Developments in Survey Methodology
Anuška Ferligoj and Andrej Mrvar (Editors)
Metodološki zvezki, 15, Ljubljana: FDV, 2000

# Event History Analysis of Interviewer and Respondent Survey Behavior

James M. Lepkowski[1], Vivian Siu, and Justin Fisher[2]

**Abstract**

The survey interview may be viewed as a longitudinal sequence of conversational exchanges between an interviewer and a respondent. Interviewer and respondent behavior across interview exchanges reflect the dynamics of the interview and the mutual effects of interviewer behavior on respondents, and respondents on interviewer behavior. Event history models can be used to examine the timing of these behaviors and whether respondents modify the way they answer questions in response to interviewer behaviors, or whether interviewers modify their interviewing techniques in response to respondent behaviors.

A total of 297 interviews from a sample survey of members of a health maintenance organization in a metropolitan area in the United States were, with subject permission, tape recorded. Survey interviewers not participating in the survey interviews were trained to listen to the tapes and record the presence of approximately 30 different types of interviewer and respondent behaviors at each question asked in the interview. Respondent behaviors such as laughter during an exchange and interrupting the reading of the question are examined as events occurring during the interview using standard event history analysis methods. Cox proportional hazard models illustrate the association of respondent and interviewer characteristics on duration times. Duration times to interruption vary across gender and race of respondents and gender and race of interviewer, and by the occurrence of the event (i.e. first, second, third, etc.). The occurrence of a question characteristic as a time varying covariate for interruption is also examined. Interrupting question reading is shown empirically to be a function of exposure to questions that exhibit wording that is lengthy or contains numerous clauses that qualify the topic of the question.

[1] Institute for Social Research, University of Michigan, USA.
[2] Westat, Inc., Rockville, MD, USA.

# 1 Introduction

Situations arise that interrupt the interview. Interviewers are trained to respond then in ways that maintain the standardized format of stimuli and responses (Brenner, 1982, 1985; Sykes and Collins, 1992). They must remain non-directive and must avoid social interaction outside the task of interviewing (Brenner, 1982; Fowler and Mangione, 1990; Sykes and Collins, 1992). However, they are expected to maintain a socially effective interaction with the respondent. The interviewer must persuade the respondent to take part in the interview and then keep the subject engaged until it is completed. The interviewer uses verbal or, in face-to-face interviews, non-verbal methods to motivate the respondent to report adequately.

The respondent, on the other hand, is not aware of her/his role in the process (Brenner, 1982). Respondents have no formal training for their role, and must be trained to respond to the survey questions in the standardized framework. Given the conversational setting of the interview, respondents may alter the stimulus-response sequence in any number of ways, including by requesting clarification, giving inadequate answers, or digressing on unrelated topics. The interviewer is trained to respond to these departures in ways that will return the interview to a standardized stimuli-response process as quickly as possible within the conversational setting.

Deviations from the standardized interaction have been studies extensively in an effort to understand the extent to which departures lead to reduced quality data. Researchers led by Cannell and colleagues at the University of Michigan in the 1960's and 1970's applied observational techniques to the survey interview in order to study deviations from the standardized goal. The observation process methodology was adapted from interaction process analysis that had been used to investigate small group processes and structure (Bales, 1950). The methodology, interaction or behavior coding, relied on direct systematic observation of the question and answer process. An important discovery was that survey interviewers frequently did not administer questions as specified, nor were other techniques used according to instructions (Brenner, 1982; Morton-Williams, 1979). Interviewer failure to conform to instructions or respondents to provide concise answers that the interviewer could code was evidence that a question or question sequence failed (Cannell et al., 1992; Morton-Williams and Sykes, 1984; Oksenberg et al., 1991). That is, deviations from ideal role behavior are related to characteristics of the questions. Thus, the three actors in the standardized interview, the interviewer, the respondent, and the questions, all have an effect on the completion of the survey interview.

While data generated from interaction or behavior coding of a survey interview has proven useful for diagnosis of questions that pose problems for interviewers and respondents, and for monitoring the performance of interviewing staff, it has been difficult to use behavior codes to study the nature of the interview. Complexity of the interaction among the three elements in the process has made it difficult to analyze behavior coding data. The shear amount of information is also a formidable barrier. For example, a one-hour interview might involve an average of 120 question-answer exchanges between an interviewer and a respondent. More than 1,000 behavior codes would typically be generated from such an interview. Coding extended sequences of behavior to understand the conversational dynamics of the process is difficult using basic statistical analysis tools such as cross-tabulations.

Interviewer behavior is expected to alter respondent behavior during the course of an interview. Further, some speculate that respondent behavior changes how interviewers conduct an interview; interviewers may tailor their behavior to adapt to respondent abilities and sensitivities discovered during the interview. An examination of the sequencing of interviewer and respondent behavior over the course of an interview might reveal whether and to what extent such altered behavior occurs.

Survival or event history models offer a method for studying the influence of two interrelated phenomenon over the course of process. Event history models provide a framework for examining whether an outcome phenomenon may be related to a time varying covariate. In the context of coded data from a survey interview, a specific interviewer or respondent behavior may be examined relative to the occurrence and timing of another interviewer or respondent behavior observed during the course of the interview. An event history model factors the time varying nature of the relationship between the two behaviors directly into the model.

