**A Multi-Method Evaluation
of the Use of an Event History Calendar**

Joanne Pascale
Alice McGee*

National Centre for Social Research*
London, England

Statistical Research Division
U.S. Census Bureau
Washington, D.C. 20233

A Multi-Method Evaluation of the Use of an Event History Calendar

Joanne Pascale, US Census Bureau
Alice McGee, National Centre for Social Research

Abstract[1]

The English Longitudinal Study of Aging (ELSA) is conducted every 2 years with persons 50 years old and older and covers a range of topics, including family structure, health and disability, economic circumstances and retirement. After two waves of data collection, the research team and the study's sponsors were interested in capturing retrospective data on major life events over the respondent's entire life cycle – such as changes in residence, marital status and employment – beginning at the respondent's birth up to the present. In September, 2005, a pilot study was carried out to evaluate the use of an event history calendar (EHC) as a means of collecting these retrospective data.

In order to evaluate the EHC a number of methods were employed. First was a respondent debriefing. This was a set of questions administered to each respondent immediately after the interview and covered topics such as their general engagement with the EHC (i.e.: how much they looked at and followed along with the EHC as the interviewer filled it in), which "domains" or topic areas they chose to start with and why, whether they preferred to go forward or backward in time, the usefulness of landmarks in recall, and so on. The second evaluation method was an interviewer diary. This was a hard-copy list of questions that the interviewer filled in soon after conducting the interview, and covered some of the same topics as the respondent debriefing but from the interviewer's perspective. Third, at the end of the field period interviewers were brought together for a general debriefing on their experience using the EHC. Finally, all pilot cases were tape recorded and "quasi-transcripts" were written. These were not word-for-word transcripts but rather running accounts of the interview flow, focusing on sections where changes and key dates in life events were reported. The purpose of the quasi-transcripts was to map the respondent's "journey" through the EHC to enable analysis of actual behaviors, such as switching across domains, going forward and backward in time, use of landmarks and so on.

This paper will present results from each of these evaluation sources to address some of the gaps in our current understanding of the EHC method. We examine two themes. First is the interviewers' and respondents' experiences with the EHC – whether they felt engaged with the calendar, how useful it seemed to be as a recall aid, how easy or difficult the interviewer found it to administer and so on. The other theme is landmarks and navigating through the EHC – that is, the extent to which landmarks came into play, who (interviewers or respondents) introduced them and why, and the effect of the landmarks. We also examined preferences and behavior

---

regarding the sequencing of topic areas and the direction of reporting in time (past to present or vice versa).

1.   Introduction

The English Longitudinal Study of Aging (ELSA) is a panel study of people aged 50 and over and their partners, begun in 2002 and carried out every two years. It is funded by the National Institute of Aging and a consortium of UK government departments, and field work is conducted by the National Centre for Social Research (NatCen), based in London. Each wave covers a variety of topic areas including: (a) economic well-being (household income, employment, wealth, retirement and pensions), (b) household and family structure and the transfer of resources, (c) physical health, disability, life expectancy and cognitive function, (d) social networks, support and participation, and (e) psychological and social well-being. It is the first study in the UK to bring together and explore the full range of topics required to understand the economic, social, psychological and health elements of the aging process. ELSA's overall aim is to explore the relationships between these topic areas as people plan for, move into and progress beyond retirement. ELSA also aims to measure change in the health, social and economic circumstances of elderly people in England (Banks et al, 2006).

Prior to Wave 3 the majority of questions in the survey covered the circumstances of respondents' lives from the time they were first interviewed (respondents ranged in age from 46 up to their upper 90s at the time) until the present day. Evidence suggests that aspects of individuals' early lives have a significant impact on their health, economic circumstances and quality of life in later years (Kuh et al, 1997). Thus a decision was made to add a topical component to Wave 3 that would capture key events over the entire life course, beginning at birth. Researchers considered various approaches for collecting these data, including an event history calendar (EHC) – a data collection method used to gather retrospective data. Rather than a conventional list of survey questions, the EHC uses a calendar, with time going across the top in columns and key events of interest in the survey going down the side in rows. In contrast to a traditional linear questioning approach, the EHC method uses a series of semi-structured questions and probes to encourage respondents to report from autobiographical memory using strategies such as sequential retrieval (e.g.: "First I had Job X, then I got Job Y, then I took Job Z."), as well as "parallel" retrieval or cross-referencing one event with another (e.g.: "When I took Job Y I moved to town X."). The method has been shown to improve the accuracy and quality of retrospective reports for some topic areas (Belli et al, 2004; Belli et al, 2001).

Given the challenge before ELSA researchers – collecting data over the entire life course – and the promise of the EHC method, a pilot test was carried out to explore the feasibility of using the EHC technique to collect retrospective life history data for Wave 3. Alongside this general aim some additional methodological research questions were formulated for the pilot to address. First was interviewers' and respondents' use of and receptivity to the technique – whether they enjoyed the approach, whether the EHC tool seemed to aid in recall of events, and whether respondents felt it was important to visually follow along as the interviewer filled out

the grid. The second set of research questions had to do with the use of "landmarks". These are key events, either in the respondent's own life, known as internal landmarks (e.g.: got married, took a new job), or major world events, known as external landmarks (e.g.: JFK's assassination). Little is known about the "mechanics" of these landmarks – that is, when they are introduced as memory aids, who introduces them (the interviewer or the respondent), what interview circumstances seem to prompt the use of landmarks, what types of events trigger recall of other events, and generally how successful these landmarks are in helping respondents recall dates accurately. Anecdotal evidence suggests that external landmarks are not especially useful though the reasons for this are largely unknown. It may be that interviewers do not introduce them or do not introduce them at appropriate times, or it may be that the landmarks are not generally salient to respondents and thus do not help them recall dates in the way they are meant to. Finally, our research set out to examine the time dimension of reporting (going forwards or backwards in time), as well the sequence of topic areas asked about. To address these questions we employed four different evaluation methodologies (described in further detail below): interviewer diaries and respondent debriefings (carried out following each individual interview), interviewer debriefings (carried out at the end of data collection), and analysis of tape recorded interviews. This paper reports on the findings from the pilot test.

