PROBLEMS WITH QUESTION WORDING BY AGE OF RESPONDENT

Sandra L. Edwards, University of Utah
University of Utah, Salt Lake City, Utah 84132

KEY WORDS: Survey questions, respondent characteristics, interviewer behavior

Introduction

Behavior coding of taped interviews often identifies problems that interviewers or respondents have with question wording (e.g., Cannell and Oksenberg, 1988; Fowler and Mangione, 1990; Fowler, 1992). Problems with question wording appear in both the way questions are asked and in the interviewer and respondent behavior that follows a question. For example, a lengthy question might lend itself to interruption by the respondent before the interviewer can read it, or a question might be worded in a way that makes an interviewer feel foolish when asking it. A question might use a word that many respondents do not understand thus leading to frequent requests for clarification, or a question might ask for information from a respondent that is very hard to remember.

This paper uses data from interviewer behavior coding done as part of a health study to look at interviewer behavior with regard to the way questions are asked and the amount of question repeating, clarifying, and probing behaviors after each question by age of respondent in order to identify any differences in problems with question wording by age of respondent.

Methods

The data for this paper come from a multicenter case-control study being conducted at the University of Utah in conjunction with Kaiser Permanente Medical Care Program of Northern California and the University of Minnesota. Computer-Assisted-Personal Interviews (CAPI) are being conducted with cases and controls 30-79 years of age at each center. The interview contains two parts: a quantitative food frequency diet history questionnaire and questions regarding demographics, medical history, occupation, physical activity, family history, and other lifestyle behaviors. The diet history questionnaire is administered in the middle of the interview. The interview takes about two and a half hours to administer. To date, 1,675 interviews have been completed.

Each interview is audio tape-recorded after receipt of respondent permission. One hundred twenty randomly selected taped interviews have been coded to date using the system developed by Cannell (Cannell, Lawson, and Hausser, 1975; Cannell and Oksenberg, 1988). The diet history questionnaire is structured in such a way that it cannot be coded using this system. The nondiet questionnaire contains 1,004 potential items to be coded. All behavior coding is done at the University of Utah by one coder using direct data entry onto a computer. The primary purpose of the coding is to check for consistency in conducting the interview between interviewers at each center and to provide ongoing feedback to interviewers as to their interviewing skills.

Interviewer behavior coding of the interviews also allowed for methodological analysis. Each question was scored in the categories of question reading, probing, and other behavior according to percent of acceptable and unacceptable behaviors. Problems with question wording were defined in two ways: "Problem Questions-Asking" and "Problem Questions-Probing."

Problem Questions-Asking were identified by calculating the "Q-Error" (Presser, 1992), the number of times the question was misread divided by the total number of times it was asked in the 120 coded interviews for each question. Questions with a Q-Error of higher than .1, that is, questions read in an unacceptable way more than 10% of the time, were defined as "Problem Questions-Asking" (PQ-A). To aid in analysis, these questions were also categorized according to: "What"--questions that asked a respondent to specify "what" following a yes/no screening question, "Length"--the number of words in the question, "Series"--whether the wording is largely identical to a question or series of questions immediately preceding it, and "Other"--questions that did not fit into any of the above categories. Questions asked less than 10 times in the 120 interviews and all questions in the family history section were excluded from analysis.

In the second analysis, each question was scored according to the number of interviewer behaviors following it excluding code 58's, which are predominately interviewer behaviors such as ok, m—m—m, uh—h—h. Any questions for which the average number of behaviors after them was greater than one were considered to be "Problem Questions-Probing" (PQ-P). Questions asked less than 10 times in the 120 interviews and all questions in the family history section were excluded from this analysis.

Of the 120 interviews, 56 were completed with respondents 30-64 years of age and 64 with respondents

Table 1. Percentage and Range of Unacceptable Question Asking for PQ-A

	Age 30-64	Age 65-79
Number of PQ-A	26	42
Range of % unacceptable question asking for PQ-A	10.3%-56.3%	10.5%-63.6%
Mean % unacceptable question asking PQ-A	19.7%	21.2%
Total Number of Times Asked	922	1306
Number of Interviews	56	64

65-79 years of age. The categorization into these two age groups was arbitrary as no set definition of "older" exists (Botwinick, 1984).

