COMPARISON OF COMPUTERIZED EVENT HISTORY CALENDAR AND QUESTION-LIST INTERVIEWING METHODS: A TWO-YEAR HOSPITALIZATION HISTORY STUDY

Mario Callegaro, Mandi Yu, Fei-Wen Cheng, Erik Hjermstad, Dan Liao, Robert F. Belli
University of Nebraska-Lincoln; Gallup Research Center 200 N. 11th Street, 4th floor, Lincoln, NE 68508-0214
Contact author: Mario Callegaro. Email: mca@unlserve.unl.edu

Keywords: Retrospective Reports, Standardized Interview, Hospitalization History, Validation Data, Computerized Event History Calendar

Introduction

Retrospective reports are often used by survey organizations in order to collect data on health related events. The majority of studies in the field of heath utilization are conducted using a standardized questionnaire administered either face-to-face or by telephone. Survey researchers classify these kinds of questions as behavioral or factual questions, as opposed to attitudinal or opinion reports (Sudman & Bradburn, 1974). Behavioral questions have, in principle, a “true value” and the report of the respondent can be checked against it to determine its accuracy. Figure 1 depicts the general theoretical framework used in discussing the studies reviewed in this paper

Reports of health related events can be correct, meaning that the report is identical to the true value of that behavior, or they can be underreported or overreported. Underreporting can occur involuntary due to forgetting or telescoping, or voluntarily because of embarrassment or social desirability. By voluntarily, we mean that the original answer is edited and/or distorted on purpose before being communicated to the interviewer. By telescoping we mean that the original event can be recalled as belonging outside the reference period, forward, or backward in time, thus creating underreporting. Overreporting can be voluntary for the above reasons, or can be due to telescoping, in which, events that occurred outside the reference period are brought into the reference period. Forward telescoping is when respondents allocate events to more recent time periods than when they actually occurred. Backward telescoping is when respondents allocate events to more remote time periods than when they actually occurred.

Variables that affect accuracy of reports

Cannell and colleagues conducted one of the first studies with the capability to compare survey data with hospital records (Cannell, Fisher, & Bakker, 1965). Respondents with higher education, more severe conditions, longer stay, and surgical cases had less underreporting. The percent of underreporting of hospitalization events fluctuated between 2.5 and 6% in the range of 10 to 40 weeks between discharge and interview. Underreporting increased very rapidly up to 42% one year after the event. The percent of hospital episodes underreported was also a direct function of the threat or the embarrassment of the single events. Percent of underreporting was 10% for non-threatening health episodes, 14% for somewhat threatening episodes and 20% for highly threatening hospital episodes. Similar results were found by Balumuth (1965). Interestingly, traditional background variables such as gender, age, socio-economic status, and ethnic background of the respondents did not have any significant effect on the accuracy of the recall.

Other researchers concentrated their attention on studying the impact of the type of condition on the accuracy of the report. Two studies detected underreport: Madow (1973) found that conditions of less subjective importance, with little impact on the subjects are most likely to be underreported. Jabine (1987) observed that events associated with chronic conditions and usage of health care resources are more likely to be underreported. Two other studies detected overreport: Loftus and colleagues (1990) found that when asking for a specific physical examination such as blood pressure check, mammogram, or child physical exam, respondents overreported by a median of 3.7% across those three procedures. The question referred to the past two months before the interview. However, when the same question was asked for a six-month reference period, the median of overreporting rose to 12.5%. Findings in the same direction are reported by Sudman et al. (1994).

Survey researchers have noted the prevalence of forward telescoping over backward telescoping (Sudman & Bradburn, 1973). Forward-telescoping was found to be more common than backward-telescoping for serious conditions, whereas minor conditions exhibited more backward-telescoping (Means, Nigam, Zarrow, Loftus, & Donaldson, 1989). Loftus, Smith, Klinger, & Fiedler (1992) investigated the conditions leading to under or overreporting of medical conditions. On one question they asked to recall *specific visits* made over the previous 12 months with the outcome of having more than 60% of actual visits underreported. In another question they asked to estimate the *total number* of visits in the last 12 months. The underreport was reduced to 23%, on average. On the contrary, when people were asked to report if they had a specific procedure, i.e. a flu shot, they overreported by 13%.

Q-list format and EHC format

A majority of the questions used in the reviewed studies can be classified as quantitative autobiographical questions. In this case the respondents are performing five cognitive tasks in order to answer the question: comprehend the meaning of the question, retrieve the particular instances of the requested behavior(s), determine if the event(s) they just retrieved from memory occurred during the reference period, judge the quality of the retrieved behavior(s), and formulate the response (Schwarz, 1990). "Respondents are likely to begin with some fragmented recall of the behavior under study and to apply various inference rules to arrive at a reasonable estimate" (Schwarz, 1990, p.107). Error can arise at each stage, but the most demanding stage is the retrieval process, which is a function of the extent of the reference period and the number of events that happened during that period.

