Health Status Measurement via Event History Calendars vs. Conventional Questionnaires: "Long Time, No Status Change"

Debra R. Miller, Clarissa R. Steele, Ipek Bilgen, Robert F. Belli[1]
[1]University of Nebraska-Lincoln, UNL Gallup Research Center, 201 North 13th Street, Lincoln, NE 68588-0241

Acknowledgement
This research has been jointly funded by the National Institute on Aging and National Institute of Child Health and Human Development, AG/HD17977.

Abstract
In retrospective event history calendar (EHC) interviews, a response may be incomplete. Response and nonresponse are not mutually exclusive, as a particular question captures information on multiple time periods. This paper explores the extent to which responses are complete in EHC interviewing by evaluating life-course health status responses from event history calendar and standardized conventional questionnaires. The aim is to better understand the number and nature of item nonresponse occurrences, don't know responses and related interviewer probes. Previous analysis indicates that event history calendars elicit more accurate reporting than conventional questionnaires relative to annual panel data. The present behavior coding reveals fewer don't know responses, more directive probes and more failures to probe among EHC than conventional interviews relative to health changes.

Key Words: event history calendar, standardized conventional questionnaire, verbal behavior coding

1. Introduction

1.1 Sources of Survey Error

Various sources of survey errors can be considered relevant to measurement or representation (see Groves, Fowler, Couper, Lepkowski, Singer, & Tourangeau, 2004, p. 48). Measurement sources or survey errors include validity, measurement error and processing error; representation sources include coverage error, sampling error, nonresponse error and adjustment error. As with more conventional or standardized forms of interviewing, item nonresponse and measurement error are particularly relevant to the quality of the health status section of the Panel Study of Income Dynamics (PSID) life-course history calendar.

Nonresponse error is a function not only of a missing response, but also of true values differing significantly from observed values. Nonresponse error therefore pertains to missing responses that, if present, would systematically differ from observed responses. Two forms of nonresponse include unit nonresponse, in which no responses are available for a sample member, and item nonresponse, in which a respondent provides answers to some but not all questions.

Groves equates measurement error with "'observational errors' … [arising] from the interviewer, the instrument, or the respondent" (1989/2004, p. 16). One source of measurement error relates to the influence of interviewer expectations, a source particularly pertinent to probing in event history calendars. Groves defines measurement error as "a component of variance in the observed values of indicators, not corresponding to variability in the true values of the underlying measures" (1991/2004, p. 18). Measurement error may artificially increase or decrease variance relative to true values if interviewers fail to probe.

Much has been written regarding survey error in conventional standardized interviews, but not in regard to event history calendar interviews. The longitudinal nature of diaries and event history calendars may compound the dimensionality of observed versus actual values. This paper discusses potential item nonresponse and measurement error among respondents who answered at least some of the survey questions. Based on prior coding, the initial hypothesis of the current coding and analysis was that event history calendar (EHC) interviews would exhibit a higher incidence of missing probes regarding health status change in comparison to conventional interviews.

1.2 Conventional Interview Background

Autobiographical memory, the memory people have of their own life experiences, serves as a basis for event history calendar interviewing. This aspect of memory can be thought of as "the psychological history of the individual self" (Conway, Wang, Hanyu & Haque, 2005, p. 739), as referring to events from earlier times in a person's life (Baddeley, 1992), or as being an "internal or inward memory" (Halbwachs, translated 1980, p. 52). Autobiographical memory can be characterized as episodic or generic (Brewer, 1994). Episodic memory involves specific episodes or events that have happened "in particular places at particular times" (Tulving, 2002, p. 3). Generic memory involves "imaginal properties that result from experiencing" multiple similar events (Brewer, 1994, p. 12). Details blur together across experiences, causing generic memories to differ from actual episodes or events.

Human memories can be thought of as stored in clusters by means of which specific memories are retrieved (See Robinson, 1986, p. 173 regarding temporal clustering). The goal of event history calendar interviewing is to access respondents' thoughts as originally stored in autobiographical memory, resulting in respondents becoming more highly engaged in the reporting process, which in turn aims to maximize validity and minimize error.

