The Use of Landmark Events in EHC-Interviews to Enhance Recall Accuracy

Wander van der Vaart[1,2], Tina Glasner[3] and Robert F. Belli[4]

[1] University for Humanistics, Kromme Nieuwegracht 29, 3512 HD Utrecht, Netherlands
[2] GGZ inGeest / VU university medical center - Institute of psychiatry and mental health, A.J. Ernststraat 887, 1081 HL Amsterdam, Netherlands
[3] Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, Netherlands
[4] University of Nebraska Lincoln, 223 Burnett Hall, NE 68588 Lincoln, Nebraska, USA

Abstract

This paper examines one of the central characteristics of calendar methods: the use of landmark events. Landmarks are salient events from people's lives that aim to facilitate recall of other events and particularly their dates.

Using data from the Event History Calendar as embedded in the PSID-survey in 1998, we explore what types of landmarks respondents generated and whether the use of landmarks is related to recall accuracy. Data from the same respondents collected one year earlier were employed as a standard of comparison to assess the quality of the retrospective reports.

The findings show that respondents used a great variety of landmark events, that the number and types of landmarks were related to socio-demographic factors and that the retrieval of landmarks showed typical recency and heaping patterns. No clear relationships between landmark usage and recall accuracy were found.

Key Words: calendar methods, landmark events, recall accuracy

1. Introduction

1.1 Landmark Events in Calendar Methods

Autobiographical memory, in its widest sense, contains information about the events, which have taken place in a person's life. Among those events, there are "special periods of time" (Shum, 1998) that are more memorable than other periods. These events have been called temporal landmarks, because in a similar way to spatial landmarks, they can be used as reference points, when trying to access other locations in memory. Temporal landmarks can stem from different thematic domains and they can be public as well as personal events. Any event that can be dated fairly accurately and that provides an anchoring point in time can serve as an idiosyncratic memory landmark to a person.

In past studies, the inclusion of public (Loftus & Marburger, 1983) as well as of personal landmarks (Van der Vaart & Glasner, in press) has been shown to enhance response quality in surveys. The employment of landmark events forms an important characteristic of so-called calendar methods in social and medical surveys (e.g., Freedman et al., 1988; Sobell et al. 1988). These methods can be used in addition - or as an alternative - to the standard survey method with chronologically ordered question lists. Calendars provide

respondents with a graphical time frame that aims to facilitate access to long-term memory (Belli, 1998) when answering retrospective questions. Most instruments consist of two-dimensional grids, in which one of the axes denotes the time dimension while life themes ("work", "residence", "education", "health" etc.) are specified on the other axis.

In recent years the use of calendar techniques such as the Event History Calendar has been growing rapidly in the social and medical sciences (for a review see: Belli, Stafford & Alwin, 2009; Glasner & Van der Vaart, 2009). The outcomes of evaluation studies indicate that those instruments have beneficial effects on the quality of retrospective data. Also, experimental studies have demonstrated that calendar instruments improved recall accuracy regarding the number of reported events, dates and characteristics of events (Belli et al., 2001, 2004; Van der Vaart, 1996, 2004; Van der Vaart & Glasner, 2007). However, it remains unclear in which way specific design characteristics of calendars contribute to these beneficial effects. Our study concentrates on one important characteristic: the use of landmark events. It aims to gain more knowledge about how to identify and select autobiographical events that can be used as landmarks in calendar instruments. The following questions will be explored:

1. How many and which type of landmark events do respondents report in case of unrestricted recall instructions?
2. How are the reported landmarks distributed within the reference period and within calendar years?
3. Do different types of respondents use different types of landmarks?
4. Are numbers and types of reported landmarks related to recall accuracy?

First we will present some relevant issues from studies into autobiographical memory.

1.2 Landmark Events and Autobiographical Memory

Three main issues form autobiographical memory research may help us to understand what type of events might function as landmarks in surveys.

First, only a few events are stored in autobiographical memory with a specific date, often referred to as a "time tag" (e.g. Friedman, 1993). Events that are rehearsed, commemorated or celebrated regularly (like a birthday, one's own wedding, a nation's Liberation Day), or events that occur on "easy to remember" dates, such as 02/02/2002, are most likely to be time-tagged (Janssen et al., 2006). As a consequence, dating most other events involves a certain amount of reconstruction (Friedman, 1993). Obviously, reconstructing dates inferring information from un-tagged, and potentially falsely-dated events, increases the risk that the target event is dated incorrectly.