It is the purpose of this paper to examine an application of event history models to behavior-coded data from a survey interview. This research is exploratory, uncovering limitations in the behavior coding process for applying the methodology, and limitations in the methodology when studying the interviewer-respondent interaction. While findings are preliminary, there is satisfactory evidence that event history models of behavior-coded data provide insight into the survey interview process.

# 2 Methods

The Health Field Study (HFS) is a validation survey concerning health and health care utilization. Conducted in 1993 by the Survey Research Center at the University of Michigan for the National Center for Health Statistics, the primary objective was to evaluate questions currently in use or under consideration for the redesign of the National Health Interview Survey (NHIS). While many of the items selected for the HFS are replications or adaptations of existing NHIS questions, the questions in our analysis were written specifically for the HFS. Forty-five interviewers, half of whom had no previous interviewing experience, were trained to administer the HFS.

A total of 2,006 respondents ages 14 years or older were interviewed in their homes on a variety of health-related topics. Respondents were drawn from membership of a Health Maintenance Organization (HMO) in a large metropolitan area. The sampling frame included individuals who had one or more health care utilizations during the previoius 16 months. Youth ages 14 to 17, African-Americans, and persons 65 years and older were over-sampled. Of 2,006 members of the sample, 970 (48.4%) are white, 945 (47.1%) are African American, and 99 (4.4%) are of other races. The mean age of survey respondents is 43 years (with a standard deviation of 18.5 years), and approximately one-half of the respondents are female. A total of 429 (21%) respondents had not completed high school, while 26% completed high school and 52% had some college level education.

Nearly all interviews were audiotape recorded, except those for which the respondent did not give permission or there was a failure in the recording process. Following data processing, a sample of tapes was selected for coding of verbal behaviors during the survey interview. Of the 1,834 tape recordings available, a systematic sample of 317 taped interviews was selected for a coding operation, controlling for respondent age, gender, and race and the date of interview. After deleting 21 taped interviews for which the number of audible questions was less than 30, a sample of 296 tape recorded interviews were available for coding.

We employed a small staff of survey interviewers (different from those who conducted the interviews) to code a variety of verbal and other behaviors from the audio taped interviews. Five coders, all with interviewing experience, were trained in groups to code the behaviors of interest. Group sessions combined with practice coding exercises were used to improve coding reliability. Follow up group sessions concentrated on resolving coding differences. A single coder was used on each tape (i.e. only one coder completed the codes for a given tape).

An *a priori list* of behaviors, which appear from previous research to best characterize the presence of cognitive difficulties in answering question objectives, were identified (Belli and Chardoul 1997; Belli and Lepkowski 1996; Belli, Weiss, and Lepkowski 1999; Fowler and Cannell 1996). The subsequent investigation focuses on two respondent behaviors, interruptions and respondent laughter. The code *interrupts* was assigned if the respondent interrupted with an answer during the initial reading of the question. Respondent or interviewer *laughter* is indicator of conversational behaviors that characterize personal attachment between interviewer and respondent.

Table 1 summarizes the coding scheme to which each question-answer exchange between interviewer and respondent was assessed. The coding process was designed to record the presence or absence of these behaviors for each question asking exchange. For convenience of reference, behaviors are grouped on the basis of interviewer question-asking behaviors (including repeating questions), respondent answering behaviors, interviewer probing and feedback behaviors, and the conversational behaviors of laughter and digressions.

**Table 1:** Verbal behavior codes, Health Field Study behavior coding subsample.

*Interviewer question-asking codes*

**Q1** Exact: reads exactly as written or makes insignificant changes.
**Q2** Significant changes: makes wording changes that can affect written question meaning.
**Q3** Other changes: verifies, states, or suggests an answer; reads non-applicable question; skips applicable question.
**Q6** Exact repeating of question
**Q7** Significant changes in repeating question

*Respondent answering codes*

**R1** Interruption: interrupts question with an answer.
**R2** Clarification: expresses uncertainty, requests question repetition, or seeks clarification.
**R3** Qualified response: qualifies answer with phrases such as about, I guess, maybe, etc.
**R4** Uncodable/inadequate response: response does not meet question objectives.
**R5** Standard codable response.
**R6** Expressions of don't know that occur before a final codable response is given.
**R7** Respondent corrects a response to a previous question

*Interviewer probing codes*

**P1** Adequate probing: probing is non-directive and sufficient.
**P2** Directive probing: at least one probe is directive.

*Interviewer feedback codes*

**F1** Acceptable short: neutral and appropriate short phrase (1-3 words) such as "Thank you."
**F2** Acceptable long: neutral and appropriate longer phrase such as "Thanks. That's useful information for our study."
**F4** Unacceptable short: offers short phrase that may indicate approval for the content of the response.
**F5** Unacceptable long: offers longer phrase that may indicate approval for the content of the response.
**F6** Unacceptable reward: approval for a don't know response, refusal, digression, interruption, or inadequate final answer. Includes a digression that follows a respondent digression.