2.   Methods

2.1   Interviewers, Respondents and Instrument

In September 2005 a pilot test of the EHC method was carried out with four experienced NatCen interviewers and 18 respondents. The aim of the EHC pilot was to capture major life events from birth up to the present, covering the following domains: residences, siblings and parents, partners, children and grandchildren, and work. The respondents (seven men and 11 women), ranged in age from 50 to 81, with an average age of 67, and seven were already part of the ELSA panel, while 11 were newly-recruited. The instrument for the pilot test consisted of a hard-copy "life grid" – a card-stock fold-out grid 12" high by 32" long, with years going across the top in columns and domains going down the side in rows. For each domain, interviewers probed for the year of key events (such as birth and death of siblings, start and end date of jobs, etc.), and certain details (e.g.: names of siblings and employers). For siblings and parents interviewers were given a short script which contained some basic questions (e.g.: "What was the name of [your mother's] first child?", "In which year was he/she born?", "Were your parents ever separated?", "In which year were they first separated?"). However, for all other domains there was no script per se; rather, interviewers were instructed to probe for start and end dates, as well as details, based on the cues printed in the calendar itself. For example, for the residence domain, the calendar displayed: "Started/stopped living in each residence: X" and "Details: Road/Town/Country." See Figure 1 for details. Filling in the life grid took between 35 and 60 minutes, and interviews were conducted face-to-face in respondents' homes.

Figure 1: Event History Calendar Life Grid (partial)

		1900	1901	1902	1903	1904...
Age						
External event			Queen Victoria died			
What to Write						
Where lived	Started/stopped living in each residence: X					
	Details: Road/Town/Country					
Siblings and parents	Sibling birth or death: X Parents separated, divorced, died: X					
	Siblings born, died, eg: Jane born, Fred died Parents' separations: Parents sep & Sep. end Parents divorced: Parents div. Parents died: Mother died, Father died					
Partners	When married or started living with partner: X When stopped living together or ended rel.: X					
	Marriages, eg: Mar. John Living together (not married) eg: Lived w. Ian Separations: Sep. & Sep. Ended Divorced: Div. Widowed: Widow					
Children and Grand-children	Children born or died: X First and last grandchild born: X					
	Child born, died, eg: Sarah born, James died Grandchildren: 1st grandch., Last grandch					
Work	Start and end of each job: X					
	Details: Job title or employer					
Other key events	Important events not reported above: X					
	Details of event					

2.2   Interviewer training

Interviewers were given a one-day training on the content and use of the EHC instrument. Topics included certain details of the domain areas (e.g.: that only residences lived in for 6 months or more were to be recorded), as well as instructions on how to introduce the life grid and its features to respondents and how to fill out the grid. This involved explaining the layout, showing the years going across the top (along with respondent's age), domains down the side, and the "external events" – national or world events that were pre-printed on the calendar for use as recall aids. After orienting the respondent to the EHC and to the task, interviewers were instructed to ask respondents which topic they'd prefer to start with. The training also covered

the scripted questions in the siblings and parents section, mentioned above, but otherwise the actual questions asked, the order of the topics and the sequence in time (forward or backwards) was left open to the interviewer and respondent to negotiate.

2.3   Evaluation Methods

As mentioned above, there were four evaluation methods employed:

A.   Interviewer Diary: this was a hard copy set of questions that interviewers filled out soon after each interview and covered the following topic areas:

1.   Setting: whether the interviewer was able to display the life grid in front of the respondent so that both could follow along as it was being filled in.
2.   Other household members: whether the respondent was interviewed alone or whether other household members were present, and whether they participated in the interview.
3.   Topic selection: which topic they started with and why.
4.   Changing domains: whether they stayed with the same topic from the respondent′s birth continuing up to the present, or whether they switched to another topic.
5.   Time sequence: whether they went forwards or backwards in time.
6.   Correcting mistakes: whether they had to go back and change the dates of events, and if so what the circumstances were.

Diaries were collected for all 18 cases.

B.   Respondent Debriefing: this was a semi-structured interview conducted with the respondent immediately after the main interview on the following themes:

1.   Setting: were respondents able to see the life grid as interviewers were filling it in, and if so, did they find it helpful.
2.   Topic selection: whether the topic they started with was the most useful; if not which topic would have been more helpful.
3.   Time sequence: whether they had a preference for going forward or backward in time; which they think was more useful.
4.   Recall: whether certain topics were more difficult than others in terms of remembering dates.
5.   Life Grid as a Recall Aid: whether identifying and writing down the year of one event helped recall the year of other events; whether external landmarks were helpful.
6.   Accuracy: the degree to which respondents felt their answers were accurate, and whether this varied by topic area.
7.   Emotional: whether any part of the interview was upsetting; whether respondents felt any questions were too intrusive or private.

The respondent debriefings were meant to be conducted and tape recorded for all 18 cases but only 10 cases were available on tape.

C.   Interviewer Debriefing: a semi-structured discussion with interviewers at the conclusion of data collection to gather interviewers′ feedback on the interviewing experience. All four interviewers were present for the debriefing.