The standardized question list (Q-list), has many advantages. There is a reduction in errors arising from change in question wording (Fowler and Mangione, 1990), a reduction of interviewer variability, a minimization of training, and a reduction in the length of the interview, which leads to a reduction in costs (Groves, 1989). Moreover, a closed question format is easier and faster to code.

At the same time, the standardized Q-list has some drawbacks. Standardization and a closed question format reduce, if not eliminate, the conversational rules that assure the understanding between the interviewer and the respondent. In other words the equation: *sameness of words = stability of meaning* is not a guarantee for the comprehension of the question or, ultimately, for the validity of the measurement instrument. In order to fully understand the question, conversational flexibility should be introduced into the standardized interview (Suchman and Jordan, 1990). The real challenge is to embed conversational flexibility into a survey interview without increasing measurement error so the validity of the research can be increased (Schober and Conrad, 2002).

There is evidence that for autobiographical events, and for the case of health reports, interviewer departure from the scripted questions does not adversely affect the quality of the reports (Belli and Lepowski, 1996, Belli, Lepowski and Kabeto, 2001). Methodologies that can focus on overcoming retrieval and comprehension difficulties instead of strictly adhering to a standardized survey question list format (Q-list), may improve the quality of factual reports (Fowler, 1995, p.180).

Potentially, one of these methodologies is the Event History Calendar (EHC) approach (Freedman et al, 1988). The EHC interview utilizes timelines, decomposing strategies and personal and official calendar landmarks (the birth of a child, national holidays, and historical events). The entire interview is organized around the calendar of the time reference period under investigation. The interviewer guides the respondent in filling in the time lines, starting with remembering personal landmarks. After filling in the personal landmarks, other domains are filled in such as residence history, household composition, employment and unemployment history, and education. The interviewer allows respondents' flexibility in the use of retrieval strategies, such as moving forward in time, backward in time, or across contemporaneous events. Every event or "spell" is entered into the time lines with a starting date and an ending date. In the current research, thirds of months were chosen as units. Each spell can happen before, after, or contemporaneously with another spell. For example, a relocation event can be contemporary with a new job event. The calendar and time lines can also be made visible to the respondent. The entire process of compiling the calendar focuses, by its nature, on coherence, sequential organization, and the reduction of missing data. Some events are unlikely to happen together, such as having two or three jobs at the same time without a plausible explanation. Every event should happen after the preceding one and before the following one. For example, you are unemployed, you look for a job, and when you find it, you are employed. Lastly, each event in a time line should be near another event without missing data (something must have happened in that time frame). Interviewers follow a script where the order of the questions is not predetermined, but adapts to the respondent's recollection. For instance, if the respondent starts with an event of unemployment, the interviewer will try to get the most precise start date and end date, and will ask about the event that followed unemployment.

EHC enhances the quality and the precision of retrospective reports because it matches three basic types of mechanisms that are associated with the structure of autobiographical memory (Belli, 1998): "top down", sequential and parallel cueing. According to Barsalou (1988), and Conway (1996), autobiographical memory is organized hierarchically where at the top exist memories of lifetimes periods, arranged in thematic domains connected to the self-concept of the person (family, job, education, etc.). For each domain temporal changes and updates are stored (for example the transition from student to worker). In the middle of the hierarchy reside memories for general events and at the bottom, memories for specific events (see also Conway and Rubin, 1993 for an extended description). Top down cueing taps into the relationship between the top of the hierarchy and its bottom, where, for example, general events are nested into a self-concept domain. Sequential cueing taps into the chronological sequence of events that belong to the same domain. Parallel cueing taps into the links that exist among domains and themes.

The flexible interviewing style of the EHC method allows for the proper memory cues and adapts itself to the current interview. The respondent feels more engaged in the conversation because each question is "personalized" or tailored. Furthermore, the entire interview is conducted in a narrative style, which is more compatible with the manner that autobiographical memory is structured (Brown and Schopflocher, 1998; Schank and Abelson, 1995) than a standardized survey interview. Flexibility, adaptability and tailoring also enhance comprehension (Schober and Conrad, 2002) by allowing the full resources of communication in a collaborative construction of meaning (Suchman and Jordan, 1990).

There is also another by-product of the EHC technique that enhances the quality of the reports. During an EHC interview, it may be more difficult to satisfice. When faced with survey questions, people can use a satisficing approach (Krosnick, 1991) in order to provide a plausible answer. The time line of the calendar demands for coherence and precision, discouraging guessing and reporting plausible answers. It also requires that the reported information is exhaustive, and, in its computerized version, highlights inconsistencies giving feedback to the respondent (Belli, 2000). Moreover, the narrative style and the tailoring of the EHC method enhance the motivation, engagement, and interest of the respondent.