*Interviewer conversational codes*

**X1** Interviewer introduces digression: digressions are verbal comments that are not directly related to satisfying question objectives
**X2** Interviewer laughs

*Respondent conversational codes*

**X8** Respondent digresses
**X5** Respondent laughs

The behavior codes were recorded as simple dichotomous indicators of whether they occurred. They were stored in variable length records, one record for each interviewer-respondent question-answer exchange, and a variable number of behaviors per exchange. Coding staff recorded behavior within an exchange in the order the behaviors actually occurred in the exchange. For example, the following sequence of behaviors occurred: the interviewer read the question exactly as written, but during the reading the respondent interrupted with a response; the interviewer repeated the question reading; the respondent gave a qualified answer that could not be coded directly; the interviewer probed adequately; the respondent gave a standard response; the interviewer provided a short acceptable feedback. The coding sequence for this exchange would be Q1, R3, Q6, R3, P1, R5, and F1.

Under formal event history modeling, time information about the occurrence of each behavior is required, if time to events is to be examined. No time stamp information was recorded during these face-to-face paper and pencil interviews. However, the total length of time for each interview was recorded to the nearest minute. As a proxy measure of time, each behavior was assumed to require the same amount of time, the total interview length in seconds divided by the number of behaviors occurring during the interview. Typically, each behavior was assigned a time value of 6-8 seconds. Time to an event or behavior is the cumulated number of behaviors of fixed length that had occurred prior to the behavior of interest.

Time was the occurrence of a behavior was needed both for the target event as well as a time varying covariate. Consider, for example, the relationship between the timing of a respondent interruption of a question reading (R1) and the time when the respondent was read a question that from question review was noted to be one likely to lead to an interruption. Time to the respondent behavior was coded by counting the number of behaviors preceding the interruption but since the last interruption (or beginning of the interview). The behavior count was multiplied by the average length in seconds of a behavior. Similarly, the start of the question reading of a potentially “interruptable” question was noted with respect to the behavior count since the last interruption or the beginning of the interview.

Considerable data processing is required to convert the sequenced behavior data stored in variable length records into formats required for event history model software. Macros were written in SAS to convert the behavior-coded data into time to event or discrete time hazard data suitable for analysis.

Further, a complete treatment of all 23 behaviors shown in Table 1 was not attempted. Two behaviors, interruption of the question reading (R1) and respondent laughter (X5) were selected for analysis in the present research. Event history models are presented only for interruption of the question reading, including an examination of the dynamic relationship between the reading of potentially interruptible questions and interruption of a question reading.

Descriptive measures of basic time to event data, such as survival curves of Kaplan-Meier estimates of conditional probabilities, was generated from the coded data. Model estimation was based on discrete time intervals rather than the continuous time coded from the behaviors due to uncertainty about the exact timing of behaviors. The interviews were divided into discrete time periods of 30 seconds each for model estimation purposes, with an average 1-hour interview generating 120 time intervals of data. The presence or absence of the target behavior (interruption of the question reading by the respondent, R1) in each 30-second interval was coded, as was the presence or absence or counts of time varying covariates (potentially interruptable questions). Discrete time hazard models were estimated using standard logistic regression models. (Continuous time hazard models have also been fit, but results are not shown here.)

Each behavior occurs multiple times in an interview. If the phenomena generating each behavior are statistically independent, the multiple occurrences of a behavior within an interview can be analyzed as independent events. An *a priori* expectation of independence across multiple occurrences is not warranted. Event history models fit to multiple occurrence data examined whether there were interactions of time varying covariates with the occurrence, indicating a lack of independence across occurrences.

# 3 Results

The frequency of each behavior obtained in the coding process for all 296 respondents across all questions asked is shown in the last two columns of Table 2. Some behaviors occur very infrequently, such as significant changes in question wording during a repeated asking of the question. Others, such as exact question reading, occur very often. The frequencies shown in Table 2 are similar to those observed in other studies of interviewer and respondent behaviors in surveys conducted by the Survey Research Center.

For purposes of measuring the reliability of code assignments between coders, two coders completed 24 tapes once each. Codes from these double-coded tapes were compared to assess inter-coder reliability. Kappa statistics ($\kappa$), which control for chance agreement in comparisons between coders, provide an estimate of inter-coder reliability (see Fleiss, 1973). Values of $\kappa$ between 0.21 and 0.40 are an indication of a fair level of agreement; values between 0.41 and 0.60 indicate moderate agreement; values between 0.61 and 0.80, substantial agreement; and values between 0.81 and 1.0, almost perfect agreement (Landis and Koch, 1977). Table 2 presents $\kappa$ values for the coded behaviors in the sample of 296 respondents. Almost all of the codes reached what would be considered as fair or moderate agreement. Only the interviewer behavior Q7 (making significant wording changes in a repeated reading of a question during an exchange) failed to reach an acceptable agreement level of $\kappa$. For purposes of the reported research, only interruptions of the question reading (R1) and respondent laughter (X5) are examined further. Both of these had moderate levels of observer agreement.