Second, it has generally been found that when older adults are asked to provide autobiographical memories from their lives without restrictions on content or time period, roughly half the reported events occurred within the most recent year of life (Holmes & Conway, 1999; Rubin et al., 1998); a phenomenon called the "recency effect". In addition, there is robust and substantial empirical evidence that people tend to report a relatively great number of events that have occurred in adolescence and early adulthood. This latter phenomenon is referred to as the "reminiscence bump" (Holmes & Conway, 1999; Rubin et al., 1998). Landmark events thus will not be distributed evenly over a reference period. Studies suggest that recall of more important events show less bias than

less important events and are better recall cues for other events (Brown & Schopflocher, 1998).

Third, most models of autobiographical memory state that thematically organized lifetime periods serve as the primary organizational units to autobiographical knowledge (Burt, Kemp & Conway, 2003; Conway & Pleydell-Pearce, 2000). Especially important are those lifetime periods that are central to one's concept of self and its changes across the lifespan. From these models it follows that landmarks will be most effective if they are personal events and focused on the central lifetime periods of the target population.

In sum, autobiographical memory research suggests that important, domain related, personal events are the most suitable landmark events.

2. Methods

2.1 Data
The landmarks examined here, originate from a study in which a paper-and-pencil Event History Calendar was used in a telephone survey (Belli, Shay & Stafford, 2001). In order to compare methods, 617 participants in the Panel Study of Income Dynamics (PSID) were randomly assigned to being interviewed either with an EHC, or with a standardized conventional questionnaire. The current study will focus on the 309 respondents in the EHC condition who had to provide personal landmarks at the beginning of the interview. Data collection took place in May and June 1998, using paper-and-pencil telephone interviewing in both conditions. The reference period covered the years 1996 and 1997. Answers from the previous wave of the panel study were used as control data for establishing the accuracy of respondent's reports regarding events that took place in 1996.

2.2 The Event History Calendar
The calendar instrument used in the study comprised six thematic domains (residence, household, employment, unemployment, and aid entitlements) and covered a reference period of two years. At the start of the telephone interview, immediately after the general introduction of the study, respondents were asked to list landmark events from the end of 1995 tot the date of interview in 1998. Interviewers read out the following instructions (Belli et al., 2001):

"Today, it is [May, June] of 1998. We are especially interested in events that had happened from around 2½ years ago up to those that happened around 5-6 months ago. We are especially interested in where and with whom you have lived, your employment and unemployment history, and your income during this time. First though, are there any events in the past few years, from the end of 1995 to the present, that stand out in your mind that you are able to date either exactly or approximately? These might be family, financial, job, social, or health related events."

Examples of probes that the interviewer could use for eliciting more personal landmarks included "births, deaths, divorces or marriages, a vacation, health-related events such as an accident, major purchases that you have made, a job promotion or pay rise, a residence or job change" (Belli, Shay & Stafford, 2001). Interviewers also had a list of public holidays, such as Christmas and Thanksgiving, which they could offer the respondents as dating cues.

2.3 Respondents and Data collection

We analyzed landmarks from 231 usable interviews. For 230 respondents we had background information on socioeconomic characteristics and the respondent's answers on the main questionnaire. There was a significant difference in age between the 134 female (mean age = 42.2 years) and 96 male respondents (mean age = 47.8 years). Due to oversampling of minority households in the Panel Study of Income Dynamics, 41% of respondents in our sample were black, 54% were white, and 5% reported another race. Even though in the general PSID, different respondents from the same household can be interviewed in subsequent waves, only respondents who had also taken part in the previous wave (1997) were interviewed for the methodological study.

On average, interviewers and respondents spent 3.4 minutes on collecting the memory landmarks, and another 13.9 minutes on the rest of the interview. Data entry after the interview took an average of 29.4 minutes. The interview time and the time spent on the landmark domain were only weakly related ($r=.20$, $p=.02$). The time needed for data entry is mainly related to interview time ($r=.49$, $p=.00$) and less to the time required for the landmark domain ($r=.18$, $p=.01$). Of the socio-demographic variables, only age appeared to significantly predict the length of the interview, which was shorter for older respondents ($b=-.291$, $t(212)=-4.55$; $p<.001$) and the time it took to collect the landmarks, which was longer for older respondents ($b=.17$; $t(213)=2.67$; $p<.01$). In contrast, data entry time was related to all socio-demographic variables. Taking into account the length of the interview and the total number of reported landmark events, data entry was somewhat lengthier for women ($b =.12$; $t(211)= 1.98$; $p=.05$), white respondents ($b=.11$, $t(211)= -1.76$; $p=.08$), and respondents in the higher income categories ($b=.13$, $t(211)= 2.15$; $p=.03$).