Studies comparing the quality of EHC and question list methodologies, using validation data as external criteria, showed better quality retrospective reports in the EHC condition. In the first study of this kind, Belli, Shay and

Stafford (2001) showed that EHC reports were more precise on moves, income, weeks unemployed, and weeks missing work resulting from self illness, the illness of another or the combination of the two. They also found that the length of the interview was not significantly different between the two data collection methodologies. Respondents of the Panel Study of Income Dynamics (PSID) were randomly assigned to one of the interview methods and asked about social and economic behaviors that happened during the previous two year of the interview. Since the respondents were panel members, the data previously collected were used as standards of comparison. Yoshihama and colleagues (in press) found that the EHC method facilitated the recall of domestic violence victimization. In particular, the EHC technique elicited more reports of lifetime experiences of intimate partner abuse (93-96%) than the standard question list format (61-64%). EHC was especially precise in revealing abuses that occurred early in respondent's lives.

The present study is the first of its kind to compare retrospective health utilization reports with hospital records validation data using a computerized event history calendar technique. The Computer Assisted Interview EHC software (C-EHC) used in this study is copyrighted to the Regents of the University of Michigan. The instrument is built to reflect the structure of autobiographical memory. There are hierarchical timeline domains such as residence, employment, unemployment and landmark events presented as *tabs* in a left to right order. The smallest unit is the third of a month. The software is also designed to flag contradictory information, for example reporting to look for a job and being employed at the same time. For an extended description of the software see Belli (2000). In our particular study, in the monitor condition, the respondent was allowed to see the computer screen and, at the same time, a paper calendar was available.

Population and sample

The target population in this study consisted of residents of Washtenaw County, Michigan ages 40 to 75 years old who were registered in the university of Michigan Health system and had been hospitalized at least once between October 2000 and March 2001.

The sample was recruited using invitation letters asking whether residents were interested in participating in a research study which was associated with hospitalization, surgeries, and doctor's office visits. The respondents were self-selected on a first-come, first-serve basis and they received an incentive of 50 dollars for completion of interviews. Overall, 600 letters were sent out, and 82 people were interviewed from February to May 2002. The initial two interviews were not usable due to software application problems.

The medical records for 74 participants of the valid sample (N=80) were obtained to validate their EHC and Q list responses for the health care utilization time line. The demographics were also included in the medical records. Two cases with EHC instruments were lost due to technical problems during data transfer. One respondent showed extreme discrepancy between his report and medical records; that observation was treated as an outlier and removed from the analysis. The effective sample size was 71 for statistical tests. Among those 71 participants included in the analysis, 53.5% are men and 46.5% are women, 80.3% are Whites, 15.5% are African Americans and 4.2% are of other races. The average age of the participants was 56.9 years. Hospital records are assumed to be the "true" value but clerical and software errors are always possible (Holth, 1998).

Study 1

This study was designed to test different methods of asking people about their health care utilization, and to examine which methods lead to the higher quality retrospective reports. The response accuracies of hospitalization with computerized EHC and standardized paper-pencil questionnaire instruments were examined by comparing the responses with the corresponding medical records.

This experimental design included two within-subjects factors: method of interview (EHC instrument and Q-list instrument) and year (2000 and 2001), as well as one between-subjects factor, the order of presentation of interview methods (Q-list first or EHC first).

During the same interview, all participants were measured by both computerized EHC and paper-pencil Q-list instruments about their hospitalizations from 2000 through 2001. The order effect was controlled by random assignment, with one group administrated EHC instrument first and Q list questionnaire second, and the other went with Q list first and EHC instrument second. The two groups of respondents were comparable demographically. The 32 respondents (average age 56 years) whose first interview was EHC consisted of 18 women and 14 men, 26 white whites and 4 African Americans (2 for other races). The other 39 respondents (average age 57.7 years) whose first interview was Q-list consisted of 20 women and 19 men; 31 white and 7 African Americans (1 of other races)

Data processing

Two dependent variables involving hospitalization were analyzed in this study: number of annual inpatient hospitalization visits (hospitalized times) and total number of annual nights in the hospital (hospital nights). In Q-list interviews, the two variables were obtained directly from the questionnaire. For EHC, these two variables were computed based on the data extracted from the EHC software. When we could not obtain the hospital name, we treated that case as belonging to The University of Michigan Health system.

