done in an episode format to facilitate event-history analyses. This process is very demanding and requires a lot of manpower, especially in the first couple of panel years when the data cleaning procedures are still in development.

All in all, the setup costs of a DI-EHC are not negligible: development, programming, interviewer training, and data handling procedures will require more resources than with a traditional CAPI. Nevertheless, in the long run, costs reduce to the level of a standard interview. On the other side, data quality is improved from the beginning and is even likely to improve with each further wave. Hence, the longer the planned duration of a longitudinal study, the higher the rate of return from a DI-EHC.

References

Balan, J., Browning, H. L., Jelin, E., & Litzler, L. (1969). A computerized approach to the processing and analysis of life histories obtained in sample surveys. *Behavioral Science, 14*(2), 105-120.

Beaulé, A., Dascola, M., & Liu, Y. (2009). *Development and Programming of an Employment Event History Calendar in the Panel Study of Income Dynamics* (PSID Technical Paper Series No. 09-01).

Belli, R. F., Smith, L. M., Andreski, P. M., & Agrawal, S. (2007). Methodological Comparisons Between CATI Event History Calendar and Standardized Conventional Questionnaire Instruments. *Public Opinion Quarterly, 71*(4), 603-622. doi:10.1093/poq/nfm045

Belli, R. F. (1998). The Structure of Autobiographical Memory and the Event History Calendar: Potential Improvements in the Quality of Retrospective Reports in Surveys. *Memory, 6*(4), 383-406. doi:10.1080/741942610

Belli, R. F. (2000). Computerized Event History Calendar Methods: Facilitating Autobiographical Recall. *Proceedings of the Survey Research Methods Section, ASA,* 471-475.

Belli, R. F., & Callegaro, M. (2009). The emergence of calendar interviewing: A theoretical and empirical rationale. In R. F. Belli, F. P. Stafford, & D. F. Alwin (Eds.), *Calendar and time diary methods in life course research* (pp. 31-52). Thousand Oaks: SAGE Publications Inc.

Belli, R. F., Lee, E. H., Stafford, F. P., & Chou, C.-H. (2004). Calendar and question-list survey methods: Association between interviewer behaviors and data quality. *Journal of Official Statistics, 20*(2), 185.

Belli, R. F., Shay, W. L., & Stafford, F. P. (2001). Event history calendars and question list surveys: A direct comparison of interviewing methods. *Public Opinion Quarterly, 65*(1), 45-74.

Belli, R. F., Stafford, F. P., & Alwin, D. F. (Eds.). (2009). *Calendar and time diary methods in life course research*. Thousand Oaks: SAGE Publications Inc.

Brüderl, J., Hank, K., Huinink, J., Nauck, B., Neyer, F. J., Walper, S., . . . Wilhelm, B. (2015). The German family panel (pairfam): ZA5678 Data file Version 6.0.0. *GESIS Data Archive, Cologne*. doi:10.4232/pairfam.5678.6.0.0.

Brüderl, J., Hajek, K., Herzig, M., Huyer-May, B., Lenke, R., Müller, B., . . . Schumann, N. (2015). *pairfam Data Manual: Release 6.0*. Munich.

Burkhead, D., & Coder, J. (1985). *Gross Changes in Income Recipiency from the Survey of Income and Program Participation* (Proceedings of the American Statistical Association, Social Statistics Section No. 351-356). Washington, D.C.

Callegaro, M. (2007). *Seam Effects Changes Due to Modifications in Question Wording and Data Collection Strategies: A Comparison of Conventional Questionnaire and Event History Calendar Seam Effects in the PSID*. Lincoln, NE: University of Nebraska.

Callegaro, M. (2008). Seam Effects in Longitudinal Surveys. *Journal of Official Statistics, 24*(3), 387-409.

Caspi, A., Moffitt, T. E., Thornton, A., Freedman, D., Amell, J. W., Harrington, H., . . . Silva, P. A. (1996). The life history calendar: a research and clinical assessment method for collecting retrospective event-history data. *International Journal of Methods in Psychiatric Research, 6*(2), 101-114.

Chan, A. Y. (2009). *The 2008 SIPP Event History Calendar (EHC) Field Test: Respondents' Reactions to the Interview* (Statistical Research Division Research Report Series (Survey Methodology) No. 03). Washington, D.C.

Drasch, K., Kleinert, C., Matthes, B., & Ruland, M. (2016). Why Do We Collect Data on Educational Histories Over the Life Course the Way We Do ? Core Questionnaire Design Decisions in Starting Cohort 6 – Adults. In H.-P. Blossfeld, J. von Maurice, M. Bayer, & J. Skopek (Eds.), *Methodological Issues of Longitudinal Surveys: The Example of the National Educational Panel Study* (pp. 331–347). Wiesbaden: Springer Fachmedien Wiesbaden. Retrieved from http://dx.doi.org/10.1007/978-3-658-11994-2_19

Eggs, J., & Jäckle, A. (2015). Dependent Interviewing and Sub-Optimal Responding. *Survey Research Methods, 9*(1), 15. doi:10.18148/srm/2015.v9i1.5860

Eisenhower, D., Mathiowetz, N. A., & Morganstein, D. (1991). Recall error: Sources and bias reduction techniques. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement Errors in Surveys* (pp. 125-144). Hoboken: John Wiley & Sons.