2.4 Landmark Coding Scheme

The landmark events that respondent had reported in the 231 interviews were marked in the printed transcripts and keywords or key phrases were entered into an SPSS database, together with the date of the event. Based directly on the respondents' description of events, two authors made their own first classification scheme. These schemes aggregated similar individual events into categories. After discussing the differences in both schemes, which were small, the authors formulated 17 event categories. Using this classification, all events were coded independently by the two authors and an assistant. Based on three coders the inter-coder reliability (Hayes & Krippendorff, 2007; Krippendorff, 2004) was Krippendorff's Alpha = 0.92 for all categories including 'unclear' and 'not applicable' (535 units, nominal measurement level; 231 respondents). Among each pair of coders Alpha ranged from 0.91 to 0.93 and Cramer's V ranged from V= 0,89 to V=0.90. All events that were classified differently by two coders were reconsidered and, after reaching agreement, classified into one of the 17 landmark categories (see Table 2). Inspired by Holmes and Conway (1999) we also constructed a more concise classification by merging the 17 categories into 7. In most cases we worked with this 7- category scheme (see Table 3).

3. Results

3.1 Research Question 1: The Number and Types of Landmarks

3.1.1 The Number of Landmarks per Person
In the 231 telephone interview transcripts we found a total of 535 valid landmarks events. The mean number of landmark events per respondent was therefore 2.32 (sd=1.90), the median number was 2. The number per person ranged from 0 to 11. Table 1 illustrates that respondents reported more events for 1997 than for 1996 and that a large proportion of respondents reported no landmark at all.

Table 1: The Number of Landmarks Reported in Total and for the Years 1996 and 1997.

	1995-1998	*1997*	*1996*
Mean No of events per respondent	2.32	1.18	.76
sd	1.90	1.16	0.92
No landmarks	39 (16.9%)	74 (32.2%)	112 (48.9%)
1 landmark	46 (19.9%)	77 (33.6%)	72 (31.4%)
>= 2 landmarks	146 (63.2%)	78 (34.1%)	45 (19.7%)

The cumulative percentages of all landmark events show that 17% of the respondents reported no landmarks at all, 60% reported at most 2 landmarks, and 95% of the respondents reported at most 5 landmark events.

3.1.2 The Number of Landmarks per Topic
Not only does the number of reported events differ greatly per respondent (see also the next section), but Table 2 also illustrates that there are large differences in the numbers of reported landmark events between topics.

Table 2. Frequency Distribution of the Landmark Events: 17 Categories.

Landmark category	*Content key-words*	*N*	*%*
Family and relationships	Family events, marriage, divorce, personal relationships, conflict, raising children	83	15,5
Holiday specific	Holidays (including weekends and trips) with additional information (like destination, activity, persons etc)	79	14,8
Work	Work events, jobs, unemployment, time off	70	13,1
Births	Births	63	11,8
Deaths	Deaths	54	10,1
Health	Health events	40	7,5
Education	Education events, formal schooling	33	6,2
Holiday generic	Holidays (including weekends and trips) without any specification	30	5,6
Housing	Moving and buying, selling, building houses	29	5,4
Purchases	Consumer purchases: buying, selling	17	3,2
Leisure activities	Leisure events and activities (including sports, music, other hobbies, parties, pets, volunteer work)	14	2,6

	Table 2. continued		
Accidents	Accidents	7	1,3
Public events	National and/or political events, public holidays	6	1,1
Birthday own	R's own birthday	4	,7
Other events	Other landmarks	4	,7
Birthdays others	Another person's birthday	2	,4
Total		535	100,0

In line with earlier findings (Howes & Katz, 1992; Holmes & Conway, 1999), respondents reported very few public events, although there is a chance that this can be attributed (in part) to the instructions, which do not mention public events.