Event history calendar (EHC) interviewing can be distinguished from conventional interviewing by the increased opportunity for memories to be retrieved via top-down (general-to-specific), sequential, or parallel associations among contemporaneous events (See Dijkstra, Smit & Ongena, 2009, p. 257; Belli, 1998, p. 395). Event history calendar interviews have been conducted since Balan, Browning, Jelin, and Litzler (1969) reported their use in collecting quantitative retrospective life-course information (see Belli & Callegaro, 2009). Since that time, other calendar-related studies have included work reported by Freedman, Thornton, Camburn, Alwin, & Young-DeMarco in 1988; Neighborhood History Calendars used in Nepal (see Axinn, Barber, & Ghimire, 1997; Axinn & Pearce, 2006), and the English Longitudinal Study of Aging (ELSA) reported by Pascale & McGee (2008). A two-year Calendar Methods Study conducted among a subset of panel members from the Panel Study of Income Dynamics (PSID) at the

University of Michigan has shown greater accuracy for history calendar than conventional interview data across domains of residence changes, number of jobs, earned income, weeks unemployed, and weeks away from work due to personal illness or illness of another in comparison with original concurrent panel data (Belli, 2003; also see Belli, Shay & Stafford, 2001).

The event history calendar condition of the Health and Economic Measures Study has been reported to have decreased memory failure than did the conventional retrospective questionnaire condition (see Figure 1 reprinted from Bilgen & Belli, 2008), possibly due to increased respondent engagement. Among a total of seven domains, the average number of domains with memory failures was initially greater in the event history calendar condition for the year 1984. Thereafter, the average number of domains with memory failures in the event history calendar condition decreased relative to the conventional questionnaire condition until the most recent year of 1997, at which time the number of discrepancies between retrospective and annually-collected concurrent data converged between the two conditions.

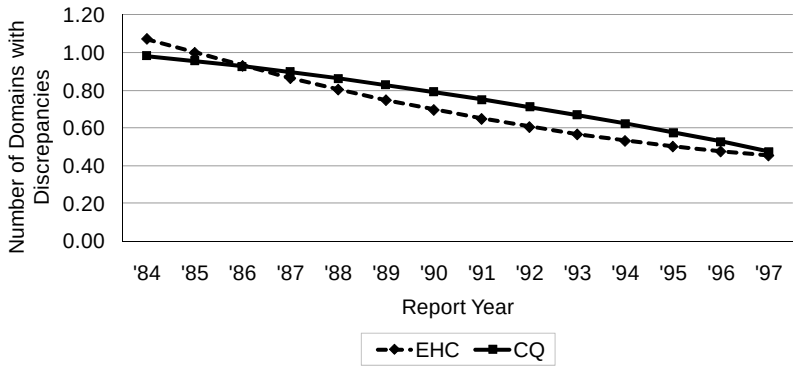

Figure 1: Panel versus Retrospective Discrepancies

1.3 Comparing Sources of Error

The current analysis initially intended to consider, define and compare item nonresponse between EHCs and conventional interviews. In a cross-sectional study, data missing for a question could relate either to failure of an interviewer to ask a question, or alternatively to a respondent's lack of knowledge or choice not to answer an entire question. However, in longitudinal, retrospective studies, whether administered via history calendar or conventional interviews, multi-level data entails time as a predictive unit of analysis. Item nonresponse, then, is not cross-sectional (see Singer & Willett, 2003). Hence, data could be missing for an entire question or simply for a portion of a question, in which case the response could be thought of as incomplete rather than entirely missing. Respondents could know the answer to a question for part of their life course without knowing the answer for the entire life course. For instance, portions of life-course data could be missing if a respondent was unsure of an initial or a final health status, or if they were uncertain about whether or when a status had changed (see Dasbach, 1994, p. 1776 regarding low frequency of don't know responses in self-reported health status of persons with diabetes compared to others of similar age).

Evaluation of nonresponse error in event history calendar interviewing must inevitably consider the relationship between nonresponse and measurement error. In the

longitudinal context of event history calendar retrospective reports, we expected an absence of probing which in turn was expected to lead to incorrect values. Missing values were expected to occur less often than incorrect values. Because incorrect values could lead to measurement error, and missing values could lead to nonresponse error, it was expected that, in the context of an event history calendar, measurement error would be more likely to arise than would be the case for nonresponse error.