Fields, J., & Callegaro, M. (2007, December). *Background and Planning for Incorporating an Event History Calendar into the Re-Engineered SIPP*. Federal Committee on Statistical Methodology Meeting,

Freedman, D., Thornton, A., Camburn, D., Alwin, D., & Young-DeMarco, L. (1988). The life history calendar: A technique for collecting retrospective data. *Sociological Methodology, 18*(1), 37-68.

Glasner, T. (2011). *Reconstructing event histories in standardized survey research: Cognitive mechanisms and aided recall techniques*. Oisterwijk: Uitgeverij BOXPress.

Glasner, T., & van der Vaart, W. (2009). Applications of calendar instruments in social surveys: a review. *Quality & Quantity, 43*(3), 333-349. doi:10.1007/s11135-007-9129-8

Hoogendoorn, A. (2004). A Questionnaire Design for Dependent Interviewing that Addresses the Problem of Cognitive Satisficing. *Journal of Official Statistics, 20*(2), 219-232. Retrieved from http://www.jos.nu/Articles/abstract.asp?article=202219

Huinink, J., Brüderl, J., Nauck, B., Walper, S., Castiglioni, L., & Feldhaus, M. (2011). Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual framework and design. *Zeitschrift für Familienforschung, 23*(1), 77-101.

Jäckle, A. (2008). *The causes of seam effects in panel surveys* (No. 2008-14).

Jäckle, A. (2009). Dependent interviewing: A framework and application to current research. In P. Lynn (Ed.), *Methodology of Longitudinal Surveys* (pp. 93-111). Chichester: Wiley.

Jäckle, A., & Eckman, S. (2016). *Is that still the same? Has that changed? On the accuracy of measuring change with dependent interviewing* (Understanding Society Working Paper Series No. 2016-06). Colchester.

Jäckle, A., & Lynn, P. (2007). Dependent interviewing and seam effects in work history data. *Journal of Official Statistics, 23*(4), 529.

Kominski, R. (1990). *The SIPP event history calendar: Aiding respondents in the dating of longitudinal events* (Survey of Income and Program Participation Working Paper Series No. 132).

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-236. doi:10.1002/acp.2350050305

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New Directions for Evaluation, 1996*(70), 29-44. doi:10.1002/ev.1033

Lynn, P., Jäckle, A., Jenkins, S. P., & Sala, E. (2012). The impact of questioning method on measurement error in panel survey measures of benefit receipt: evidence from a validation study. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 175*(1), 289–308. doi:10.1111/j.1467-985X.2011.00717.x

Lynn, P., & Sala, E. (2006). Measuring Change in Employment Characteristics: The Effects of Dependent Interviewing. *International Journal of Public Opinion Research, 18*(4), 500–509. doi:10.1093/ijpor/edl013

Moore, J. (2012). *Analysis of Recorded Interviews in the 2010 SIPP-EHC Field Test* (SEHSD Working Paper No. 2012-17).

Moore, J., Bates, N., Pascale, J., & Okon, A. (2009). Tackling seam bias through questionnaire design. In P. Lynn (Ed.), *Methodology of Longitudinal Surveys* (pp. 73-92). Chichester: Wiley.

pairfam. (2015). *The German Family Panel (pairfam), Codebook Anchor, Wave 6, 2013/2014: Release 6.0.*

Pascale, J. (2009). *Event History Calendar Field Test Field Representative Focus Group Report* (Statistical Research Division Study Series (Survey Methodology) No. 02). Washington, D.C.

Reimer, M., & Matthes, B. (2007). Collecting Event Histories with TrueTales: Techniques to Improve Autobiographical Recall Problems in Standardized Interviews. *Quality & Quantity, 41*(5), 711-735. doi:10.1007/s11135-006-9021-y

Rips, L. J., Conrad, F. G., & Fricker, S. S. (2003). Straightening the Seam Effect in Panel Surveys. *Public Opinion Quarterly, 67*(4), 522-554. doi:10.1086/378962

Sayles, H., Belli, R. F., & Serrano, E. (2010). Interviewer Variance Between Event History Calendar and Conventional Questionnaire Interviews. *Public Opinion Quarterly, 74*(1), 140–153. doi:10.1093/poq/nfp089

Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys*: Chicago: Aldine.

Trahms, A., Matthes, B., & Ruland, M. (2016). Collecting Life-Course Data in a Panel Design: Why and How We Use Proactive Dependent Interviewing. In H.-P. Blossfeld, J. von Maurice, M. Bayer, & J. Skopek (Eds.), *Methodological Issues of Longitudinal Surveys: The Example of the National Educational Panel Study* (pp. 349-366). Wiesbaden: Springer Fachmedien Wiesbaden. Retrieved from http://dx.doi.org/10.1007/978-3-658-11994-2_20

Uhrig, N., & Sala, E. (2011). When Change Matters: An Analysis of Survey Interaction in Dependent Interviewing on the British Household Panel Study. Sociological Methods & Research, 40(2), 333-366.

van der Vaart, W. (2004). The time-line as a device to enhance recall in standardized research interviews: A split ballot study. *Journal of Official Statistics, 20*(2), 301-318.

Watson, N. (2009, July). *Disentangling Overlapping Seams: The experience of the HILDA Survey.* HILDA Survey Research Conference, University of Melbourne.