**Table 2:** Kappa values for interviewer and respondent behaviors (n=24 subjects from 296 total).

| Item | Kappa-Value | ASE[3] | Average | SD |
|---|---|---|---|---|
| **Q1Q1** | 0.511 | 0.028 | 0.946 | 0.141 |
| **Q2Q2** | 0.432 | 0.032 | 0.043 | 0.121 |
| **Q3Q3** | 0.467 | 0.099 | 0.012 | 0.042 |
| **Q6Q6** | 0.722 | 0.033 | 0.051 | 0.09 |
| **Q7Q7** | 0.181 | 0.157 | 0.002 | 0.016 |
| **R1R1** | 0.652 | 0.031 | 0.037 | 0.08 |
| **R2R2** | 0.783 | 0.023 | 0.096 | 0.127 |
| **R3R3** | 0.538 | 0.035 | 0.065 | 0.106 |
| **R4R4** | 0.390 | 0.033 | 0.115 | 0.19 |
| **R6R6** | 0.643 | 0.071 | 0.013 | 0.062 |
| **R7R7** | 0.769 | 0.092 | 0.011 | 0.038 |
| **P1P1** | 0.676 | 0.022 | 0.179 | 0.242 |
| **P2P2** | 0.283 | 0.050 | 0.025 | 0.075 |
| **P3P3** | 0.303 | 0.066 | 0.008 | 0.032 |
| **F1F1** | 0.700 | 0.016 | 0.282 | 0.241 |
| **F2F2** | 0.441 | 0.064 | 0.041 | 0.083 |
| **F4F4** | 0.314 | 0.045 | 0.138 | 0.216 |
| **F5F5** | 0.548 | 0.046 | 0.015 | 0.061 |
| **F6F6** | 0.430 | 0.050 | 0.018 | 0.057 |
| **X1X1** | 0.323 | 0.093 | 0.008 | 0.039 |
| **X3X3** | 0.574 | 0.040 | 0.016 | 0.056 |
| **X8X8** | 0.371 | 0.044 | 0.047 | 0.132 |
| **X5X5** | 0.639 | 0.031 | 0.028 | 0.069 |

Descriptive results were generated to examine the nature of multiple occurrences of these behaviors, and whether multiple occurrences within the same interview might be treated as independent events. Three types of measures were generated: the frequency of occurrence of each in the interviews, the time to each occurrence examined for all occurrences and for first, second, etc., occurrences, and the timing of occurrences within each interview.

Table 3 presents the frequency of multiple occurrences for the two behaviors of interest across the 297 interviews. Almost one-quarter of the subjects did not interrupt question reading at any point in the interview, while nearly one-third interrupted question reading six or more times. Thus, the distribution of

[3] Approximate standard error.

interruptions is skewed to the right, reflecting the tendency in some interviews for the respondent to interrupt question reading frequently. Respondent laughter occurred in more than four-fifths of the interviews. The number of respondent laughter behaviors is, on average, higher than that for interruptions, with nearly two-fifths of the interviews having six or more respondent laughter events. It is difficult to determine from these marginal distributions whether the substantial proportions of interviews with a higher frequency of occurrences is due to respondent characteristics, interviewer characteristics or behavior, or a conversational style reflecting the contributions of respondent and interviewer characteristics.

**Table 3:** Frequency of interrupted question reading and respondent laughter, Health Field Study behavior coding sample (n=296).

| | *Interruptions* | | *Laughter* | |
|---|---|---|---|---|
| No. | Number of subjects | Percent | Number of subjects | Percent |
| 0 | 69 | 23.3 | 46 | 15.5 |
| 1 | 45 | 15.2 | 40 | 13.5 |
| 2 | 37 | 12.5 | 31 | 10.5 |
| 3 | 19 | 6.4 | 29 | 9.8 |
| 4 | 22 | 7.4 | 20 | 6.8 |
| 5 | 16 | 5.4 | 16 | 5.4 |
| 6+ | 89 | 30.1 | 115 | 38.9 |

The median time to the first, second, etc., occurrence, since the last occurrence (or the beginning of the interview) of each type of behavior is shown in Table 4. For those interviews with no interruptions, the median length of time to the completion of the interview is 53 minutes, somewhat shorter than the average of 60 minutes for all interviews. Of course, one might attribute this shorter length to the fact that question reading was not interrupted, and interviewers were not forced (through standardized interviewing practice) to reread the entire question again. On the other hand, interruptions might have been expected to speed the interview process along, and such a finding is something of a surprise. Not surprisingly, as the number of interruptions increases, the length of time to the interruption decreases. That is, more interruptions means there will, on average, be less time between interruptions. These findings are difficult to interpret, though, since the subject pool decreases for each successive number of occurrences. Only 104 of the 296 subjects for whom time to occurrence could be analyzed had six or more occurrences. This sample of 104 may differ substantially from the 296 – 104 = 192 subjects who had fewer than six occurrences of interrupting.