In the C-EHC method, thirteen percent of the respondents could not report (or were not asked) the number of nights for at least one hospitalized visit. To deal with this item nonresponse problem, we calculated the average number of nights per number of thirds of a month for those EHC cases in which we had both nights and thirds (average = 3.37 nights per third of month) and then for those EHC cases that did not have nights, we imputed by multiplying the number of thirds of a month by 3.37.

According to the repeated measure analysis (between subjects factor: order; within subject factor: year), mean hospitalized times (F_{171} =.094, p=.760) and hospitalized nights ($F_{1,71}$=.241, p=.625) are comparable across the two order groups. On average, respondents in both order groups were hospitalized once and stayed hospitalized around 4 nights per year.

Three measures of response accuracies were tested: (1) the mean raw differences between the reported hospitalized times and those in the medical records; (2) the mean absolute value of the differences between the reported hospitalized times and those in the medical records; and (3) the correlation between reported hospitalized times and those in the medical records (Burton and Blair, 1991). Mean raw differences can reveal the directions of the overall response errors, where negative values indicate underreports and positive values suggest over-reports (Cannell, Fisher, and Baker, 1965). The second measure, mean absolute difference, does not allow positive and negative response discrepancies to cancel each other, thus telling the magnitudes of the responses errors. For the third measure of accuracy, both Pearson's moment correlation and Spearman's Rho correlation were computed between the reported hospitalized times and medical records across respondents within each order group.

All three measures are reported in Table 1, which includes the mean raw differences, and absolute differences of hospitalized times by year and interview method respectively, as well as the correlation coefficients for Pearson's r and Spearman's Rho.

Differences among the two order groups in terms of the first two measures of accuracy are analyzed using the repeated measures analysis of variance, a 2 (year; 2000, 2001) ×2 (method; C-EHC, Q-list) ×2 (order; C-EHC first – Q-list second, Q-list first – C-EHC second) ANOVA of two within-subjects factors (year and method) and one between-subjects factor (order). All the effects of the mean raw differences and mean absolute differences were not significant at .05 level. Considering the small cell sample sizes, on an exploratory analysis basis, the main effect of method collapsing over order of administration ($F_{1,69}$ =3.718, p=.058), was found to be significant at the .10 level. EHC interview method (marginal mean=.00) produced lower mean raw response errors than the Q-list (marginal mean=.09). The EHC performed better than the Q-list in aiding recall of hospitalized times that happened during the past two-year period. Respondents showed a tendency to over report their health care utilization with the Q-list interview method. This result supports the hypothesis that the EHC interview method is more likely than the Q- list to lead to better quality retrospectives report on health care utilization**.**

One-sample T tests were used to test the significance of the mean raw and absolute differences for each combination of factors: method, year and order. We did not observe any overreport or underreport within each combination of factor method by year by order, given all cell means for raw differences in hospitalized times were not significant from zero. It is interesting to notice that all their corresponding absolute differences were found to be significant. This is due to the fact that overreporting and underreporting can cancel out each other when computing a mean. The absolute value solves the problem, but at the same time, the direction of the bias is lost. The overall levels of accuracy exhibited by the correlations are very high. The z tests of independent correlation coefficient comparisons did not show any differences among order groups.

The same measures of accuracy were used to analyze the responses on hospitalized nights. The statistics for the mean raw and absolute differences of hospitalized nights by year and method, as well as the correlation coefficients for Pearson's correlation and Spearman's Rho are shown in Table 2.

No significant effects were found for the repeated measures analysis. All the mean raw differences for each combination of method by year by order were not significant. However, given the small sample size, it should be noticed that there is a tendency that people were more likely to underreport hospitalized nights. The significant one-sample T tests for all cells mean absolute differences of hospitalized nights revealed the existence of overall report error at the individual level.

Table 3 also shows that the hospitalized night reports correlate strongly with medical records. There is a significant interaction effect of year by method on the correlations between reports and records. When focusing on the first interview, either EHC or Q-list, in 2000, EHC responses exhibited stronger relationship with records (r =+.954, n=32, p<.05, two tailed) than Q list responses (r =+.799, n=39, p<.05, two tailed). However a reverse pattern was observed in 2001, with a stronger accuracy of responses found with Q list (r =+.928, n=39, p<.05, two tailed) than EHC instruments (r =+.805, n=32, p<.05, two tailed). Both differences were significant (Fisher's z=3.121, p=. 0018 and z=-2.129, p=.0333), which provides support for the hypothesis that EHC performs better than Q-list in a more remote year, while Q-list performs better in a more recent year. In this study, remote time period refers to health care events that happened in 2000, roughly 1 to 2 years previous the interview data, and recent time period refers to those in 2001, roughly up to 1 year before the interview. In order to control for outliers, Spearman's Rho was also computed. However the pattern observed using Pearson's r was not found.