Even in the presence of a robust number of status changes, an unasked probe in multi-level data (and event history calendars in particular) could cause the shape of a respondent-level trend be flatter at a particular point in time than would be the case if the probe had been provided and the respondent had experienced a change during the corresponding period of time. Assume, for example, that a data file includes no missing points but that a probe regarding a final change in health status was not provided. For the time period in which a change was not probed, data would be expected to represent measurement error as opposed to missing data or item nonresponse (see Figure 2).

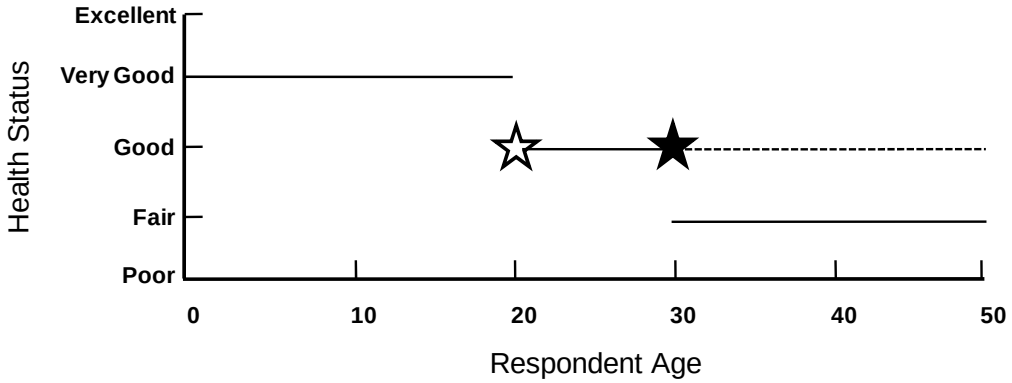

Figure 2: Example of Measurement Error versus Item Nonresponse

☆ *Status change provided by respondent following interviewer's probe*

★ *Status change perceived by respondent but not provided due to interviewer failure to probe for additional status change*

—— *Respondent's perception of health status*

- - - *Data values representing measurement error for a time period regarding which changes were not asked and for which respondent perceived a different status level than was provided for the previous time period*

The purpose of the present analysis, then, is to consider the presence, nature, causes and impact of missing measures in the health status section of the Health and Economic Measures study, for life-course history calendar relative to conventional questionnaire interviews.

2. Method

The present analysis involves interview transcripts from the 2002 Health and Economic Measures Lifecourse Validation Study, which consisted of telephone interviews among a subset of panel members who had participated in the PSID since 1980 (see Belli, Smith, Andreski, & Agrawal, 2007). Participants were randomly assigned to either an EHC or

conventional interview condition. Interviewers (n=26) were matched based on conventional interview experience and randomly assigned to an event history calendar or conventional interview condition. Retrospective domains included residence, marriage and cohabitation, children, education, labor, parent/guardian, parental socioeconomic status, and health history (see Belli, Smith, Andreski, & Agrawal, 2007).

2.1 EHC Instrument (Health and Economic Measures Study)

Public landmark events were provided with the event history calendar interviews to serve as temporal anchors. At the beginning of each domain, interviewers were provided with a loosely structured script (see Appendix A – Health Domain Script – Event History Calendar) which they could read verbatim or upon which they could improvise wording. The present report focuses on the health status question of the health history domain. The health status question was comprised of a five-point rating scale and was asked immediately after a question on disabling health conditions and hospitalization in order for the latter to provide information by which interviewers could probe the health status.

2.2 Conventional Instrument (Health and Economic Measures Study)

In the conventional interview, the series of health status questions was presented to interviewers in an automated sequence eliminating the need to think about the sequence or flow of questions and answers. Respondents were asked their health status from birth till the age of seven on a five-point rating scale, followed by a question asking whether their health had consistently stayed the same since that time (see Appendix B – Health Domain Script – Conventional Questionnaire). If a respondent answered negatively, they were asked the year of the change, followed by iterations of the rating question and the question asking whether health status has remained the same, until a positive response indicated that their health status had remained the same.