**Table 4:** Time to event for interrupted question reading and respondent laughter, Health Field Study behavior coding sample (n=296).

| | *Interruptions* | | *Laughter* | |
|---|---|---|---|---|
| Events | No. subjects | Median seconds to event | No. subjects | Median seconds to event |
| 0 | 69 | 3180 | 46 | 3000 |
| 1 | 227 | 423 | 250 | 697 |
| 2 | 182 | 276 | 210 | 339 |
| 3 | 145 | 206 | 179 | 284 |
| 4 | 126 | 291 | 150 | 229 |
| 5 | 104 | 195 | 130 | 190 |
| All | 227 | 158 | 250 | 184 |

Laughter exhibits a similar distribution of time to next occurrence. Interviews with no respondent laughter are shorter than those with one or more occurrences, and the time to each occurrence decreases as the number of occurrences increases. This inverse relationship is exactly expected for time to event data, and properly modeled using event history type models.

Selection bias from one level of occurrences to another in Table 4 may be partially addressed by examining time to occurrence separately for interviews with only one, with exactly two, etc., behaviors. Tables 5 and 6 present median time to occurrence for interviews with up to five occurrences of interruptions or laughter, respectively. For interviews with only one interruption (n=45), one might expect for random timing and uniform distribution of the interruption that it would occur midway through the average 3,600 second interviews, or at 1,800 seconds. The single interruption tends to occur, though, in the first half of the interview. For interviews with two interruptions, one might expect randomly generated uniformly spaced interruptions at 1,200 seconds successively. Again, time to interruption is somewhat shorter than expected. The number of interviews for exactly three, exactly four, and exactly five interruption interviews decreases to a point where comparative analysis is tentative, at best. Still, time to interruption tends to be shorter than expected. Not shown, of course, are times-to-interruption for interviews with more than five interruptions. If the times for interviews with fewer numbers of interruption are below expectations, those for interviews with more interruptions may actually show longer times to the first, and shorter times to the subsequent interruption. Heterogeneity of subject time to event prevents further elaboration of these ideas here, but they are returned to in the concluding section.

**Table 5:** Median time to interrupted question reading by number of times the subject interrupted during the interview, Health Field Study behavior coding sample (n=296).

*Median time to interruption, in seconds*

| Event | Only 1 (n=45) | Only 2 (n=37) | Only 3 (n=19) | Only 4 (n=22) | Only 5 (n=16) |
|---|---|---|---|---|---|
| First | 1439 | 900 | 311 | 490 | 262 |
| Second | | 1053 | 892 | 200 | 570 |
| Third | | | 620 | 299 | 58 |
| Fourth | | | | 784 | 192 |
| Fifth | | | | | 161 |

**Table 6:** Time to respondent laughter by number of times the respondent laughed during the interview, Health Field Study behavior coding sample (n=296).

*Median time to laughter, in seconds*

| Event | Only 1 (n=40) | Only 2 (n=31) | Only 3 (n=29) | Only 4 (n=20) | Only 5 (n=16) |
|---|---|---|---|---|---|
| First | 2099 | 1400 | 1130 | 974 | 623 |
| Second | | 948 | 407 | 579 | 381 |
| Third | | | 528 | 409 | 240 |
| Fourth | | | | 341 | 555 |
| Fifth | | | | | 209 |

For respondent laughter (see Table 6), the median time to occurrence is closer to expectation for subjects with only one or two interruptions. It may be that for these subjects the time to laughter is randomly generated uniformly across time, thus generating the kind of times shown in Table 6. For interviews with more occurrences, though, the pattern show for interruptions begins to repeat itself, with the time to the subsequent respondent laughter decreasing (for the most part) as the number of occurrences increases.

One phenomenon that could drive this process is a sequence of questions toward the end of the interview that fatigued respondents may interrupt, or laugh at. Tables 7 and 8 examine this possibility by giving a count of interruptions and respondent laughter across five-minute (300 second) intervals of the interviews. The number of subjects declines as the length increases because some subjects may end the interview before others. For both interruptions and laughter, the number of occurrences in five-minute intervals remains steady, despite declining numbers of interviews with larger numbers of events. For each time interval, the average number of occurrences per interview shows a clear increasing pattern for both interruptions and laughter. It appears that both behaviors occur more frequently toward the end of the interview. This may be a phenomenon due to respondent characteristics, how the interviewer delivers the questions, or the characteristics of the questions being asked.

**Table 7:** Number of interrupted question readings by time intervals during the interview, Health Field Study behavior coding sample (n=296).

| Interval | Number of interruptions | Subjects surviving | Mean no. per person |
|---|---|---|---|
| 0-300 | 165 | 296 | 0.56 |
| 301-600 | 94 | 296 | 0.32 |
| 601-900 | 109 | 295 | 0.37 |
| 901-1200 | 109 | 292 | 0.37 |
| 1201-1500 | 124 | 292 | 0.43 |
| 1501-1800 | 105 | 290 | 0.36 |
| 1801-2100 | 126 | 282 | 0.45 |
| 2101-2400 | 149 | 272 | 0.55 |
| 2401-2700 | 153 | 250 | 0.61 |
| 2701-3000 | 122 | 219 | 0.56 |

**Table 8:** Number of times respondent laughed by time interval in the interview, Health Field Study behavior coding sample (n=296).