Results
Question Asking

Table 1 contains the number, range of unacceptable question asking behavior, and mean percentage of unacceptable question asking behavior for questions identified as PQ-A by age of respondent. Twenty-six questions were identified as PQ-A for respondents 30-64 years of age. Forty-two were identified for respondents 65-79 years of age.

Figure 1 summarizes the distribution of the types of question asking behaviors for the PQ-A by age of respondent. While the overall percentage of acceptable question asking behavior for the PQ-A was the same (80.2% versus 80.2%) for the two age groups, interviewers had a tendency to read these questions as written to respondents 65-79 years of age and with minor wording changes (often this was skipping optional phrases) with respondents 30-64 years of age. The percentage of questions read with major changes was the similar for each age group. Interviewers tended to not read questions to respondents 30-64 years of age versus making a statement about an answer the interviewer expected to respondents 65-79 years of age.

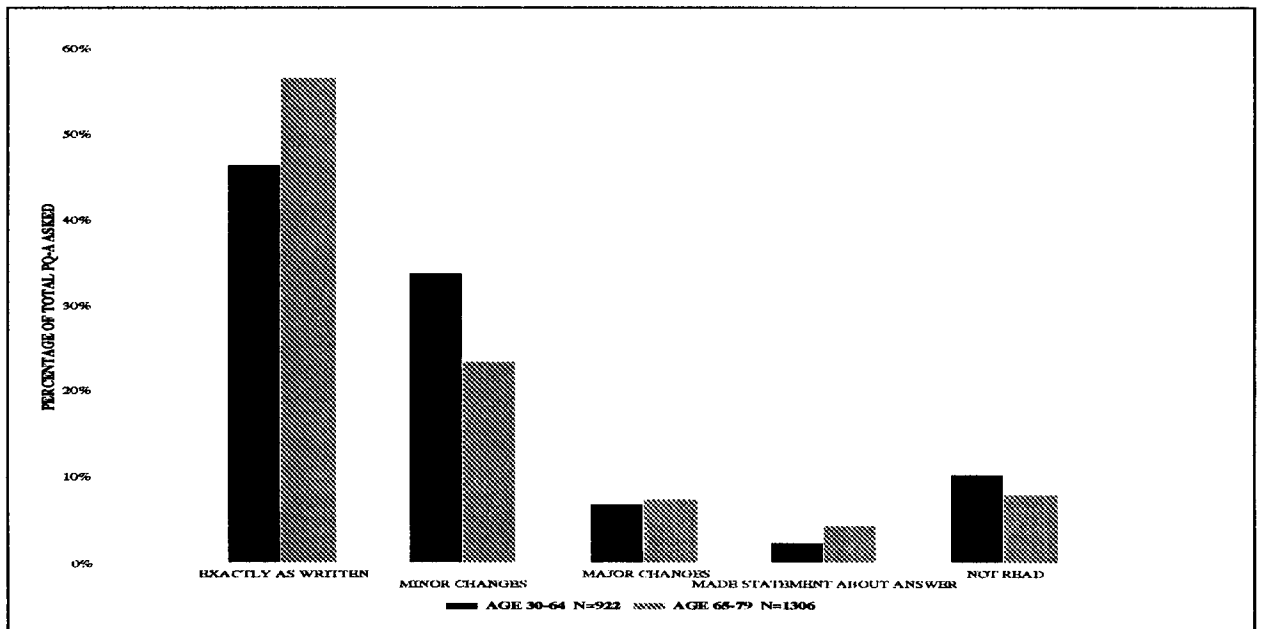**Figure 1. PQ-A Question Asking Behavior**
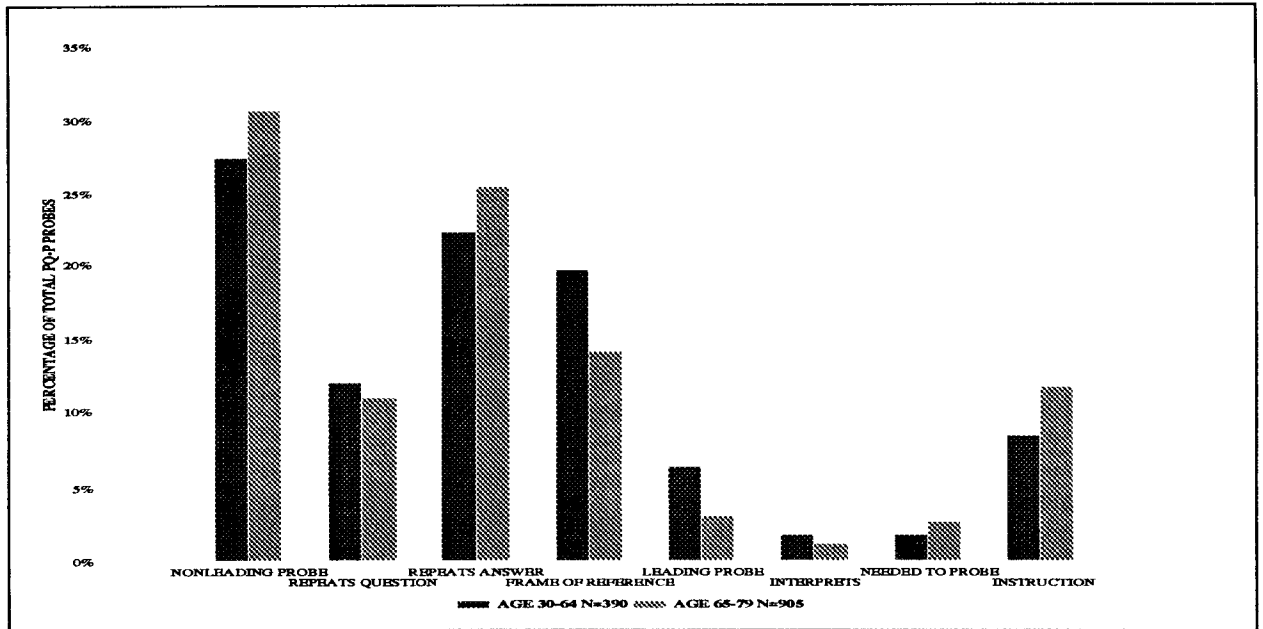
Figure 2. PQ-P Probing Behavior