2.3 Verbal Behavior Coding

Verbal behavior coding is sometimes described as "interaction coding" (Groves, 1989/2004, p. 382) or "interaction analysis" (Dijkstra, 2008, p. 54), and can be distinguished as having three main purposes, i.e., testing questionnaires, studying the performance of interviewers, or evaluating the interviewer-respondent interaction (see Dijkstra, 2008, p. 54). The present analysis used behavior coding to identify interviewer and respondent behaviors, and hence to compare the functionality of event history calendars and standardized conventional interviews in regards to measurement error.

One difficulty in evaluating measurement error based on behavior coding is that the occurrence of a coded behavior may not necessarily reflect measurement error in which the provided response differs from the true value (see Groves, 1989/2004, p. 387). In the case of the present event history calendar interviews, coded behavior may at times have been desirable and beneficial; conventionally undesirable interviewer behavior may not necessarily have resulted in data differing from reality. As Oksenberg, Cannell & Kalton have indicated, behavior coding is limited in its ability to identify causes of behaviors (1991). The authors of the present analysis attempted to code interviewer behaviors that could have an impact, either positively or negatively, on a respondent's answers.

Analysis was conducted via verbal behavior coding based on transcripts of interview recordings for which the coding scheme is shown in Table 1 (see also Appendix C – Verbal Behavior Coding Examples).

Table 1: Verbal Behavior Coding Scheme

Acronym	*Short Name*	*Full Description*
INSR	Full Scale Not Read	Full scale (five consecutive points) not read in any single turn of the health status section; a scale could be successfully read from positive to negative, or from negative to positive.
RDK	Resolved DK	"Don't know" mentioned but resolved before end of health status section; multiple occurrences coded in regards to change of status, status level and/or year/age of change.
UDK	Unresolved DK	"Don't know" mentioned and remained unresolved at the end of the health status section; multiple occurrences coded in regards to change of status, status level, or year/age of change.
DIR	Directive	Interviewer directively (non-neutrally) probed, implying potential for bias; multiple occurrences coded in regards to change of status, status level, or year/age of change.
NPR	No Probe	Interviewer failed to probe when respondent information was missing or uncertain; multiple occurrences coded in regards to change of status, status level, or year/age of change.
DISI	Disability Reference (Interviewer)	Interviewer references a particular disability status or definition if respondent has not already mentioned the same disability
DISR	Disability Reference (Respondent)	Respondent references a particular disability status or definition if interviewer has not already mentioned the same disability

2.3.1 Participants

Interviews for the Health and Economic Measures (HEM) study were conducted retrospectively in 2002 among participants from the Panel Study of Income Dynamics (PSID). Both EHC and CQ conditions were administered by means of Computer-Assisted Telephone Interviewing (CATI). Of the 626 interviews conducted, 311 (AAPOR definition 1 respondent level cooperation rate of 93%) were conducted by an EHC, with the remaining 315 (AAPOR definition 1 respondent level cooperation rate of 96%) conducted by conventional questionnaire. The present analysis looks only at the health status portion of the health domain, a question intended to gain a subjective measure of respondents' perceptions of their health over their lifetime.

Due to inaudible and/or incomplete audio recordings, 297 (95%) of the 311 EHC interviews were transcribed and randomly assigned to one of three replicates. Of the 315 conventional interviews, 291 (92%) were transcribed and randomly assigned to one of three replicates. A subset of replicates could then be selected to save costs of transcribing, coding, and analyses. Among the 195 EHC interviews assigned to the first two of three replicates, 190 had a health status section sufficiently audible for transcribing and verbal behavior coding. Among the 195 conventional interviews assigned to the first two of three replicates, 192 had a sufficiently audible health status section.

Table 2 shows demographic information for coded respondents. In order to interview respondents who had participated in the Panel Study of Income Dynamics since 1980 or earlier (see Belli, Smith, Andreski & Agrawal, 2007), sample members were 45 years of

age or older. Descriptive statistics for EHC respondents from the original Health and Economic Measures Study included an average age of 62.5 years (SD = 11.8); 50.3% of EHC respondents were male and 49.7% female; 80.5% were white, and 19.5% were nonwhite. Descriptive statistics for conventional interview respondents include an average age of 63.4 years (SD = 10.9); 52.1% of conventional questionnaire respondents were male and 51.9% female; 83.6% were white, and 16.4% were nonwhite.