3.2 Research Question 2: The Distribution of the Reported Landmarks over Time

3.2.1 Frequency Distribution by Year
The primary reference period of the study covered two years, 1996 and 1997, but respondents were told to recall events from the end of 1995 up until the time of the interview (May/June 1998). Table 3 shows the distribution of all 533 dated life events over the recall period per type of landmark category.

Table 3: Number of Landmarks by Year

Landmark	Every year	1994	1995	1996	1997	1998	Total
1 holidays	10			39	55	4	108
				22.3%	20.3%	7.5%	20.3%
2 health			3	11	24	9	47
			18.8%	6.3%	8.9%	17.0%	8.8%
3 family/relat	4	2	2	29	43	9	89
		50.0%	12.5%	16.6%	15.9%	17.0%	16.7%
4 birth/death		1	6	43	55	12	117
		25.0%	37.5%	24.6%	20.3%	22.6%	22.0%
5 work/educ		1	3	28	60	11	103
		25.0%	18.8%	16.0%	22.1%	20.8%	19.3%
6 housing/leis			2	21	31	6	60
			12.5%	12.0%	11.4%	11.3%	11.3%
7 others/uncl				4	3	2	9
				2.3%	1.1%	3.8%	1.7%
Totals (100%)	14	4	16	175	271	53	533

Table 3 demonstrates that the distribution of life events over the recall period (Chi-Square=55.30, df=42, p=0.082) exceeds a chance expectation. A disproportionate number of landmark events date from the more recent part of the reference period, the year 1997. The fact that the number of reported life events decreases relatively rapidly, even within a short reference period, is in line with the literature on long-term autobiographical memory for important events.

However, there are differences between domains (see Figure 1). The recency effect is especially pronounced for 'work and education' and 'health' events, which both show a

decline of more than 50% of reported events within one year. For the other main categories, 'births and deaths', 'family/relationships' and 'housing', but also for reports of holidays, memory decay seems to be more gradual.
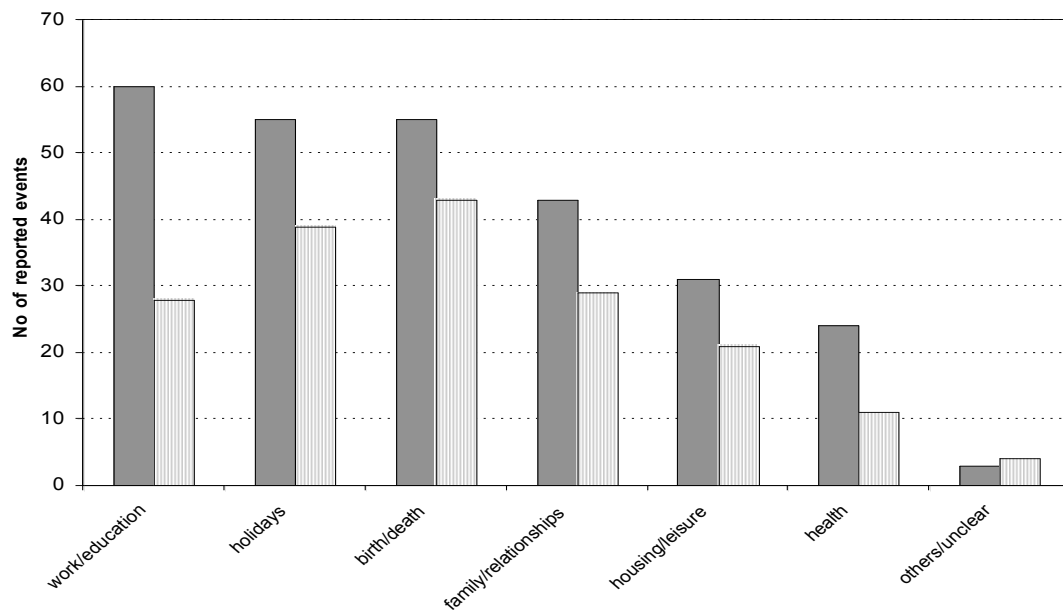

Figure 1: Number of Reported Landmarks per Year (left column: 1997; right: 1996)