Table 2. PQ-A Question Type Frequencies
and Percentages

Question Type	Age 30-64		Age 65-79	
	#	%	#	%
What	10	38.5%	12	28.6%
Series	9	34.6%	22	52.4%
Length	4	15.4%	4	9.5%
Other	3	11.5%	4	9.5%
Total Number of PQ-A	26	100%	42	100%

The frequency of types of questions for the PQ-A are in Table 2. While the frequency of "What" and "Series" PQ-A were almost evenly split in the 30-64 year old respondents, "Series" questions accounted for half of the PQ-A in the 65-79 year old respondents. Every "What" PQ-A for 30-64 year old respondents appeared as a PQ-A for 65-79 year old respondents. Each "Series" PQ-A for 30-64 year of respondents appeared as a PQ-A for 65-79 year old respondents except two. Only one of the PQ-A categorized as "Length" was the same for the two groups. In the "Other" category, the same three questions appeared in the 30-64 year old respondents as the 65-79 year old respondents. An additional one question appeared as a PQ-A for 65-79 year old respondents.

Probing

Table 3 contains the number, range, and mean of probing behavior following the questions identified as PQ-P by age of respondent. Fewer questions were identified as PQ-P than PQ-A. Twice as many questions were identified as PQ-P for the 65-79 group than the 30-64 group. As with the PQ-A, the range was wider in the 65-79 group and the average mean number of behaviors higher. Six questions appeared on both

Table 3. Range and Mean of Probing
Behaviors Following PQ-P

	Age 30-64	Age 65-79
Number of PQ-P	8	17
Range of probes	1.118-1.554	1.032-1.826
Mean	1.2	1.34

lists. Figure 2 shows the frequency of the types of probes for all PQ-P asked by age.