Table 1: Demographic Information for Verbal Behavior Coded Respondents

Condition	*Completes*	*Behavior Coded*	*Mean Age*	*% Male*	*% White*
EHC	311	190	63	50	84
CQ	315	192	63	52	81

Two coders were involved, one of whom (the first author) served as the primary coder. For the first replicate of transcripts (92 EHC and 93 CQ transcripts), coding was dependent in that the two coders practiced and came to final agreement on the first replicate of transcripts. For the second replicate of transcripts (98 EHC and 99 CQ transcripts), coding was determined independently.

3. Analysis and Discussion

To measure reliability between the two coders, Cohen's kappa coefficients were calculated (see Thorndike, 2005). Evaluating the level of agreement for dichotomous variables, separate kappa statistics were calculated for each of the seven codes that could have been assigned. For the sake of kappa determination, code values were determined dichotomously according to whether or not a particular code was assigned for an individual respondent at any time during their entire life course. Kappa values were determined based on the second, independently-coded replicate. Values for Kappa coefficients usually range between 0 for complete disagreement and 1 for complete agreement (Alwin, 2009, pp. 286-287), with minimal recommended values of .40, and preferable values greater than 0.75 (Fleiss, Levin & Paik, 2003, p. 604). As indicated in Table 3, the "no probe" code failed to reach a minimum cut-off level of a 0.40 Kappa value. The remaining agreements ranged between 0.43 for "unresolved don't know" and 0.82 for "scale insufficiently read".

Table 2. Cohen's Kappa Coefficients

Abbreviation	*Short Name*	*Event History Calendar + Conventional Interviews Kappa (n=382)*
INSR	Full Scale Not Read	0.82
RDK	Resolved DK	0.65
UDK	Unresolved DK	0.43
DIR	Directive	0.67
NPR	No Probe	**0.34**
DISI	Disability Reference (Interviewer)	0.73
DISR	Disability Reference (Respondent)	0.61

Differences between average occurrences of coded behaviors in the event history calendar versus conventional questionnaire conditions were observed to be significant relative to a chance probability level of at most .05, with the exception of unresolved don't knows, which were significant relative to a probability of .10. The mean occurrence of each code (shown in Table 4) reflects the average number of times per transcript a particular code was assigned by the primary coder. For the insufficiently read scale code (assigned a maximum of one time per interview), the average number of interviews with codes assigned was significantly (p=.002) higher in the event history calendar condition. Codes other than insufficiently read scale were assigned multiple times per interview according to the number of times the behavior occurred. For resolved and unresolved don't know, the average number of times codes were assigned was higher in the conventional questionnaire condition, with observed significance values of p=0.018 and p=0.098, respectively. Directive for change was assigned much more frequently in the event history calendar than in the conventional questionnaire condition, as indicated by p=0.000. No probe was also assigned more frequently in the event history calendar condition, as indicated by p= 0.016. Codes for Disability Reference by Interviewer and Respondent were assigned significantly more times in the event history calendar than the conventional questionnaire interviews (p=0.000 and p =0.003, respectively).

Table 3. Proportion of Transcripts with Master Codes Assigned

	Event History Calendar		*Conventional Questionnaire*		
Abbreviation	*# of codes*	*Mean (n=190)*	*# of codes*	*Mean (n=192)*	*t-test*
INSR	14	0.0757	2	0.0106	3.1428**
RDK	6	0.0324	18	0.0952	2.3863*
UDK	2	0.0108	7	0.0370	1.6563
DIR	175	0.9459	28	0.1481	7.8272**
NPR	60	0.3243	36	0.1905	2.4281*
DISI	36	0.1946	1	0.0053	5.8847**
DISR	43	0.2324	20	0.1058	3.0259**

**Significant at the p<.05 level* ***Significant at the p<.01 level*

Behaviors originally hypothesized to occur more frequently in the event history calendar interview condition (resolved and unresolved don't know) occurred less frequently. Other coded behaviors occurred more frequently in the event history calendar than in the conventional questionnaire condition. Interviewer directives and references to the disability section were particularly high in the event history calendar condition. It is interesting that although behaviors considered undesirable in conventional interviews are present more often in the even history calendar condition, they are accompanied by more accurate memory recall. Perhaps some of the coded behaviors considered undesirable in conventional interviews were influenced by the nature of the health status question asked, whereas the accuracy of memory recall shown in Figure 1 reflects the health history domain as only one of seven domains averaged for each question.