D.   Tape Recordings: Interviews were tape recorded for later analysis. Due to timing and
        budget constraints, rather than full transcripts "quasi-transcripts" were produced. These
        were not word-for-word verbatim texts but rather a running account of the respondent-
        interviewer interactions, focusing on discussions which involved dating events, such as:
        what events got reported, when landmarks were used, how they were introduced and
        who introduced them, and the circumstances leading up to these reports. Thus certain
        parts of discussions (such as whether a training position "counted" as a job) were left
        out. Though the aim was to produce quasi-transcripts for all 18 cases, only 13 full cases
        were produced (and in one additional case the quasi-transcript was produced for the
        residence section but then the tape became inaudible).

The interviewer diaries, respondent debriefings and quasi-transcripts were analysed using a
content analysis approach based on "Framework," an analytic tool developed by the Qualitative
Research Unit at NatCen. A matrix was set up with various themes and sub-themes going
across the top in columns and individual cases going down the side in rows. Under each sub-
theme a summary was made of what had occurred for each particular case. Thus data could be
read horizontally as a complete case record for an individual, or vertically by theme, looking
across all cases. Results from the interviewer debriefing were typed up into a short report.

3.   Results

To the extent that the same themes were addressed by multiple evaluation methods, results
have been triangulated across the methods. In some cases, however, themes were addressed by
only one method.

3.1   Interview Setting

According to the interview diaries, respondent debriefings, and interview debriefing, most of
the time interviewers sat next to the respondent so that they could both see the life grid as it
was being filled in. However, according to the respondent debriefings, most respondents said it
was not helpful to see the life grid or follow along as the interviewer filled it out (though one
respondent said it was useful to watch the process), and interviewers reported during the
debriefing that while some respondents looked on, others did not want to. These results are
rather mixed, and somewhat compromised by the fact debriefings were available for only 10 of
the 18 respondents. Nevertheless, there is no overwhelming evidence that respondents have a
strong desire or need to visually follow along as the interviewer fills out the life grid.

3.2   Other household members

According to the interviewer diaries, in most cases (16) respondents were interviewed alone,
but in two cases the respondent's wife, mother or mother-in-law interjected a few dates. Based
on the quasi-transcripts, one case in particular involved a fair amount of participation from the
respondent's wife, particularly when the respondent had difficulty in the residence section. This
may have had a non-trivial effect on the quality of the data. Given the relatively low prevalence

of participation by other household members, however, a study involving a much greater number of respondents would be needed in order to draw conclusions about the nature and implications of other household members' participation in the interview.

3.3   Topic selection

The interviewer debriefing, diaries and quasi-transcripts indicated that interviewers usually followed the default order of topics – that is: residences, siblings, parents, partners, children and then work. In the diaries, interviewers offered a variety of reasons for beginning with residence, noting that it was not a contentious subject, instructions were easy to follow, it was generally easy for respondents to remember, and that often respondents had no preference but when they did, they chose the residence section as the starting topic. One respondent, however, preferred to start with children because she said those dates are easily and accurately remembered and thus good to relate other events to. The quasi-transcripts further indicated that in six of the 13 cases available, the interviewer began with residence without asking the respondent for their preference, in four cases the interviewer gave the respondent a choice, and in three cases the interviewer suggested a topic in a leading way. Among the cases where respondents were given a choice, only one respondent voiced a strong opinion, noting that she felt confident of her ability to recall residence history. In the other cases respondents did not indicate strong preferences or opinions, noting "I'll do exactly as you want" and "I don't care." In the respondent debriefings, subjects were about evenly split across three opinions: some said the default order of topics was fine, some said it would make no difference, and some said it was very useful to start with residence, noting "it brought back other memories" and "other events fitted into it."

Also according to the diaries, in roughly half the cases interviewers said they stayed with the same topic from start to finish, while in the other half they said they stayed with the same topic but entered dates for other domains if they were mentioned along the way. In one case the interviewer said she stayed with the same topic mostly due to unfamiliarity with the instrument but after having administered several interviews she would fill in other events as they were mentioned. The quasi-transcripts indicated that once a particular topic was begun, the interviewer stayed with it up to the present day. And there were several indications that, when respondents did mention key events from other domains in passing (e.g.: I moved to Place X when I got married), interviewers entered those dates in the other domains. This type of behavior, however, would not necessarily be evident on the tape or the quasi-transcript since interviewers may well have made these recordings without verbalizing to the respondent what they were doing.

3.4   Time sequence

According to the interviewer diaries, in all cases they went forward in time, but in seven cases they went backward in time part way through the interview, usually to clarify, narrow down or correct a date. During the debriefing, half the respondents said they preferred going forward in time, and one noted that going backward in time would have been problematic. Four

respondents were indifferent to the time sequence, and one said that on hindsight, going backward in time would have been easier.

The quasi-transcripts indicated that indeed the interviewers usually started from the past and moved forward in time. In the residence section, for example, they usually started by asking where the respondent was born and moved forward in time asking about each sequential residence. In the siblings, partner, children and work sections interviewers generally asked about the first sibling, partner, child or job and then the second, third, and so on. In terms of switching gears and going backward in time, the quasi-transcripts generally corroborated the interviewer diary findings, though only three instances of going backward in time were evidenced (this could well be a missing data issue, as quasi-transcripts were available for only 13 cases, while diaries were available for all 18 cases). In two of these three cases there was explicit evidence that the respondent was having difficulty (one was reporting the fifth of seven residences (age 28 at the time); the other was reporting the 11th of 13 places he'd lived (age 40 at the time)) and the interviewer suggested switching gears and starting with the present and going backward in time. In the third case there was no evidence of respondent difficulty but the interviewer made the suggestion to switch. In all three cases the respondents seemed to have no trouble reporting events going backward in time, and in the two problematic cases the tactic seemed to aid the respondent in dating the events.