3.2.2 Frequency Distribution by Month

The reference period in this study is too short to show 'reminiscence bumps', but it's well known that on a more restricted time scale the dates of retrieved events often show 'heaping', for example on specific months within years (Becker & Diop-Sidibe, 2003). In line with the presence of a heaping mechanism, Table 4 demonstrates that the distribution of events over months in our study deviates strongly from a chance distribution (Chi-Square=145,01, df=66, p<0.001). A relatively high number of events were reported to have taken place in either the summer months or in December. Furthermore, it appears that these 'heaping effects' in our data are domain specific. Not surprisingly, more holidays are reported for June, July, and August, than for the rest of the year. Health events, on the other hand, are overrepresented towards the end of the year. Perhaps not surprisingly, most family events were reported for the summer months, and towards the end of the year. The same is true for the category of housing events, leisure, and purchases and, to a lesser degree, also for births and deaths. Comparing the distribution of events over month for 1996 and 1997 shows that there are no large differences between both years (Chi-Square=15.10, df=12, p=0.24).

Table 4. Type of Landmark by Month (years 1996 and 1997; landmarks, for which no month was indicated, were excluded from the analysis)

Month	Holidays	Health	Family/ relations.	Births/ deaths	Work/ education	Housing/ leisure	Others/ unclear	Total
1	1.1%	6.7%	5.6%	1.0%	3.5%	7.7%	0.0%	3.5%
2	4.5%	3.3%	1.4%	6.3%	2.4%	0.0%	0.0%	3.3%
3	0.0%	10.0%	4.2%	3.1%	5.9%	7.7%	0.0%	4.2%
4	3.4%	10.0%	5.6%	9.4%	7.1%	3.8%	16.7%	6.5%
5	10.1%	6.7%	9.7%	4.2%	18.8%	1.9%	0.0%	9.1%
6	14.6%	6.7%	11.1%	13.5%	16.5%	13.5%	0.0%	13.3%
7	28.1%	6.7%	16.7%	6.3%	4.7%	11.5%	0.0%	12.8%
8	12.4%	6.7%	12.5%	11.5%	11.8%	7.7%	16.7%	11.2%
9	7.9%	6.7%	4.2%	10.4%	11.8%	9.6%	16.7%	8.8%
10	5.6%	6.7%	9.7%	10.4%	2.4%	7.7%	16.7%	7.2%
11	3.4%	13.3%	5.6%	11.5%	5.9%	17.3%	0.0%	8.4%
12	9.0%	16.7%	13.9%	12.5%	9.4%	11.5%	33.3%	11.9%
N	89	30	72	96	85	52	6	430

These findings demonstrate that, if landmarks are self-generated by respondents in an unrestrained way, dates may be biased by domain specific heaping effects. Though not all occurrence of heaping has to be an artifact (e.g., the heaping of holidays in the summer), the distribution over the months is uneven.

3.3 Research Question 3: Relationships Between Respondent Characteristics and Landmark Usage

Ranging from 0 to 11, the number of reported landmarks differs greatly per person and the question is whether this is related to socio-demographic factors and whether respondents who report more events from one domain also report more events from other domains.

There are several weak, but significant, relationships between socio-demographic variables and the number of reported life events (Table 5). The correlations between socio-demographics and landmarks are similar for 1996 and 1997 reports separately. Gender appeared to be related to the total number of reported landmark events. On average, female respondents (N=133) reported significantly more landmark events than male respondents (N=93) (overall: female 2.59 – male 2.01, t=2.26, p=0.03). This difference was even more pronounced for events that had taken place in 1996 (female .92 – male .55, t=3.02, p=0.003). Differences between men and women within the subcategories of events were not significant, except for the holiday category, in which women reported slightly more events.

On average, white respondents reported the most landmarks of all racial groups. While white respondents reported 2.68 events, black respondents reported 1.96 (N=92) and respondents of other races reported 2.17 landmarks (N=12) (F=3.91, df=2, 223, p=.02). Taking into account age, gender, and income, the differences between white and non-white respondents as a group are statistically significant for the total number of reported

events as well as for events within specific years. With regard to the sub-categories, non-whites reported significantly fewer work and education events.

At first sight, age seems to be negatively related to the total number of reported landmarks. When taking in account race, gender, and income of the respondent, however, this relationship becomes non-significant (the men in our sample were significantly older than the women). The results on specific domains show a positive relationship between age and the number of reported health events. On all others domains, age is negatively correlated with the number of reported landmarks, even though this correlation is only significant regarding the number of housing events (p= .03).

Income is not related to the total number of landmarks reported, but there is a near-significant positive relationship with the number of work and education events (p = .06), as well as a near-significant negative relationship with the number of reported births and deaths (p = .08).