4. Conclusions and Recommendations

Lessons learned from this analysis of the health status question between event history calendar and conventional questionnaire interviews point toward three recommendations. The first two relate to interviewer training, and the final relates to further analysis. First, interviewers need to be trained to clearly and proactively ask about a final health status

change without assuming that a change mentioned by a respondent represents a final change until the respondent indicates no additional changes. Second, interviewers need definitive and specific guidance during training and feedback sessions, particularly regarding which aspects of an event history calendar interview may allow more flexibility and which require more structure. One example of necessary structure would include reading all response options in a five-point scale. Another example of a necessary structure would include asking respondents whether they have had a change in health status until they indicate they have not. How this is asked or determined could be flexible.

In future analysis, verbal behavior reflected in transcripts needs to be compared with observed data values in order to determine the extent to which coded behaviors may or may not be problematic. Future work could also match respondents who were not asked about health status changes to respondents with similar demographic characteristics and health status during the years about which the questionnaire asked. If a larger respondent base were used, those with complete status changes could be used to estimate the number or portion of missing status changes among matched respondents for whom a status change was not probed. Further consideration could be given as to whether health status can change (or be perceived as changing) when a person is not disabled. Further data analysis that could impact the answer to these questions will likely be forthcoming in regards to the Health and Economic Measures Study. Consideration could also be given to whether or not, within the same study, similar results arise in domains other than health status.

Appendices

Appendix A – Health Status Script – Event History Calendar

We would also like to know how your general health has been over your entire lifetime. For each year of your life, would you say that your health had been excellent, very good, good, fair, or poor? It may help to focus on those years in which there was a CHANGE in your general health from one state to another.

Appendix B – Health Domain Script – Conventional Questionnaire

HH-28 Now I'd like to ask about your general health. Thinking back to your early childhood, from birth until you reached the age of 7**,** would you say that your health was excellent, very good, good, fair, or poor?

Excellent.....1
Very good.....2
Good.....3
Fair 4
Poor 5

HLTHCHG1

HH-29 You've indicated that during early childhood, your health was (fill HH-28). Since early childhood, has your health stayed at this level?

Yes.....1
No.....5

Flow Check HH-29: If yes go to HH-33.

HTHCHGYR

HH-30 In what year did your health change?

— — — —
Year

GENHLTH2

HH-31 When your health changed in (fill HH-30), would you say it became excellent, very good, good, fair, or poor (eliminate the response option given in HH-28)?

Excellent.....1
Very good.....2
Good.....3
Fair 4
Poor……5

HLTHCHG2

HH-32 Since your health changed in (fill HH-30), has your health consistently stayed at this level?

Yes.....1
No.....5

Appendix C – Verbal Behavior Coding Examples

Acronym	*Short Name*	*Example*
INSR	Full Scale Not Read	I: Then how would you consider your general health over your entire lifetime? *R: Mm, Laugh-R, I'd say about uh, I'd say good.* I: Were there ever periods--any periods of time where your health might have been um, less than good, or fair or--? *R: Well, this really just little minor things, um, uh Laugh-R one little case of stress here and there when I had abdominal problems, (*M-pos*) other than that, no.* I: So, that was just a temporary thing? *R: M-pos.* I: So you would consider, during your entire lifetime then you would consider your health to be good? (*Yes)* Is that correct? *R: M-pos.* I: Then the next question that we have has to do with um your smoking history.
RDK	Resolved DK	I: Okay, so you would say good, for this time period right now. *R: Yeah, you know, well now, I'm pretty--well, hold it, I--well, I don't know how--I don't know how to measure good, maybe that's--.* I: Well, I have excellent--. *R: Because I got the--uh--uh--um--wait--no--uh--uh--um---being treated for blood pressure is normal, but it's chemically normal. (*M-pos*) And, uh--diabetes. So, I--I guess [***Unintelligible***] call it good, but, you know…* I: Well, the range that I have is excellent, very good, good, fair, and poor. *R: I'll say fair. (*All righty*) At this stage.*
UDK	Unresolved DK	R: *I don't even remember the year when I broke all my ankle bones (hmmm) Because then I got a medal plate and 2 screws in my hip. I don't what year that was put in.* I: Wow, hmm, that sounds painful. R: *All on the same side.* I: [pause] Yeah … R: *I don't remember the year though* I: Let's just go ahead and skip to the next one then. Did you ever smoke cigarettes?