Discussion

These data suggest that the number of questions with a Q-Error higher than .1 is greater for respondents 65-79 years of age. In addition, there were more questions with an excess of interviewer behavior following them in the 65-79 year old respondents.

The reasons for these differences cannot be established from this research. Older adults have been shown to have different ways of encoding and comprehending information (Botwinick, 1984), but question asking should not be affected by differences in encoding or comprehension of information. All interviewers were instructed to read each question as written and to finish reading the question even if the respondent interrupted with an answer. Making a statement about the answer the interviewer excepted based on something the respondent said or not reading the question and recording the answer was coded as unacceptable and interviewers were given immediate feedback through the quality control process regarding this behavior. Despite this training and supervision, differences in the way questions were asked by age of respondent did exist. This data would seem to support a hypothesis of Presser (1992) that respondent characteristics contribute to interviewer's departing from question wording. This was particularly true with the "Series" questions. While Fowler and Mangione (1990) did not find the misreading of questions to be related to the size of interviewer effect, Fowler (1991, p.266) states "some wording changes affect answers, others do not, and we lack good generalizations about which changes matter most." In addition, training interviewers to read questions as written and then giving them questions that are difficult to read or often skipped because the interviewer to feels foolish when reading serves to give a mixed message to interviewers and respondents as to the importance of standardization.

Differences in probing behavior are even more difficult to access. Due to time and money constraints the respondents' responses could not be coded. Without these codes, it is impossible to tell if the excess probing is due to difficulties a respondent was having in answering the question due to encoding or comprehension differences as they relate to age, a result of respondent characteristics influencing interviewer behavior, or some interaction of the two. What is clear, however, is the effect problems with question wording has on the data. As part of another analysis of study data, the ρ_{int} was calculated for selected questions. No calculations were done by age, but the ρ_{int} was significant at the .05 level for each of the PQ-P for which it was calculated.

As the population of the United States grows older, the number of older respondent's in population-based studies will increase. As this number grows, it is important for those of us who design questionnaires and train and supervise interviewers to be aware of the potential difference in problems questions by age of respondent. This applies not only to the application of cognitive research and questionnaire pretesting, but to

pursuing the notion that respondent characteristics, in this case age, contribute to interviewer departure from question wording and excess probing. While the scope of this paper was not large enough to identify specific solutions, I hope it demonstrates the need to due further research on the interview process based on respondent age.

References

Botwinick, Jack. 1984. *Aging and Behavior*. New York: Springer

Cannell, Charles F., Sally A. Lawson, and Doris Hausser. 1975. *A Technique for Evaluating Interviewer Performance*. Ann Arbor: Survey Research Center, University of Michigan.

Cannell, Charles F. and Lois Oksenberg. 1988. "Observation of Behaviour in Telephone Interviews." In *Telephone Survey Methodology*, ed. R. Groves et al., pp. 475-95. New York: Wiley.

Fowler, Floyd J. 1991. "Reducing Interviewer Error Through Interviewer Training, Supervision, and Other Means." *In Measurement Errors in Surveys*, ed. P. Biemer et al., pp.259-78. New York: Wiley

Fowler, Floyd J. 1992. "How Unclear Terms Affect Survey Data." *Public Opinion Quarterly* 56: 218-31.

Fowler, Floyd J. and Thomas W. Mangione. 1990 *Standardized Survey Interviewing*. Newbury Park: Sage.

Kelsey, Jennifer L., Linda A. O'Brien, Jeane Ann Grisso, and Susie Hoffman. 1989. "Issues in Carrying Out Epidemiologic Research in the Elderly." *American Journal of Epidemiology* 130: 857-66.

Presser, Stanley and Zhao, Shanyang. 1992. "Attributes of Questions and Interviewers As Correlates of Interviewer Performance." *Public Opinion Quarterly* 56: 236-240.