3.5   Recall and the EHC as a Memory Aid

Respondents were somewhat inconsistent in their debriefings regarding the ease or difficulty of recalling events in general. Five said there were no topics they found particularly difficult, while one found the interview challenging in general. Four respondents said they found the work section easy, two said residence, and one said children/grandchildren was easy. However, five respondents said the residence section was difficult, but three of these respondents said it was reporting on postcodes and house numbers that was particularly difficult, not dates. Not surprisingly, three respondents noted that more recent events were easier to remember than those in the more distant past. One respondent noted that in general it was harder to recall years than events; that is, he could remember the sequence of events but had a harder time reporting the year the events happened. Regarding the usefulness of the EHC per se, five respondents said they didn't think the life grid was helpful, two said it was (one noting that some events – particularly marriage, children and house moves – helped trigger recall of other events).

During their debriefing, interviewers noted that respondents generally could recall dates of events accurately, and felt the EHC helped this process. They discussed cross-referencing dates across domains to jog the respondent's memory or to check dates, noting that dates of births, deaths and marriages were generally easy to remember, and that jobs were harder to remember. They also noted that the residence section often took a long time but respondents could generally remember when they moved residences, and this section was generally very helpful for dating events in other domains.

The quasi-transcripts indicate heavy use of landmarks and cross-referencing across domains

and suggest that the life grid approach may have been more useful than was suggested by the respondent debriefing (results on landmark usage are discussed in more detail below).

3.6   Accuracy and Correcting Mistakes

In the respondent debriefings all but one respondent said they were very confident in their answers. The one respondent who differed said she was "not terribly"confident about the residence and parents sections. In the diaries, interviewers noted several instances of correcting mistakes in dates that had been entered earlier based on subsequent questions about other domains. For example, one respondent changed start and end dates of her job after relating it to her children and where she was living at the time, another changed the birth year of a grandchild after reporting the year she changed jobs, and another changed the birth year of her child due to dates reported in the residence section. The quasi-transcripts also indicated several instances of interviewers and respondents catching and correcting mistakes when cross-referencing dates across domains.

3.7   Emotional Issues

In the respondent debriefings six subjects said the interview was not upsetting or difficult emotionally, while the others noted particular events or periods of life they found somewhat difficult, such as questions on childhood, parents separating and divorcing, and discussing lost family members.

3.8   Landmarks

According to both the interviewer diaries and debriefing, external landmarks helped in a few cases but mostly added interest and helped lighten the mood. The quasi-transcripts, however, offer a more rich, complete set of data on how often landmarks came into play, and the circumstances surrounding their use. The remainder of this section is dedicated to those results.

3.8.1   Frequency of Use

Overall, across all interviewers, respondents and topic areas, a total of 314 landmarks were used (both internal and external). Respondents used landmarks about twice as often as interviewers; in total respondents used 221 landmarks and interviewers used 93. Both respondents and interviewers used internal landmarks more frequently than external landmarks, but respondents were much more likely to use internal landmarks than interviewers. Among respondents, 86% of landmarks used (189 out of 221) were internal, while among interviewers 57% (53 out of 93) were internal.

Table 1: Frequency of Landmark Use

	Internal		External		Totals				
	Resp	**Iwer**	**Resp**	**Iwer**	**Internal**	**External**	**Resp**	**Iwer**	**Overall**
Residence	69	7	28	29	76	57	97	36	133
Siblings	9	1	0	0	10	0	9	1	10
Parents	14	0	1	8	14	9	15	8	23
Partners	13	3	0	1	16	1	13	4	17
Children	16	1	0	0	17	0	16	1	17
Work	68	41	3	2	109	5	71	43	114
Total	189	53	32	40	242	72	221	93	314

3.8.2   Landmark Use by Topic Area

Overall, most of the internal landmarks (45%) were used in the work section, followed by the residence section (31%), children and partners (7% each), parents (6%) and siblings (4%) (see Table 2). This overall pattern was driven mostly by interviewers. Most of the internal landmarks they used (77%) were in the work section, followed by residence (13%), followed by partners (6%). Among respondents, their use of internal landmarks was about evenly split between residence (37%) and work (36%), with the remaining landmarks being fairly evenly spread across the children section (8%), parents (7%), partners (7%) and siblings (5%). Among the external landmarks, the vast majority (79%) were observed in the residence section (see Table 3). Another 13% were used in the parents section, and 7% were used in the work section.

This general pattern held across interviewers and respondents, but respondents were somewhat more likely than interviewers to use external landmarks in the residence section than other sections. Among respondents, 88% of their external landmarks were used in the residence section and 9% were used in the work section. Among interviewers, 73% of their external landmarks were used in the residence section and 20% were used in the parents section.

Table 2: Internal Landmark Use by Topic Area

	Residence		Siblings		Parents		Partners		Children		Work		Total
	n	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	
Resp	69	37	9	5	14	7	13	7	16	8	68	36	189
Iwer	7	13	1	2	0	0	3	6	1	2	41	77	53
Total	76	31	10	4	14	6	16	7	17	7	109	45	242

Table 3: External Landmark Use by Topic Area

	Residence		Siblings		Parents		Partners		Children		Work		Total
	n	%	n	%	n	%	n	%	n	%	n	%	
Resp	28	88	0	0	1	3	0	0	0	0	3	9	32
Iwer	29	73	0	0	8	20	1	3	0	0	2	5	40
Total	57	79	0	0	9	13	1	1	0	0	5	7	72

The average number of "turns" or exchanges between respondents and interviewers varied by topic area. By far the greatest number of turns was observed in the residence section (23 turns per case on average), while for all the other five sections the average number of turns ranged from 5 to 9. The number of changes or events also varied by topic area, and again the greatest number of changes was in the residence section (an average of 10 residences per case). The work section was also high (7 jobs per case), while the average number of children was 3, and the average number of siblings and partners was one each. Given this variation, another way to measure the use of landmarks by topic area is a "landmark per turn" ratio – that is, the sum of all landmarks used divided by the total number of turns taken by topic area. This measure shows an overall ratio of 0.31 landmarks per turn, with the residence section the highest (0.42), followed by work (0.36), then parents (0.29). Siblings, partners and children ranged from 0.14 to 0.16. This suggests that landmarks come into play more often in the residence and work sections, even while controlling for higher activity (i.e.: more interaction between respondents and interviewers, and more reports of events and dates) in those sections in general.