Study 2

In study 2, the between-subjects factor, view calendar condition, is included to test the differences of response accuracies between the view calendar and non-view calendar condition with the EHC instrument. By random, a half sample was selected to view the computerized EHC instrument and a paper calendar of the reference period. No calendar (computerized or paper) was available for the other half sample in the EHC interviews, which simulated telephone interviewing. Similarly, the repeated measures analysis is used with one within-subjects factor, year, and one between-subject factor, monitor (view calendar vs. non-view calendar).

Table 4 shows the mean raw and absolute response errors on hospitalized times and hospitalized nights for view calendar and non-view calendar groups as well as their respective correlations with medical records.

The interaction between monitor and year was significant on mean absolute errors of the hospitalized times reports. Mean absolute response error on hospitalized times for year 2000 was larger for respondents under the non-view calendar condition (mean= .42) than the view calendar condition (mean =. 16). Consistent results were found with Pearson's correlations (Fisher's z=2.845, p=. 0044). This result supports the hypothesis that viewing the calendar could improve the accuracy of people's retrospective reports of hospitalized times. However, the significant effect of viewing the calendar on the improvement of response accuracies did not extend to the absolute responses errors in 2001. The improvement is more significant in retrospective reports about health care events that happened during a more remote time period than a more recent time period. Interestingly, people reported recent hospitalized nights more accurately without viewing the calendar than when viewing the calendar, according to the comparison between Pearson's r's (in 2001, Fisher's z=-2.104, p=.0354). For reports on health care events that happened in the remote year, the monitor effect on accuracy improvement is supported by Spearman's Rho correlations between hospitalized nights and records in 2000 (Fisher's z=2.529, p=.0114).

Results discussion

Overall, the findings in this study had important implications for the potential to implement EHC as an interviewing method to improve response accuracies. First, EHC produced more accurate reports on health care utilization that happened during the past two-year period than standard questionnaires. Another noteworthy attribute is the performance of EHC to improve the precision of retrospective health care utilization reports for events that happened during remote time periods. Third, allowing respondents to view the computerized calendar could help improve the recall of health events happened during remote time periods. Last, Q-list performed better for more recent years.

Study limitations

There are several reasons for the failure to find substantive differences between C-EHC and Q-list methodologies. Overall our data resulted in high levels of accuracy in both Q-list and C-EHC conditions, even when respondents reported hospitalization events that occurred over two years ago. This led to a ceiling effect where the differences between the two methodologies are blurred. Our particular sample also had an "easy" heath history, with less than one hospitalization event per year, on average, and a mean of roughly 4 overnights of stays per year. For these reasons, our analyses are in contrast with those of Cannell et al. (1965), where they found underreporting of 42% after one year of the hospitalization event(s). One possible reason is because respondents were aware of the topic of the interview beforehand due the to invitation letter. People are more likely to rehearse and recall before interviews in self-recruiting surveys, thus mediating method effect and year effect. Cannell and colleagues (1965) prevented the respondents from rehearsing hospitalization information because they were able to conduct the interview without telling them the real purpose of the study, and, subsequently, to obtain validation data from the hospitals. On the positive side this study can give initial support to the usage of invitation letters not only to increase the response rate, as the majority of the literature demonstrates, but also to increase the precision of the recall. However in our study all respondents did receive an invitation letter and we did not have a control group; for this reason, a further experiment should validate these initial findings.

A second limitation is that the respondents were selected on a first-come-first served basis, leading us to expect that the more motivated have answered the request. Highly motivated respondents are likely to be optimizers (Krosnick, 1991), and because of that, fully devote all resources in the response process, trying to be as precise as possible.

Another explanation is the design of the C-EHC for this particular study. C-EHC was not intended to specifically ask respondents to report exactly where they were hospitalized. The location clue can facilitate respondents in remembering the health-related events (Means et al., 1991).

Conclusions

The current study is the first experiment comparing Q-list with C-EHC where retrospective reports were matched against external, independent validation data. The mentioned limitations may have attenuated the expected results but, nonetheless, C-EHC proved to be more effective and precise than Q-list methods to collect retrospective health utilization data for more remote years. We also expect EHC methodology to be more effective if the study period is increased to more remote years, for example 3 or 4 years since the interview. In conclusion the, C-EHC methodology appears to have benefits in comparison to the Q-list, especially when respondents can visualize the reference period calendar.