| Interval | Number of interruptions | Subjects surviving | Mean no. per person |
|---|---|---|---|
| 0-300 | 85 | 296 | 0.29 |
| 301-600 | 146 | 296 | 0.49 |
| 601-900 | 138 | 295 | 0.47 |
| 901-1200 | 135 | 292 | 0.46 |
| 1201-1500 | 147 | 292 | 0.50 |
| 1501-1800 | 137 | 290 | 0.47 |
| 1801-2100 | 177 | 282 | 0.63 |
| 2101-2400 | 200 | 272 | 0.74 |
| 2401-2700 | 210 | 250 | 0.84 |
| 2701-3000 | 187 | 219 | 0.85 |

Further analysis was conducted to examine whether the characteristics of the questions could account for the apparent greater frequency of interruptions toward the end of the interview. The interruptions may be occurring because there are more questions toward the end of the interview that stimulate the respondent to interrupt. Alternatively, the respondent may have been exposed to a sequence of interruptible questions that have trained them to interrupt the question reading, resulting in more interruptions later in the interview.

Event history models with time varying covariates provide a method to examine the dynamics between two phenomenon across time. Question characteristics vary from one exchange to the next. Such models can be fit using continuous or discrete time techniques for estimation. For the sake of simplicity of coding the required data, a discrete time event history model was estimated for the time varying covariate of a question characteristic and the time to respondent interruption of the question reading.

The Health Field Study questionnaire contained approximately 450 questions, of which a typical respondent would have answered 125. Project staff reviewed all 450 questions to determine if the wording of a question might be expected to stimulate a respondent to interrupt the reading. Staff examined questions for characteristics such as the use of several successive clauses that qualified the condition or phenomenon of interest, or the presentation of a qualifying statement after the question had been asked. Consider as an example of the latter characteristic the question that asked respondents if they had visited a doctor's office in the previous six or 12-month period. Immediately following the question was a qualifying statement that excluded visits to a podiatrist. This question phrasing was deemed "interruptible", since respondents may have a tendency to answer the question before hearing the qualifying statement about podiatrists. The expert review identified a total of 23 questions considered to be interruptible. Respondents typically were exposed to 6-8 of these questions during an interview.

Each interview was divided into 30-second time intervals. The presence or absence of an interruption and an interruptible question was coded for each interval. The vast majority of the 30-second intervals captured two question exchanges, while exchanges with more behaviors would be spread across two or possibly three such time intervals. All occurrences of question interruptions were included in the analysis. However, to assess whether the relationship between question interruption and interruptible questions was the same at the beginning of the interview as the end, each occurrence was numbered, and a set of dummy indicator variables for first through $11^{th}$ occurrences and 12 or more occurrences was created.

A sequence of logistic regression models were then fit to the discrete time coded data. Six models are presented in Table 9. Model 1 is a simple event history model that examines the multivariate relationship among three demographic characteristics. Persons 65 years and older are the reference category for the five age group indicators. All five coefficients are positive and statistically significant, indicating that younger persons have greater hazard of interrupting the question reading than persons ages 65 years and older. Females have a larger hazard for interruption than males, while Black respondents are no different from non-Black respondent in the hazard of interruption.

Model 2 adds eleven dummy indicator variables for the occurrence of the interruption, dropping the first occurrence indicator that serves as the reference category in the model. Age differences in the hazard of interruption disappear, while the second, third, and eighth occurrences have higher hazard of interruption than the first. Thus, the hazard of interruption is not constant across multiple occurrences, and age, gender, and race groups have similar hazard rates for interruption. It appears that persons under 65 years have a higher hazard of interruption for the second, third, and eighth interruptions, and the presence of occurrence indicator variables accounts for this effect entirely.

The covariates in Models 1 and 2 are fixed throughout the time interval represented by each interruption. Model 3 introduces a time varying covariate, the presence or absence of an interruptible question. The strong positive coefficient indicates that expert opinion about interruptible questions has identified a set of questions that greatly increases the hazard of interruption. That is, interruptible questions as identified by expert review generate shorter times to the next interruption.

Model 4 adds the demographic variables into Model 3, and Model 5 adds demographic and occurrence indicator variables. Age, gender, and race coefficients in Model 4 are virtually identical to those in Model 1, indicating that demographics and interruptible questions operate independently with respect to the hazard of interruption. Introducing occurrence indicators in Model 5 again eliminates virtually all of the age and gender differences in Model 4, while showing a continued higher hazard of interruption for several occurrences after the first. The coefficient for age group 18-34 years is statistically significant, but it is only marginally larger than the same coefficient in Model 2. The interruptible question indicator remains positive and statistically significant. Thus, interruptible questions generate a higher hazard of interruption, regardless of demographics and occurrence (first, second, third, etc.) of the interruption.

Model 5 does not rule out the possibility that interruptible questions could increase the hazard of interruption later in the interview. If interruption were a “contagious” phenomenon in which as the respondent interrupts more and more often during the interview, then the hazard of interruption could decline as interruptible questions occur later in the interview. Model 6 adds eleven interaction variables to Model 5 to examine whether the interaction between occurrence and interruptible questions decreases the hazard of interruption. One half of the interaction coefficients are negative, but none of the negative coefficients are statistically significant. On the interaction coefficient for the second occurrence is statistically significant, and it is positive. On balance, the conclusion seems apparent: interruptible questions operate in the same way throughout the interview. There is no evidence of contagion of interruption through the course of the interview.