3.8.3 Landmark Use by Type

Internal landmarks often referred to domains already within the scope of topics in the calendar. For example, in the residence section respondents would mention getting married when they moved to a certain residence. Most of the internal landmarks (184 out of 240, or 77%) referred to domains within the calendar, while the other 56 (23%) referred to events outside the scope of the calendar. Among these 184 landmarks used referring to topics within the calendar, the most frequent type of landmark used (32%) was age – the respondent's own age or the age of a family member. For example, respondents would say "I moved when I was 21."

Table 4: Landmark Use by Type

	Residence	Siblings	Parents	Partners	Children	Work	Total
Age	17	3	4	3	11	21	59
Residence	0	1	1	7	1	19	29
Siblings	3	0	0	0	0	0	3
Parents	8	0	0	2	1	0	11
Partners	14	0	0	0	1	15	30
Children	7	0	3	1	1	36	48
Work	4	0	0	0	0	0	4
Other	18	6	6	2	2	22	56
Total	71	10	14	15	17	113	240

Excluding age from landmarks within the calendar, the most frequent landmark used overall was children (38%), followed by partners and residence (24% and 23% respectively), and parents (9%). However, this ranking varied by topic area. The work section followed the overall pattern fairly closely; the most frequent landmark used was children (51%), then residence (27%), then partners (21%). For the residence section, however, the most frequent landmark was partners (39%), followed by parents (22%), then children (19%). And in the partners section the most frequent landmark used was residence (70%).

Among the landmarks outside of the scope of the calendar, there was quite a range of subject matter. The types of landmarks mentioned more than once were personal events related to the war (e.g.: evacuated, house was bombed), going to school, buying a house, going abroad, hospitalizations/illnesses/injuries, going into the military, and being made redundant.

3.8.4   Context and Effect of Landmark Use

In order to better understand the circumstances under which landmarks were introduced, and their effects, a coding scheme was developed which was based around each instance that an interviewer attempted to capture a year that a particular event occurred (e.g.: In what year did you move to Place X?; When did that job end?; In what year did you get married?). The codes reflected the respondent's first response to the request (provided a year, reported that it occurred when a previously-reported parallel event occurred, etc.), and for some of these main codes, sub-codes were developed to capture the followup interaction (e.g.: interviewer offered the date that the parallel event occurred).

In total, across all cases and topic areas, there were 372 attempts by interviewers to capture the year that an event occurred. In most cases (357, or 96% of the time) interviewers began the

exchange by asking what year a particular event occurred. For most of the remaining 4% of the cases, interviewers suggested a year based on previously-reported information (results reported below). Among the 357 cases, a total of six main codes were used that reflected the respondent's first response to the interviewer's request (see Table 5):

A: Year: the respondent provided a year straightaway, without any further elaboration

B. Parallel: the respondent provided information about a parallel event but not a year (e.g.: I lived there until I got married)

C. Year + LM: the respondent provided both a year and also mentioned landmark (LM) in passing (e.g.: I moved to place X in 1945, the year I got married)

D. Difficulty: the respondent expressed difficulty in recalling the year, uncertainty or gave a range of years

E. Narrative: the respondent gave the year or duration in a spontaneous, running narrative (e.g.: I moved to place X in 1945, then to place Y for 8 years, then to place Z for 3 years)

F. Duration: the respondent provided the duration but not the year (e.g.: I lived there for 12 years)

G. Other

Table 5: Respondent's First Response

Respondent's First Response	Residence		Siblings		Parents		Partners		Children		Work		Total	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
A. Year	41	29	11	69	10	33	15	45	17	53	31	29	125	35
B. Parallel	16	12	3	19	1	3	7	21	5	16	29	27	61	17
C. Year + LM	20	14	1	6	9	30	7	21	6	19	15	14	58	16
D. Difficulty	22	16	1	6	8	27	4	12	3	9	17	16	55	15
E. Narrative	24	17	0	0	1	3	0	0	1	3	8	7	34	10
F. Duration	14	10	0	0	0	0	0	0	0	0	7	7	21	6
G. Other	2	1	0	0	1	3	0	0	0	0	0	0	3	1
Total	139	100	16	100	30	100	33	100	32	100	107	100	357	100

Overall these results suggest that respondents did not have a terribly difficult time reporting dates. In most cases (35%) respondents provided a year straight away (code A), and in another 16% of cases they provided a year as well as a landmark (code C). However, there was some variation by topic area. In both the residence and work sections, respondents either provided a year, or the year and a landmark, 43% of the time, while in the other domains the figure ranged from 66% to 75%. This suggests that respondents may not have dates as ready at hand for the residence and work sections, and complements findings discussed above indicating heavier use

of landmarks for these two topic areas compared to the others.

When respondents did not report a year outright they often provided indirect information – either parallel events or duration (codes B, D and F) that usually led to the year – a total of 33% of cases, summing across these three codes. In another 15% of cases, though, respondents expressed some type of difficulty or uncertainty or imprecision.

For the 51% of cases where respondents clearly provided a year, there was generally no followup behavior of interest. For the other two sets of responses – providing indirect information leading to a year, or expressing difficulty – followup interviewer-respondent interaction was of considerable interest and appropriate sub-codes were developed. We first discuss cases where respondents expressed difficulty, along with followup behaviors, then we examine cases where respondents provided indirect information leading to a year, again with followup behaviors.