The project was funded jointly by the National Institute on Aging, and National Institute of Child Health and Human Development, 1R01AG17977

Table 1 Response accuracies on hospitalized times by method, year and order

		Mean raw response error		Mean absolute [a] response error		Correlation between reports and validation records [b]			
						Pearson's R		Spearman's Rho	
Order of administ.	N	Year 2000	Year 2001	Year 2000	Year 2001	Year 2000	Year 2001	Year 2000	Year 2001
EHC 1st	32	.00	.00	.25**	.31**	.619**	.790**	.676**	.783**
Q-List 1st	39	-.03	.13	.33**	.49**	.700**	.683**	.753**	.801**
EHC 2nd	39	-.10	.10	.31*	.41**	.573**	.780**	.655**	.765**
Q-List 2nd	32	.09	.19	.28**	.50**	.708**	.671**	.750**	.670**

1 Note: All tests of significance include only order group.
[a] *One Sample T test that the value of Mean Absolute Response Error is significantly different from zero at .05 level, two tailed,
** T test is significance at .01 level, two tailed.
[b] **Correlation is significantly greater than 0 at .01 level, two tailed.

Table 2 Response accuracies on hospitalized nights by method, year and order

		Mean raw response error		Mean absolute [a] response error		Correlation between reports and [b] validation records			
						Pearson's R		Spearman's Rho	
Order of administ.	N	Year 2000	Year 2001	Year 2000	Year 2001	Year 2000	Year 2001	Year 2000	Year 2001
EHC 1st	32	-.298	-.473	2.054**	1.339**	.954**c	.805**d	.735**	.809**
Q-List 1st	39	-.188	-.205	1.692**	1.436**	.799**c	.928**d	.876**	.893**
EHC 2nd	39	-.827	-.729	1.769 **	2.278**	.745**	.716**	.738**	.790**
Q-List 2nd	32	-.484	-.078	2.109**	1.453**	.942**	.807**	.674**	.777**

Note: All tests of significance include only order group.
[a] **One Sample T test that the value of Mean Absolute Response Error is significantly different from zero at .05 level, two tailed.
[b] **Correlation is significantly different from 0 at .01 level, two tailed.
[c] Fisher's Z test of difference between two independent correlation coefficients: z=3.121, p=.0018, two tailed.
[d] Fisher's Z test of difference between two independent correlation coefficients: z=-2.129, p=.0333, two tailed.

Table 3 EHC response accuracies on hospitalized times and hospitalized nights by year

		Mean raw response error		Mean absolute [a c] response error		Correlation between reports and validation records			
						Pearson's R		Spearman's Rho	
	N	Year 2000	Year 2001	Year 2000	Year 2001	Year 2000	Year 2001	Year 2000	Year 2001
Hospitalized Times									
Monitor	38	-.05	.11	.16*d	.42**e	.820**f	.730**	.756**	.745**
Non Monitor	33	-.06	.00	.42**d	.30**e	.421*f	.895**	.592**	.812**
Hospitalized Nights									
Monitor	38	-.4637	-.9071	1.3542**	1.6639**	.875**	.616** g	.872**h	.830**
Non Monitor	33	-.6924	-.3848	2.5803**	2.1097**	.900**	.846** g	.612**h	.776**

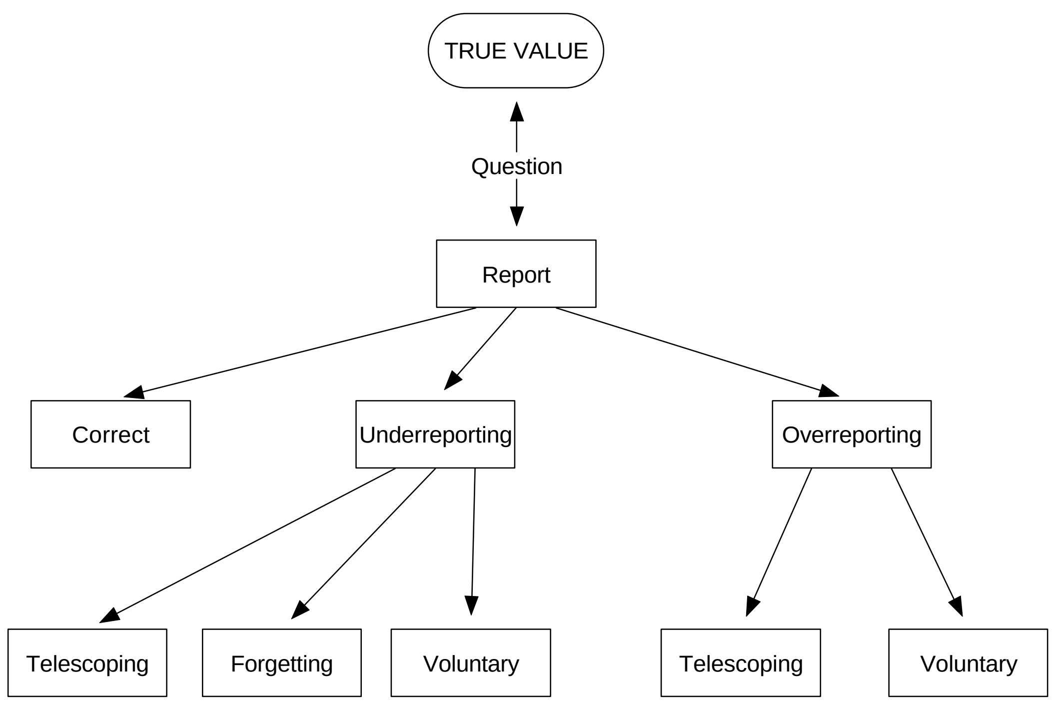
Note: All tests of significance include only order group.
[a] *One Sample T test that the value of Mean Absolute Response Error is significantly different from zero at .05 level, two tailed,
**T test is significant at .01 level, two tailed.
[b] *Correlation is significantly different from 0 at .05 level, ** Correlation is significantly different from 0 at .01 level, two tailed.
[c] The interaction effect between monitor and year on Mean Absolute Response Error is significant at .05 level: F (1, 69) =5.161, p=.026.
[d] The simple effect of monitor when year was fixed at 2000 is significant at .10 level: F (1, 69) =3.1805, p=.0789.
[e] The simple effect of monitor when year was fixed at 2001 is not significant at .10 level: F (1, 69) =.5616, p=.4562.
[f] Fisher's Z test of difference between two independent correlation coefficients: z=2.845, p=.0044, two tailed.
[g] Fisher's Z test of difference between two independent correlation coefficients: z=-2.104, p=.0354, two tailed.
[h] Fisher's Z test of difference between two independent correlation coefficients: z=2.529, p=.0114, two tailed.