Table 5: Multiple Regression (standardized betas) of Socio-Demographic Characteristics on the Number of Reported Events for 1996 and 1997 and per Six Event Categories (family/relationships & births/deaths combined, 'others' left out)

	Total	1997	1996	Holidays	Health	Family & Births/D.	Work/ed.	Hous.
Age	-.11	-.11	-.12*	-.05	.18**	-.09	-.08	-.17**
Female	.20***	.16**	.24***	.15**	.04	.13*	.07	.04
Non-white	-.25***	-.18**	-.23***	-.12*	-.10	-.13*	-.14**	-.11
Income	.01	.05	-.05	.08	.01	-.13*	.14*	-.04
R^2	.094	.06	.106	.04	.04	.05	.05	.04
F	5.38	3.4	6.11	1.94	2.35	2.6	2.88	1.86
P	.000	.01	.000	.105	.06	.038	.02	.118

*, **, ***: Significant at .10, .05, and .01 level

3.4 Research Question 4: Relationships Between the Number and Types of Landmarks and Recall Accuracy

3.4.1 The Number and Types of Landmarks and Recall Accuracy

In the present section we examine whether the use of landmark events is related to data quality. Agreement scores – treated as an indication of 'recall accuracy' - could be assessed for the 1996 reports only, by comparing the experimental 1998-reports with the 1997 reports about the preceding year. Some measures, however, reflect a slightly different reference period, which is the period between the interviews held in 1996 and 1997. It is likely that both the 1996 and 1997 landmarks could be related to data quality of the 1996 reports, although it is also reasonable to assume that 1996 landmarks would be more predictive than the 1997 ones. Due to 1997 landmarks, respondents may realize that an event did not take place in 1996 but in 1997, or the other way around. Since the reference period of the PSID-study was 1996-1997 respondents might use landmarks from both years to distinguish between the years. Moreover, due to small numbers we

were forced sometimes to merge both years and do analyses on the total number of landmarks

3.4.2 Difference Scores Correlated With Landmark Frequency

Table 6 presents the difference scores regarding eight independent measures that were available for analysis. No correlation was found between the total number of landmarks and recall accuracy (in terms of absolute differences) for 6 out of 8 measures. For the remaining 2 measures (regarding the number of weeks away from work due to illness or vacation) the (small) correlations indicate a positive, instead of a negative, relationship between the total number of landmarks and recall error. The number of 1997 landmarks shows both a positive and a negative relationship. Table 6 also presents the relationships between the difference scores and the number of landmarks in specific categories, again illustrating that reporting more landmarks mostly goes along with more recall error.

Table 6: Correlations Between the Number of Landmarks and Absolute Recall Error. (Standard deviations between brackets. Columns 4-8 involve all 1995-1998 landmarks.)

	N 96+97	N 97	N 96	Holi-days	Birth/death	Wrk/edu	Wrk	Hous/leis	N
Persons moved in					-.12 (.08)				220
Number of jobs						.17 (.01)	.15 (.03)	.14 (.04)	226
Income					-.14 (.05)	.18 (.01)	.16 (.02)		205
Wks working							.13 (.06)		213
Wks out of labor							.12 (.08)		213
Wks vacation	.14 (.05)	.12 (.07)				.14 (.04)	.13 (.05)		213
Illness self				.12 (.07)		.13 (.09)			213
Illness other	.17 (.01)	-.13 (.06)		.14 (.04)					213

Note: The subcategories 'work' (only) is presented in addition to the 7-category scheme. The category 'others' was excluded due to small numbers.

We have to be very cautious interpreting these few and weak associations, but thanks to the use of multiple measurements some patterns become visible. One salient aspect is that the results are very domain specific. Landmarks from the working domain appear to be most central to recall error in this study, which seems logical since a majority of issues is clearly work related. Next, only the very important life events involving 'birth' or 'death' are associated with a reduction of recall error. Finally, the many positive relationships with recall error raise the question what the number of landmarks is indicating: they may not reflect the availability of 'more recall aids' but may instead indicate that these respondents had more events to report in that domain and thus also are at greater risk in making recall error due to having a more difficult retrieval task.

Further exploration underpinned that most of the time there's a positive association between the reports on the activities in the life domains and the number of landmarks (all correlations are weak, ranging from r=.12 to r=.31). Usually the number of landmarks (total and per category) was only related to recall error if this was the cas.. So a greater number of landmark events may not necessarily indicate more 'recall aid' but to the contrary a 'more difficult recall task'.