Among the 15% of respondents who expressed some kind of difficulty in their initial response, there was some variation by topic area, with the parents section indicating the highest level of difficulty (27%), followed by the residence and work section (16% each) (see Table 5). In terms of followup for these cases, most of the time interviewers either offered a landmark (53% of the time) or they suggested a year, based on previously-reported information (22% of the time) (see Table 6). When interviewers offered landmarks, most respondents (20 of the 29 cases) used the landmark to help date the event, and in 3 additional cases the landmark helped them catch and correct a mistake they'd made earlier. Thus, in 23 of the 29 cases where landmarks were offered (79%), they helped respondents date the event. In the other 6 cases (21%) respondents said the landmarks didn't help. When interviewers suggested a year most of the time respondents accepted it, though in some cases respondents rejected and recalculated the year, and in some cases they didn't confirm or dispute it.

There was some evidence of missed opportunities on the part of interviewers in helping respondents date events. In 20% of cases respondents, as part of their first response to the interviewer's request for a year, expressed difficulty but continued on a spontaneous narrative where they tried to identify the year, usually using landmarks, and finally did produce a year with no further interaction with the interviewer. In some of these instances it is arguable that it may have been helpful if interviewers intervened and/or confirmed the year. And in another 4% of cases interviewers used generic probes to narrow down the date, rather than offering a landmark that may have been helpful to respondents.

Table 6: Interviewer's First Response to Respondent Expressing Difficulty (code D from Table 5)

	n	%
A. Interviewer offered landmarks	29	53
and they were used to help date event	20	
and respondent said they didn't help, then went on to make best estimate of year	6	
and respondent/interviewer caught and corrected a mistake	3	
B. Interviewer suggested a year based on previously-reported information	12	22
and respondent accepted suggestion	7	
and respondent rejected the suggestion; respondent and interviewer discussed, sometimes using landmarks, and calculated year	3	
and respondent did not confirm or dispute it	2	
C. Respondent gave spontaneous narrative, sometimes using landmarks (e.g.:, after expressing difficulty: "I think I stayed at place X till the early eighties. I think till about 1983") and interviewer coded without any probing or followup	11	20
D. Interviewer used generic probes (e.g.: "What would be your best guess" or "Would that be closer to 1971 or 1972") but did not use landmarks	2	4
E. Respondent asked when an event occurred; interviewer gave date of landmark which was used to help date event	1	2
Total	55	100

Turning now to the 33% of cases where indirect information was provided, we wanted to examine the extent to which interviewers cross-checked dates of events in other domains, conducted simple calculations (regarding duration) and fed back this information to respondents for confirmation. Thus followup codes were developed to reflect the interviewer's response to the respondent's first response, and cases coded B, E and F were combined (see Table 7).

In most cases (53%) interviewers took the information respondents gave them and calculated a year, or they probed for a year if necessary. For example, in some cases when respondents gave a parallel event, the date of the parallel event had not yet been reported (e.g.: if the interview started with the residence section and the respondent reported that they moved to Place X when they got married, the interviewer had to probe for the year married). In another 43% of cases, however, interviewers made no verbal response to the respondent's indirect report. For example, if the respondent lived at Place X in 1945 and said they lived there for 10 years before moving to Place Y, given no verbal evidence from the interviewer, we assume he/she did the calculation and simply coded Place Y as 1955. In another 3% of cases interviewers did not suggest a year but rather they offered landmarks that may have helped confirm the year the

event occurred.

Table 7: Interviewer's First Response to Respondent (codes B, E and F from Table 5)

	n	%
Interviewer suggested or probed for a year and	62	53
Respondent accepted the suggestion or gave year (if interviewer had probed for but not suggested a year)	46	
Respondent rejected the suggestion; respondent and interviewer discussed, sometimes using landmarks, and respondent recalculated year	9	
Respondent did not confirm or dispute the suggestion	7	
Interviewer did not verbally followup but presumably did the calculation and coded the year of the event	50	43
Interviewer offered landmarks (but not a year)	4	3
Total	116	100

For the 53% of cases where interviewers suggested or probed for a year, a third set of followup codes was developed to capture the respondent's reaction to the interviewer's probing. For the most part (46 of 62 cases, or 74% of the time) respondents accepted the interviewer's suggestion, or they provided a year (if interviewers had probed). However, in 9 cases (15% of the time) respondents did not accept the suggestion but rather, a discussion ensued in which both interviewers and respondents went back to previously-reported information, and sometimes introduced new information, in order to calculate the correct year. In another 7 cases of the cases (11%), however, respondents neither confirmed nor disputed the interviewer's suggestion.

As noted at the beginning of Section 3.8.4, in 15 of the 372 cases interviewers began the exchange with respondents not by asking for a year but using some other approach. Most of the time (13 cases) they began by suggesting a year, based on previously-reported information (e.g.: "So was it in 1958 that you got married?"). And among these 13 cases, most of the time (8 cases) respondents accepted the suggestion, in three cases they rejected it and recalculated a different year, and in two cases the outcome was unclear.

Finally, in some cases, even though a year was provided for a given event, the interviewer followed up by giving another landmark (e.g.: "So you moved to place X in 1971 – the year of the Suez Canal?"). Of the 372 years reported, this type of confirmation was observed for 30 cases (or 8%). Among these 30 cases, most of the time (15 cases) respondents accepted the confirmation, in 6 cases they said they had no knowledge of the event, or didn't know when it happened, in another 6 cases they simply made no comment, and in 3 cases they corrected a previous report.