Table 4 EHC response accuracies on hospitalized times and hospitalized nights by year

		Mean raw response error		Mean absolute [a c] response error		Correlation between reports and validation records			
						Pearson's R		Spearman's Rho	
	N	Year 2000	Year 2001	Year 2000	Year 2001	Year 2000	Year 2001	Year 2000	Year 2001
Hospitalized Times									
Monitor	38	-.05	.11	.16[*d]	.42[**e]	.820[**f]	.730[**]	.756[**]	.745[**]
Non Monitor	33	-.06	.00	.42[**d]	.30[**e]	.421[*f]	.895[**]	.592[**]	.812[**]
Hospitalized Nights									
Monitor	38	-.4637	-.9071	1.3542[**]	1.6639[**]	.875[**]	.616[** g]	.872[**h]	.830[**]
Non Monitor	33	-.6924	-.3848	2.5803[**]	2.1097[**]	.900[**]	.846[** g]	.612[**h]	.776[**]

Note: All tests of significance include only order group.
[a] [*]One Sample T test that the value of Mean Absolute Response Error is significantly different from zero at .05 level, two tailed, [**]T test is significant at .01 level, two tailed.
[b] [*]Correlation is significantly different from 0 at .05 level, [**] Correlation is significantly different from 0 at .01 level, two tailed.
[c] The interaction effect between monitor and year on Mean Absolute Response Error is significant at .05 level: F (1, 69) =5.161, p=.026.
[d] The simple effect of monitor when year was fixed at 2000 is significant at .10 level: F (1, 69) =3.1805, p=.0789.
[e] The simple effect of monitor when year was fixed at 2001 is not significant at .10 level: F (1, 69) =.5616, p=.4562.
[f] Fisher's Z test of difference between two independent correlation coefficients: z=2.845, p=.0044, two tailed.
[g] Fisher's Z test of difference between two independent correlation coefficients: z=-2.104, p=.0354, two tailed.
[h] Fisher's Z test of difference between two independent correlation coefficients: z=2.529, p=.0114, two tailed.

Figure 1 Classification of accuracy outcomes for heath related events



References

Balumuth, E. (1965). Health interview responses compared with medical records. *Vital and Health Statistics*, *2*(7), 1-74.

Barsalou, L. W. (1988). The connect and organization of autobiographical memories. In U. Niesser & E. Winograd (Eds.), *Remembering reconsidered: ecological and traditional approaches to the study of memory* (pp. 193-243). New York: Cambridge University Press.

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*, 383-406.

Belli, R. F. (2000). Computerized event history calendar methods: facilitating autobiographical recall. In *Pro-*

ceedings of the Section on Survey Research Methods. (pp.471-475). Alexandria, VA: American Statistical Association.

Belli, R. F., & Lepowski, J., M. (1996). Behavior of survey actors and the accuracy of response. In R. Warneke (Ed.), *Health Survey Research Methods: Conference Proceeding. DHHS Publication No. (PHS) 96-1013* (pp. 69-74). Washington DC: U.S. Government Printing Office.