Acronym	*Short Name*	*Example*
DIR	Directive	I: And then I'm interested in knowing about different, um, periods in your life in which you may have felt one of these various five choices here: excellent, very good, good, fair, or poor. Did you have periods in your life when you felt one way or another, or. *R: I always had pretty fair, I guess, I'd say.* I: Fair your entire life? *R: M-pos*. I: Did you have illnesses when you were young?
NPR	No Probe	R: *Oh, I'd say it was excellent up through 30, and then it was very good for another 20 years or 30 years maybe, and then it's been good since then.* I: All righty, almost done here. Have you ever smoked cigarettes?
DISI	Disability Reference (Interviewer)	I: Well then we'll just look at what your health status has been. Um, from the time you were a little boy until you had your heart attack, how would you say your health was?
DISR	Disability Reference (Respondent)	*R: I don't recall having any really terrible times. (Okay) Like everybody else, I've had the usual things after I got out of school, so to speak, and started working, you know. And all I can say I had—but I haven't had disabling ones. (M-pos) Like, um, the only surgery I ever had was a day surgery on a thyroid [Unintelligible]. (M-pos) And that was a overnight thing. (M-pos) In the hospital, I was hospitalized for the food poisoning once. I mean, it's just--(M-pos) I've been very fortunate.*

References

Alwin, D.F. 2009. "Assessing the Validity and Reliability of Timeline and Event History Data" in *Calendar and Time Diary Methods in Life Course Research*, eds. R.F. Belli, F.P. Stafford & D.F. Alwin, Sage, Los Angeles, pp. 277-301.

Axinn, W.G., Barber, J.S. & Ghimire, D.J. 1997. "The Neighborhood History Calendar: A Data Collection Method Designed for Dynamic Multilevel Modeling", *Sociological Methodology*, vol. 27, pp. 355-392.

Axinn, W.G. & Pearce, L.D. 2006. *Mixed method data collection strategies*, Cambridge University Press, Cambridge.

Baddeley, A. 1992, "What is autobiographical memory?" in *Theoretical Perspectives on Autobiographical Memory*, eds. M.A. Conway, D.C. Rubin, H. Spinnler & W.A. Wagenaar, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 13-29.

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

Belli, R.F. & Callegaro, M. 2009. "The Emergence of Calendar Interviewing: A Theoretical and Empirical Rationale" in *Calendar and Time Diary Methods in Life Course Research*, eds. R.F. Belli, F.P. Stafford & D.F. Alwin, Sage, Los Angeles, CA, pp. 31-52.

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

Belli, R.F., Smith, L.M., Andreski, P.M. & Agrawal, S. 2007. "Methodological comparisons between CATI event history calendar and standardized conventional questionnaire instruments", *Public Opinion Quarterly*, vol. 71, no. 4, pp. 603-622.

Bilgen, I. & Belli, R.F. 2008. "Do Different Interviewing Techniques & Different Interviewer and Respondent Characteristics Impact the Quality of Retrospective Reports?", *Sixty Third Annual AAPOR Conference*.

Brewer, W.F. 1994. "Autobiographical Memory and Survey Research" in *Autobiographical Memory and the Validity of Retrospective Reports*, eds. N. Schwarz & S.S. Sudman, Springer-Verlag, New York, pp. 11-20.

Conway, M.A., Wang, Q., Hanyu, K. & Haque, S. 2005. "A Cross-Cultural Investigation of Autobiographical Memory: On the Universality and Cultural Variation of the Reminiscence Bump", *Journal of Cross-Cultural Psychology*, vol. 36, no. 6, pp. 739-749.