16

4.   Summary and Conclusions

In terms of administration of the EHC, there is no strong evidence that respondents want or need to closely follow along visually to what the interviewer is doing. Regarding topic selection, again there was no strong evidence that respondents had preferences, and when given a choice they seemed rather indifferent. However, considering actual behavior observed in the quasi-transcripts it is possible to make some judgments about which topic areas presented the fewest challenges to respondents. In particular, the siblings and children sections showed the highest levels of providing a year (or a year along with a landmark) straightaway in response to the interviewer's request for a year, and these two sections also exhibited the lowest levels of difficulty expressed by respondents. And finally, in terms of the frequency with which events from various domains were used to help date events in other domains, children was the most common type of landmark used (after age). These results suggest that it may be strategic to sequence the domains by asking about children first, followed perhaps by siblings, to maximize the EHC's potential as a recall aid. However, there are some caveats. First, not all respondents have children. Second, asking about children first by definition begins the interview in mid-life, and forfeits the opportunity to begin at birth with a certain domain and continue sequentially up to the present, covering the respondent's entire life span. It is unknown whether this would have a negative effect on reporting overall.

With regard to time sequence, there was no compelling evidence of respondent difficulty either going forward to the present, or from the present backward. While the vast majority of cases moved forward in time this is likely due to interviewers' general tendencies to start with the past and move sequentially forward. While relatively infrequent, there were instances of switching gears and moving backward in time when a problem was encountered. This suggests that keeping the EHC flexible on the time sequence may have benefits in a non-trivial number of cases. It should be noted, however, that these results may apply only to surveys seeking reporting over such a long timeline – more than 90 years in some cases. For surveys with a shorter reference period the recall strategy in terms of time sequencing may be quite different.

Though respondents appeared somewhat indifferent about the EHC's role as an aid to recall there was, in fact, very heavy reliance on landmarks to date events, most of those landmarks were used by respondents, and most of the landmarks they used were internal. This suggests that cross-referencing parallel events does indeed get used as a somewhat instinctual strategy and complements straight-out recall of events. Interviewers also exhibited heavy use of landmarks, both internal and external, frequently offering parallel events or suggesting years when needed to ascertain the year, but also to confirm or verify a date. While respondents often accepted these suggestions from interviewers, there was a non-trivial rate of rejecting the suggestion and recalculating the correct year. There were also more than a few instances of respondents catching and correcting mistakes based on the interviewer's introduction of landmarks. These are welcome finding and suggest that respondents do not simply acquiesce to interviewers' suggestions but are in fact rather engaged in the reporting process.

Overall these results suggest that respondents did not have very much difficulty with the

interview in general, that interviewers implemented the tool appropriately (introducing landmarks frequently and suggesting appropriate years when needed), that respondents also introduced landmarks often, and that the landmarks helped date events. While there was some evidence of missed opportunities in terms of using landmarks to help respondents when they encountered some difficulty, the prevalence of these missed opportunities was fairly low and with more training and practice the level of missed opportunities is very likely to come down.

One important limitation, of course, is that these results are based on only 18 cases. The process of developing the code frame for examining the quasi-transcripts, however, along with the other methods employed to evaluate the EHC, proved valuable and will be used to examine the next version of the EHC used in the production ELSA instrument, discussed below.

5.   Post-Pilot Work

5.1   Development of a Production EHC Instrument

Following this pilot test a general decision was made to incorporate the EHC method into production Wave 3 interviews, with some modifications. First, regarding content, the production ELSA dropped the siblings and parents section and added questions on childhood and adult illnesses. Furthermore, additional detail on residences and work was collected. For residences, questions were asked about how many bedrooms and what kind of facilities there were, and who else lived at the residence where the respondent lived at age 10. For work, data on pay and taxes were collected. Finally, in the children section, rather than collect dates of birth "from scratch," data was fed forward using dependent interviewing and respondents were asked to verify the information. Another important difference between the pilot and production instruments was the mode. A decision was made to use an automated instrument (versus a hard copy design), thus a research phase of refining and adapting the EHC design within a CAPI environment was undertaken (see Hacker et al, 2007, for details).

The first test in this research phase was exploring a "calendar-focused" instrument. A flexible version of the calendar was produced which incorporated what were deemed the principle elements of the EHC, namely: being able to enter events whenever they were raised; the ability to move backwards or forwards in time; allowing respondents to choose the sequence of topic domains being discussed; and being able to revise and edit information with ease. Using a mouse, interviewers would click in the appropriate domain and year where a particular event had been reported, and the mouse click would open a box containing a series of questions regarding this event. This approach was tested by three interviewers and three respondents. A number of problems relating to the flexibility of the CAPI interview became apparent during this test. Interviewers, who were trained in traditional linear questioning models, found navigating this unstructured tool very challenging. For example, it was not possible to see the entire grid on the screen at any one time; the interviewer had to scroll to the right to view later years. Furthermore, events could be missed out where they were not recalled using any time or topic structure and this in turn made it difficult to identify and verify gaps and overlaps. Interviewers also found using this flexible EHC technically challenging and this in turn meant they (a) might not use all the functions

available to them or that (b) they would find it time consuming and difficult if they attempted to. Further interviewer training may have helped improve confidence and skills in using the flexible EHC but given the range of problems, and the scale of the planned production data collection effort, a priority was put on developing a more user-friendly tool. Evidence from the pilot also suggested that some elements of flexibility, such as sequencing of topic domains, may not be critical or necessary from the respondent's or interviewer's perspective – at least not enough to warrant the complex programming required.

Based on these findings the decision was taken to move to a more structured design incorporating elements of a standardized interview as well as the principles behind the EHC method. This design, called a "dual-focused" instrument, meant that the interviewer worked through a much more structured questionnaire, from start to finish, with questions moving from the past forward to the present within each topic. The new design was tested in two stages. First was a pilot test with five interviewers and 28 respondents. Findings showed this new dual-focused instrument to be too structured and so further alterations were made to include more of the original flexible features. For example, a function was incorporated which allowed interviewers to move between the six topic domains, though they always began at the start of that domain. For instance, if the interviewer had begun with the children and partner topics and mid-way through the respondent had begun talking about residences, the interviewer would have been able to skip forward to the residence section but would have had to start at the beginning of that section (i.e. at the respondent's birth) and move forward in time. A second function was added that allowed the insertion of additional events where the respondent remembered later on in the interview that they had left something out (e.g. a job they had only had briefly and had forgotten initially). Shortcut keys were also introduced to allow interviewers to make changes at a particular year; that is, the interviewer could use the shortcut keys to go straight to a given year in order to enter data, rather than only being allowed to enter data as a response to a scripted question.