# 4 Concluding remarks

The purpose of the analysis presented in this paper has been exploratory: to examine the use of event history concepts and models in the investigation of dynamics in the survey interview. The analysis indicates that time is an important factor in considering two types of respondent behaviors, interruption and laughter. The time to interruption and laughter varies by several characteristics of the interview, such as the number and timing of the event. The multivariate event history models reveal that time varying covariates can be explored and usefully interpreted.

**Table 9:** Event history models (proportional hazards) of time to interrupted question readings for all occurrences, Health Field Study behavior coding sample (n=296).

| | Model1 | Model2 | Model3 | Model4 | Model5 | Model 6 |
|---|---|---|---|---|---|---|
| Age <18 | +0.667** | +0.168 | | +0.672** | +0.176 | +0.206 |
| Age 18-34 | +0.790** | +0.229 | | +0.800** | +0.241* | +0.277 |
| Age 35-44 | +0.587** | +0.082 | | +0.595** | +0.086 | +0.133 |
| Age 45-54 | +0.641** | +0.041 | | +0.654** | +0.050 | +0.151 |
| Age 55-64 | +0.703** | -0.042 | | +0.726** | -0.028 | +0.039 |
| Female | +0.120* | +0.043 | | +0.126* | +0.053 | +0.070 |
| Black | -0.023 | +0.032 | | -0.167 | +0.036 | +0.048 |
| **Int. Q** | | | **+0.838**** | **+0.850**** | **+0.964**** | **-0.006** |
| 2nd occ | | +0.895** | | | +0.905** | +0.805** |
| 3rd occ | | +0.351** | | | +0.367* | +0.416* |
| 4th occ | | -0.053 | | | -0.050 | -0.089 |
| 5th occ | | +0.313 | | | +0.308* | +0.330* |
| 6th occ | | +0.212 | | | +0.225 | +0.259 |
| 7th occ | | +0.080 | | | +0.076 | +0.124 |
| 8th occ | | +0.510** | | | +0.512** | +0.444* |
| 9th occ | | -0.342 | | | -0.335 | -0.371 |
| 10th occ | | +0.098 | | | +0.071** | +0.179** |
| 11th occ | | +0.402 | | | +0.410 | +0.340 |
| 12th occ | | +0.212 | | | +0.226 | +0.303 |
| Int. Q x 2nd | | | | | | +0.529* |
| Int. Q x 3rd | | | | | | -0.229 |
| Int. Q x 4th | | | | | | +0.236 |
| Int. Q x 5th | | | | | | -0.137 |
| Int. Q x 6th | | | | | | -0.207 |
| Int. Q x 7th | | | | | | -0.412 |
| Int. Q x 8th | | | | | | +0.514 |
| Int. Q x 9th | | | | | | +0.253 |
| Int. Q x 10th | | | | | | -0.732 |
| Int. Q x 11th | | | | | | +0.452 |
| Int. Q x 12th | | | | | | -0.636 |

*0.05 <= p < 0.01
** p <= 0.01

The examples offer limited interest to survey researchers exploring the relationship between interviewer and respondent behavior in the survey interview. The analysis examined a simple relationship between a respondent behavior and question characteristics that has been observed in survey interviews in the past. Standard behavior coding is often used to identify interruptible questions by finding questions with higher frequencies of interruptions. The event history models shown in the latter part of the results can be viewed as verification that expert review yields similar results to those of behavior coding in diagnosing problem questions in an interview. It is not a surprise, then, that there is a strong positive relationship between interruptible questions identified by expert review and the occurrence and timing of interruptions in the interview.

The present investigation has little immediate practical value for the survey practitioner who seeks to improve the quality of the survey interview. However, the methodology presented in the paper offers the potential to address a number of practical and important survey methodology issues. Through relatively straightforward linear models (on the log-hazard scale), investigators can study the dynamics among interviewer and respondent behavior, interviewer behavior and question characteristics, and respondent behavior and question characteristics (as was done in this investigation). A more thorough understanding of those dynamics may lead to models that explain the nature of interviewer or respondents tailoring their behavior during the course of the interview to adjust to stimuli presented by the behavior of their partner in the interview or the questions themselves. From such models, survey research methodologists can generate improved techniques for training interviewers and writing questions that reduces undesirable respondent behaviors.

For example, our recent work has found that respondent qualification of answers, expressing uncertainty of an answer, or providing a “don’t know” response during an exchange is related to less accurate answers. Event history models can be used to assess the relative strength of association of various interviewer behaviors, such as exact question reading, digression to establish rapport, or feedback, and characteristics of questions, such as retrospective reporting periods of varying length, the number of response categories, or presentation of instruction before or after the question asking, and respondent behaviors of interest. The models could be used to assess whether increasing or decreasing the level of interviewer behavior or the presence or absence of certain question characteristics would affect respondent behavior, before and after experimental manipulation of such behavior.