Belli, R. F., Lepowski, J., M., & Kabeto, M. U. (2001). The respective roles of cognitive processing difficulty and conversational rapport on the accuracy of retrospective reports of doctor's office visits. In M. Cynamon, L. & R. A. Kulka (Eds.), *Seventh conference on health survey research methods. DHHS Publication No. (PHS) 01-1013* (pp. 197-203). Hyattsville, MD: U.S. Government Printing Office.

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

Brown, N. R., and Schopflocher, D. (1998). Event clusters: an organization of personal events in autobiographical memory. *Psychological Science* 9, 470–475.

Burton, S., and Blair, E. (1991). Task conditions, response formulation processes, and response accuracy for behavioral frequency questions in surveys. *Public Opinion Quarterly, 51*(1), 50-79

Cannell, C., F., Fisher, G., & Bakker, T. (1965). Reporting of hospitalization in the Health Interview Survey. *Vital and Health Statistics, 2*(6), 1-77.

Conway, M. A. (1996). Autobiographical knowledge and autobiographical memories. In D. C. Rubin (Ed.), *Remembering our past: studies in autobiographical memory* (pp. 67-93). New York: Cambridge University Press.

Conway, M. A., & Rubin, D. C. (1993). The structure of authobiographiocal memory. In A. F. Collins, S. E. Gathercole & M. E. Conway (Eds.), *Theories of memory* (pp. 103-137). Hove, Sussex: Lawrence Erlbaum.

Fowler, F. J. (1995). *Improving survey questions: design and evaluation*. Thousand Hoaks, CA: Sage.

Fowler, F. J., & Mangione, T., W. (1990). *Standardized survey interviewing*. Newbury Park, CA: Sage.

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

Groves, R. M. (1989). *Survey errors, and survey costs*. New York: Wiley.

Jabine, T. B. (1987). Reporting chronic conditions in the National Health Interview Survey: A review of tendencies from evaluation studies and methodological test. *Vital and Health Statistics*, 2(105), 1-54.

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

Loftus, E. F., Smith, K. D., Klinger, M. R., & Fiedler, J. (1990). A tale of two questions: Benefits of asking more than one question. *Public Opinion Quarterly, 54*(3), 330-345.

Loftus, E. F., Smith, K. D., Klinger, M. R., & Fiedler, J. (1992). Memory and mismemory for health events. In J. M. Tanur (Ed.), *Questions about questions. Inquiries into the cognitive bases of surveys* (pp. 102-137). New York: Russel Sage.

Madow, W. G. (1973). Net differences in interview data on chronic conditions and information derived from medical records. *Vital and Health Statistics, 2*(57), 1-58.

Means, B., Nigam, A., Zarrow, M., Loftus, E. F., & Donaldson, M. S. (1989). Autobiographical memory for health-related events. *Vital and Health Statistics*, 6(2), 1-37.

Means, B., Swan, G. E., Jobe, J. B., and Esposito, J. E. (1991). An alternative approach to obtaining personal history data. In P. B. Biemer, R. M. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman. *Measurements errors in surveys.* (pp. 167-183). New York: Wiley.

Holt, G., Feurer, I. D., & Schwartz, H. S. (1998). An analysis of concordance among hospital databases and physician records. *Annals of surgical oncology: the official journal of the Society of Surgical Oncology, 5*(6), 553-556.

Schank, R. C., and Abelson, R. P. (1995). Knowledge and memory: the real story. In *Advances in Social Cognition* (vol. 7), ed. Robert S. Wyer, (pp. 1–85). Hillsdale, NJ: Erlbaum.

Schober, M. F., & Conrad, F. G. (2002). A collaborative view of standardized survey interview. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer & J. van der Zouwen (Eds.), *Standardization and tacit knowledge. Interaction and practice in the survey interviews* (pp. 67-94). New York: Wiley.

Schwarz, N. (1990). Assessing frequency reports of mundane behaviors. Contributions of cognitive psychology to questionnaire construction. *Review of Personality and Social Psychology, 11*, 98-119

Suchman, L., & Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association, 85*, 232-241.

Sudman, S., & Bradburn, N. M. (1973). Effects of time and memory factors on response in surveys. *Journal of the American Statistical Association*, 68(344), 805-815.

Sudman, S., & Bradburn, N. M. (1974). *Response Effects in Surveys. A review and synthesis.* Chicago: Aldine.

Sudman, S., Warneke, R., Johnson, T., & O'Rourke, D. (1994). Cognitive aspects of reporting cancer prevention examinations and tests. *Vital and Health Statistics, 6*(7), 1-171.

Yoshihama, M., Gillespie, B., Hammok, A. C., Belli, R. F., & Tolman, R. M. (in press). Does the life history calendar method facilitate the recall of intimate partner violence? Comparison of two methods of data collection. *Social Work Research*