A second test was carried out with 11 interviewers and 59 respondents in order to explore these alterations. Following this test minor changes were made, which largely involved refining and tweaking rather than a further overhaul of the entire structure of the EHC. For instance, changes were made to improve and tighten up question wording. And the function that allowed interviewers to insert additional events (where they had initially been forgotten) was not taken forward as interviewers at the first test found this overly complex and difficult to use. The end result was a less flexible interview, now falling some way between a traditional standardized questionnaire and an EHC interview. This final version of the dual-focused calendar was used for mainstage production wave 3 interviews between February and November 2007, in which more than 250 interviewers conducted interviews with approximately 9,000 ELSA respondents.

5.2   Evaluation of Production EHC Interviews

The basic design elements of the evaluation used in the EHC pilot phase in 2005 were taken forward and applied to the production interviews in 2007. Thirty of the 250 interviewers were selected at random to take part in this evaluation, and for each interviewer eight cases were selected at random. Of these eight cases, interviewers were asked to choose four to audio

record. In total 124 interviews were recorded (as some interviewers recorded more than four cases) but for reasons not yet determined, sound files were available for only 100 of these interviews. As mentioned above, four evaluation methods will be used:

A. **Interviewer Questionnaire**: this was a hard copy self-administered questionnaire that interviewers completed soon after conducting each interview. It covered the following topic areas:
   1. Respondent's Interaction with the Calendar: whether and how often the respondent looked at the calendar on the screen, which topics they followed along for and reasons for not looking at the screen. Interviewers were also asked whether the respondent had indicated that the calendar helped them to date events.
   2. Topic Sequence: the order in which the interviewer had asked the questions and whether, in hindsight, this was the best order for that particular respondent. This section also asked whether it was problematic to ask questions about a whole topic domain before moving to the next and also whether starting with the past and moving forwards in time was a problem.
   3. Referring to Other Events to Help Respondent Remember: this section focused on the use of internal and external landmarks and whether and how these helped to date events.
   4. Memory and Recall: whether there were particular topic domains that were difficult or easy for respondents to remember and why.
   5. Emotional Issues: whether respondents had found parts of the interview upsetting and their general emotional state at the time of interview
   6. Miscellaneous: whether the respondent was interviewed alone or that another person was present, whether the respondent had any mental or cognitive impairments and whether they had appeared to enjoy the interview and if there were any areas they had liked or disliked in particular.

Interviewer Questionnaires were filled in for 124 interviews.

B. **Respondent Debriefing Questionnaire**: this was a semi-structured PAPI interview conducted with the respondent immediately after the main interview on the following themes:
   1. General reaction to calendar: how interesting and enjoyable the respondent had found the interview, how helpful the calendar was in aiding the recall of events, whether and how often the respondent had looked at the screen and whether this was helpful.
   2. Topic and Time Sequence: which was the most useful topic domain to begin with and whether the respondent wanted to stay with that topic or shift to others, whether starting in the past and moving forward to the present was helpful in remembering dates.
   3. Referring to Other Events to Help Remember: whether internal or external landmarks were helpful in assisting with dating events and whether there were any topic domains that were particularly difficult or easy

4. Anything else: respondents were also asked if there was anything they especially liked or disliked about the interview and whether they could think of ways to improve the calendar.

Respondent Debriefing Questionnaires were filled in for 122 interviews.

C. **Interviewer Debriefing**: two semi-structured discussions were held following fieldwork with the 30 interviewers who took part in this evaluation. These discussions gathered feedback on the overall interviewing experience.

D. **Behaviour coding**: 100 interviews were audio recorded using CARI software, and a behaviour coding scheme was adapted from the scheme designed and implemented by Bob Belli (2004). The adapted code scheme focuses mainly on the use of landmarks, whether they help respondents date events, who introduces them, how, when and why they are introduced. This code scheme is in the process of being applied to the 100 interviews within Blaise.

Results from all four evaluation sources described above will be triangulated and analyzed in and of themselves, and in light of the pilot findings.

REFERENCES

Banks, J., Breeze, E., Lessof, C., Nazroo, J. (eds) (2006) "Retirement, health and relationships of the older population in England: the 2004 English Longitudinal Study of Aging (Wave 2)" Institute for Fiscal Studies. http://www.ifs.org.uk/elsa/

Belli, R.F., Lee, E.H., Stafford, F.P. and Chou, C. (2004) "Calendar and Question-List Survey Methods: Association between Interviewer Behaviours and Data Quality" Journal of Official Statistics, Vol 20, No. 2, pp. 185-218

Belli, R.F., Shay. W.L. and Stafford, F.P. (2001) "Event History Calendars and Question List Surveys: A Direct Comparison of Interviewing Methods" Public Opinion Quarterly, 65: 45-74 (2001)

Hacker, E., Cox, K., Lessof, C. and Miceli, C. (2007) "Retrospective Data Collection." Internal report, National Centre for Social Research, unpublished.

Kuh D., Power C., Blane D. and Bartley M. (1997) Chapter 8: Social pathways between childhood and adult health. In Kuh D, Ben-Shlomo Y, (Eds.) A lifecourse approach to chronic disease epidemiology: tracing the origins of ill-health from early to adult life. Oxford: Oxford University Press.