An obvious next step in these investigations will be to use event history models to examine the dynamic features of interviewer behavior, respondent behavior, and question characteristics. A set of approximately 30 interviewer and respondent behaviors are already coded and are being examined. Question characteristics will be coded following models available in the literature so that these can be linked to interviewer and respondent behavior. It may also be useful to introduce additional fixed covariates to these models that reflect respondent and interviewer characteristics which could confound the dynamic relationships to be explored using time varying covariates.

Several methodological features of the investigation can be readily strengthened. Continuous time models incorporating time varying covariates can and will be used for future models. Further exploration of the pooling of multiple occurrences of the same behavior will be examined to determine whether separate models ought to be constructed for subsets or single occurrences (such as the first). Proper variance estimation is an important issue for multiple occurrences since the occurrences from the same interview share fixed characteristics. Variances of estimated coefficients will tend to be underestimated using standard discrete time or continuous time event history model software. Jackknife repeated replication variance estimation will be examined to determine the effects of pooling multiple occurrences when using continuous time hazard models.

## References

[1] Bales, R.F. (1950): *Interaction: A Method for the Study of Small Groups*. Massachusetts: Addison-Wesley.

[2] Belli, Robert F. and Stephanie A. Chardoul (1997): The Digression of Survey Actors in a Face-to-Face Health Interview. Paper presented at the annual conference of the American Association for Public Opinion Research, Norfolk, VA.

[3] Belli, Robert F. and James M. Lepkowski (1996): Behavior of Survey Actors and the Accuracy of Response. *Health Survey Research Methods: Conference Proceedings* (69-74). DHHS Publication No. (PHS) 96-1013.

[4] Belli, Robert F., Paul S. Weiss, and James M. Lepkowski (1999): Dynamics of Survey Interviewing and the Quality of Report: Age Comparisons. In Norbert Schwarz, Denise Park, Bärbel Knäuper, and Seymour Sudman (Eds.), *Cognition, Aging, and Self-reports*, , 303-326. Philadelphia: Psychology Press.

[5] Brenner, M. (1982): Response Effects of Role-restricted Characteristics of the Interviewer. In W. Dijkstra and J. van der Zouwen (Eds.), *Response Behavior in the Survey Interview*, 131-165, London: Academic Press.

[6] Brenner, M. (1985): Survey Interviewing. In M. Brenner, J. Brown, and D. Canter (Eds.), *The Research Interview: Uses and Approaches*, 9-36, London: Academic Press.

[7] Cannell, C.F. and Kahn, R. (1968): Interviewing. In G. Lindzey, and E. Aronson (Eds.), *The Handbook of Social Psychology*, **2**, 526-595, Reading, Mass.: Addison-Wesley.

[8] Cannell, C.F. and Oksenberg, L. (1988): Observation of Behavior in Telephone Interviews. In R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls II, and J. Waksberg (Eds.), *Telephone Survey Methodology*, 475-495, New York: Wiley.

[9] Cannell, C.F., Blixt, S., and Oksenberg, L. (1992): *Characteristics of Questions that Cause Problems for Interviewers or Respondents*. Paper presented at the American Association for Public Opinion Research, St. Petersburg, FL.

[10] Fleiss, Joseph L. (1973): *Statistical Methods for Rates and Proportions*. New York: Wiley.

[11] Fowler, F.J. and Mangione, T.W. (1990): *Standardized Survey Interviewing*. Newbury Park, CA: Sage Publications.

[12] Fowler, Floyd J. and Charles F. Cannell. (1996): Using Behavioral Coding to Identify Cognitive Problems with Survey Questions. In Norbert Schwarz and Seymour Sudman (Eds.), *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Questions*, 15-36. San Francisco: Jossey-Bass.

[13] Groves, R. M. (1989): *Survey Errors and Survey Costs*. New York: Wiley.

[14] Landis, J.R. and G.G. Koch. (1977): The Measurement of Observer Agreement for Categorical Agreement. *Biometrics*. **33**, 159-174.

[15] Mangione, T.W., Fowler, F.J., and Louis, T.A. (1992): Question Characteristics and Interviewer Effects. *Journal of Official Statistics*, **8**, 293-307.

[16] Morton-Williams, J. (1979): The Use of 'Verbal Interaction Coding' for Evaluating a Questionnaire. *Quality and Quantity*, **13**, 59-75.

[17] Morton-Williams, J. and Sykes, W. (1986): The Use of Interaction Coding and Follow-up Interview to Investigate Comprehension of Survey Questions. *Journal of the Market Research Society*, **26**, 109-127.

[18] Oksenberg, L., Cannell, C., and Kalton, G. (1991): New Strategies for Pretesting Survey Questions. *Journal of Official Statistics*, **7**, 349-365.

[19] Schaeffer, N.C. (1991): Conversation with a Purpose - or Conversation? Interaction in the Standardized Interview. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (Eds.), *Measurement Errors in Surveys*, 367-391, New York: Wiley.

[20] Suchman, L. and Jordan, B. (1990): Interactional Troubles in Face-to-Face Survey Interviews, *Journal of the American Statistical Association*, **85**, 232-241.

[21] Sykes, W. and Collins, M. (1992): Anatomy of the Survey Interview. *Journal of Official Statistics*. **8**, 277-291.