Dasbach, E.J., Klein, R., Klein, B.E.K. & Moss, S.E. 1994. "Self-rated health and mortality in people with diabetes", *American Journal of Public Health*, vol. 84, no. 11, pp. 1775-1779.

Dijkstra, W. 2008, "Behavior Coding" in *Encyclopedia of Survey Research Methods*, ed. P.J. Lavrakas, Sage, Thousand Oaks, CA, pp. 53-55.

Dijkstra, W., Smit, J.H. & Ongena, Y.P. 2009. "An Evaluation Study of the Event History Calendar" in *Calendar and Time Diary Methods in Life Course Research*, eds. R.F. Belli, F.P. Stafford & D.F. Alwin, Sage, Los Angeles, pp. 277-301.

Dykema, J., Lepkowski, J.M. & Blixt, S. 1997. "The Effect of Interviewer and Respondent Behavior on Data Quality: Analysis of Interaction Coding in a Validation Study" in *Survey Measurement and Process Quality*, eds. L.E. Lyberg, P.P. Biemer, M. Collins, et al, Wiley, Hoboken, NJ, pp. 287-310.

Fleiss, J.L., Levin, B. & Paik, M.C. 2003. *Statistical methods for rates and proportions*, 3rd edn, Wiley, Hoboken, NJ.

Freedman, D., Thornton, A., Camburn, D., Alwin, D.F. & Young-DeMarco, L. 1988. "The Life History Calendar: A Technique for Collecting Retrospective Data", *Sociological Methodology*, vol. 18, pp. 37-68.

Groves, R.M. 1989/2004. *Survey errors and survey costs*, Wiley, Hoboken, NJ.

Groves, R.M. 1991/2004. "Measurement Error Across Disciplines" in *Measurement Errors in Surveys*, eds. P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz & S.S. Sudman, Wiley, Hoboken, NJ, pp. 1-28.

Groves, R.M., Fowler, F.J., Jr., Couper, M.P., Lepkowski, J.M., Singer, E. & Tourangeau, R. 2004. *Survey Methodology*, 1st edn, Wiley, Hoboken, NJ.

Halbwachs, M. 1980. *Collective memory*, Harper & Row, New York.

Means, B. & Loftus, E.F. 1991. "When Personal History Repeats Itself: Decomposing Memories for Recurring Events", *Applied Cognitive Psychology*, vol. 5, pp. 297-318.

Oksenberg, L., Cannel, C. & Kalton, G. 1991. "New Strategies for Pretesting Survey Questions", *Journal of Official Statistics*, vol. 7, no. 3, pp. 349-365.

Pascale, J. & McGee, A. 2008. *A Multi-Method Evaluation of the Use of an Event History Calendar*, U.S. Census Bureau, Washington, DC.

Robinson, J.A. 1986. "Temporal reference systems and autobiographical memory" in *Autobiographical Memory*, ed. D.C. Rubin, Wiley, Cambridge, pp. 159-188.

Schaeffer, N.C. & Maynard, D.W. 2002. "Occasions for Intervention: Interactional Resources for Comprehension in Standardized Survey Interviews" in *Standardization and tacit knowledge: interaction and practice in the survey interview*, eds. D.W. Maynard, H. Houtkoop-Steenstra, N.C. Schaeffer & J. van der Zouwen, Wiley, New York, pp. 261-280.

Schwarz, N. 1996. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*, Erlbaum, Mahwah, NJ.

Singer, J.D. & Willett, J.B. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, Oxford University Press, Oxford.

Suchman, L. & Jordan, B. 1990. "Interactional troubles in face-to-face survey interviews", *Journal of the American Statistical Association*, vol. 85, no. 409, pp. 232-241.

Thorndike, R.M. 2005. *Measurement and Evaluation in Psychology and Education*, 7th edn, Pearson Prentice Hall, Upper Saddle River, NJ.

Tourangeau, R., Rips, L.J. & Rasinski, K. 2000. *The Psychology of Survey Response*, Wiley, New York.

Tulving, E. 2002. "Episodic Memory: From Mind to Brain", *Annual Review of Psychology*, vol. 53, pp. 1-25.

Upton, G. & Cook, I. 2006. *Oxford Dictionary of Statistics*, 2nd edn, Oxford University Press, Oxford.