3.4.2 Correlations between experimental and standard measures by landmark frequency

Since the lack of variance in the difference scores might have hampered our analysis and since the difference scores were not applicable to all of the measures we had available, Table 7 presents correlations between the 1998 experimental reports for 1996 and the 1997 standard of comparison. We compared the figures of respondents who reported one or more landmarks to those of respondents who did not report any landmarks, controlling for age, gender, race and income. It appears that for 6 out of 13 measures the correlations are higher for respondents who reported landmarks from 1996, compared to those who did not, while for 7 measure it's the other way around. Regarding the 1997 landmarks for 7 out of the 11 available measures the correlations are higher for respondents who reported landmarks from 1997, compared to those who did not report any landmarks, whereas 2 measures appear to be equal and for another 2 the correlations are lower. These findings suggest that the overall pattern in the experimental data more closely resembles the standard of comparison if respondents did report landmarks in 1997, but not if they did report landmarks in 1996, the latter being the year 'of comparison'. The 1996 landmark findings are in line with our previous analysis that showed that more landmarks went together with greater recall error. The finding that having landmarks in the year after 1996 is largely positively related to recalling 1996 events might be due to bounding impact of these later landmarks, but this is only a hypothesis. The main finding is that no clear relationships can be found between 'recall accuracy' and the number of landmarks reported. The question remains whether the landmarks functioned as recall aids or just were an indication of more activity in people's lives.

Table 7: Correlations Between the 1998 Reports and the 1997 Standard of Comparison Concerning 1996, by the Number of Landmarks Reported for 1996 and 1997 (controlled for age, gender, race and income).

	Partial correlations	N96		N97	
		0	>= 1	0	>= 1
1	Persons moved in	.52	.72	.18 [a]	.83
	N	(95)	(96)	(58)	(133)
2	Persons moved out	.58	.95	.93	.88
	N	(95)	(96)	(58)	(133)
3	Number of jobs	.69	.57	.41	.67
	N	(96)	(99)	(60)	(135)
4	Income	.94	.85	.79	.96
	N	(95)	(97)	(60)	(132)
5	Weeks worked	.77	.64	.64	.74
	N	(94)	(89)	(57)	(126)
6	Weeks out of labor	.74	.67	.68	.74
	N	(94)	(89)	(57)	(126)
7	Wks vacation	.35	.16 [b]	.28 [c]	.26 [d]
	N	(94)	(89)	(57)	(126)
8	Wks self ill	.48	.77	.63	.76
	N	(94)	(89)	(57)	(126)
9	Wks other ill	.58	.69	.69	.68
	N	(94)	(89)	(57)	(126)
10	Wks unemployed	.75	.28 [b]	-	.83
	N	(94)	(89)	(57)	(126)
11	Moved address (i)	.57	.73	.49	.74
	N	(87)	(89)	(53)	(123)
12	Food stamps (i)	.80	.62	.78	.68
	N	(96)	(99)	(60)	(135)
13	Child aid (i)	.64	1.00	-	.81
	N	(96)	(99)	(60)	(135)

All p <.001, except: a) p= .16, b) p = 01, c) p= .03, d) p=.003
Spearman correlation coefficient all other: Pearson's correlation coefficient

4 Conclusions and Discussion

The answers to our four research questions on landmark usage by respondents in a survey setting - as put forward in the introduction - can be summarized as follows.
First, in a relatively unrestrained landmark task, like in the present study, respondents use a great variety of landmark events. Also their reports show large differences in the number of landmark events per topic. We inductively came up with seven landmark categories (holidays, health, family/relationships, birth/death, work and education, housing/leisure, others) that had much in common with the primary life domains that according to current memory models are central to autobiographical memory.

Second, most events stem from the most recent year and are unevenly distributed over the months. Both the retention rates and heaping patterns are domain specific. There is a relatively large decline in the number of reported landmark events when the reference period is extend from one to two years.

Third, associations (though weak) exist between the number and types of landmarks and respondent characteristics. Female and white respondents reported more landmarks than male and non-white respondents. Age, race and income influence the types of landmarks reported.

Fourth, landmark usage was weakly related to recall accuracy and showed both positive and negative associations with accuracy. The main explanation probably is that the landmarks were hardly used by the interviewers during the core of the interview to actively cue the respondent's memory. If asking for landmarks during the introduction stimulated recall at all, it will have functioned mainly as a "warming up" procedure. However, we also found indications that the number of landmarks is related to the level of activity of respondents in a certain life domain. This could explain the negative associations: more activity goes together with more landmarks but also greater risk of recall error.

If landmarks are to be employed as aided recall tools, it is important that landmark events are dated accurately and that they do not disturb or bias retrieval processes. From both laboratory studies into autobiographical memory and questionnaire studies it appears that a free recall of life events may not lead to optimal landmark usage. Landmarks probably can be more effective when employing a structured procedure. The instruction could be regulated at least in two ways. First, in order to reduce recency and heaping effects, respondents could be requested to recall landmark events from specific time periods, like from each season or a fixed number of months. Second, in line with the important role of lifetime periods in memory, respondents could be asked to report 'domain related' landmarks, focused on specified themes that are central to the data collection.

Acknowledgements

References

Becker, S., & Diop-Sidibe, N. (2003). Does Use of the Calendar in Surveys Reduce Heaping? *Studies in Family Planning, 34*(2), 127-132.

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.

Belli, R. F., Lee, E. H., Stafford, F., & Chou, C.-H. (2004). Calendar and question-list survey methods: Association between interviewer behaviors and data quality. *Journal of Official Statistics, 20*(1), 185-218.

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

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

Brown, N.R. (2005). On the Prevalence of Event Clusters in Autobiographical Memory. *Social Cognition, 23* (1), 35-69.

Brown, N.R., & Schopflocher, D. (1998). Event Cueing, Event Clusters, and the Temporal Distribution of Autobiographical Memories. *Applied Cognitive Psychology, 12*, 305-319.

Burt, C. D. B., Kemp, S., & Conway, M. A. (2003). Themes, events, and episodes in autobiographical memory. *Memory & Cognition, 32*, 317-325.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of Autobiographical memories in the self-memory system. *Psychological Review, 107*, 261-288.

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

Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin, 113*, 44-66.

Glasner, T.J., &. Van der Vaart, W. (2009). Applications of Calendar Instruments in Social Surveys. A Review. *Quality and Quantity. 43,* 333–49.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. Communication Methods and Measures, 1, 77-89.

Holmes, A., & Conway, M.A. (1999). Generation Identity and the Reminiscence Bump: Memory for Public and Private Events. *Journal of Adult Development, 6* (1), 21-34.

Howes, J.L., & Katz, A.N. (1992). Remote Memory: Recalling Autobiographical and Public Events from across the Lifespan. *Canadian Journal of Psychology, 46* (1), 92-116.

Janssen, S., Chessa, A. & Murre, J. (2006). Memory for time: How people date events. *Memory & Cognition,* 34, 138–47.

Jobe, J.B. and Mingay, D.J. (1989). Cognitive Research Improves Questionnaires. American Journal of Public Health, 79, 1053-1055.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology.* Thousand Oaks, CA: Sage.

Loftus, E. F., & Marburger, W. (1983). Since the eruption of Mt. St. Helens, did anyone beat you up? Improving the accuracy of retrospective reports with landmark events. *Memory & Cognition, 2*, 114-120.

Rubin, D.C., Rahhal, T.A., & Poon, L.W. (1998). Things Learned in Early Adulthood are Remembered Best. *Memory & Cognition, 26* (1), 3-19.

Shum, M.S. (1998). The Role of Temporal Landmarks in Autobiographical Memory Processes. *Psychological Bulletin, 124* (3), 423-442.

Sobell, L. C., Sobell, M. B., Leo, G. I., & Cancilla, A. (1988). Reliability of a timeline method: Assessing normal drinkers' reports of recent drinking and a comparative evaluation across several populations. *British Journal of Addiction, 83*, 393-402.

Van der Vaart, W. (1996). *Inquiring into the past. Data quality of responses to retrospective questions.* Ph.D.-thesis. Amsterdam: VU University.

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), 301-317.

Van der Vaart, W. & Glasner, T.J. (in press). Personal landmarks as recall aids in survey interviews. *Field Methods.*

Van der Vaart, W., & Glasner, T.J. (2007). Applying a Timeline as a Recall Aid in a Telephone Survey: A Record Check Study. *Applied Cognitive Psychology, 21 (2),* 217-238.