The Quality of Data Collection by an Interview on the Prevalence of Elder Mistreatment

Hannie C. Comijs, PhD
Wil Dijkstra, PhD
Lex M. Bouter, PhD
Johannes H. Smit, PhD

ABSTRACT. The aim of the present study was to investigate the quality of the interview process that was used to obtain data on the prevalence of elder mistreatment in order to improve the quality of this process in future research studies. Audiotapes of interviews with 143 victims of elder mistreatment were transcribed into verbatim protocols that were subsequently coded and analyzed. The results of the present study show that despite the fact that all interviewers were carefully selected, intensively trained, and supervised during the interview period, 4.2% of the questions were skipped, and in 4.4% of the interactions no usable answer was obtained. However, in all of these cases the interviewer did register an answer, the validity of which is questionable. Considerable differences were found between interviewers with respect to their performance and also between the various questions. Finally, suggestions are made to improve interviewer behavior.
KEYWORDS. Elder mistreatment, prevalence, question-answer sequence, interaction coding

INTRODUCTION

In the past decennium, several studies on the prevalence of elder mistreatment have been carried out (Pillemer & Finkelhor, 1988; Podnieks et al., 1990; Kurrle et al., 1992; Ogg & Bennett, 1992; Comijs et al., 1998). Estimates of the prevalence of elder mistreatment are usually based on data from standardized interviews in cross-sectional samples of the population (Pillemer & Finkelhor, 1988; Podnieks et al., 1990; Ogg & Bennett, 1992; Comijs et al., 1998). Because elder mistreatment is a relatively rare phenomenon (prevalences found in various studies on elder mistreatment vary between 3% and 6%), a few cases of inaccurate classification may seriously influence the estimated prevalence figures. Especially in research on elder mistreatment inaccurate classification might be caused by certain specific conditions that are related to the sample and the topic under study.

To obtain adequate answers from respondents a number of prerequisites should be met: interviewers should ask questions correctly, should probe if answers are inadequate, should build rapport with the respondent, etc. Therefore careful selection and training of the interviewers is of major importance. However, despite conscientious selection and training, the interviewers sometimes encounter some unforeseen problems that might influence the quality of the collected data. Firstly, asking questions about elder mistreatment, which is an emotionally charged topic for the respondents as well as the interviewers, may result in inadequately obtained scores. Among the respondents, recalling painful moments in their lives can cause confusion or emotional reactions. This, in turn, might lead to the interviewers skipping questions because they might find it difficult to ask certain respondents particular ones or interpreting inadequate, improper or vague answers to avoid explicit questioning (Sudman & Bradburn, 1982). Secondly, the advanced age of the study population (over 65 years of age) may cause difficulties in providing precise answers because of cognitive problems (Andrews & Herzog, 1986; Smit et al., 1997) or the length of the period between the moment of occurrence and the moment when the respondent is asked to recall (Groves, 1989; Czaja et al., 1994).
Finally, an interview on a topic such as elder mistreatment is difficult for both the respondent and the interviewer. The aim of the present study is to investigate the quality of the process used to obtain interview data on the prevalence of elder mistreatment in order to improve the quality of this process in future research. The study focused on inadequately obtained scores because the question was skipped or no usable answer was given. The main research questions addressed were:

1. What percentage of scores was inadequately obtained in interviews on elder mistreatment?
2. Are inadequately obtained scores related to
   - the interviewer,
   - characteristics of the question,
   - length of the reference period, or
   - cognitive functioning of the respondents?

**METHOD**

**Sample**

The participating subjects were recruited from a sub-sample of respondents in the Amsterdam Study of the Elderly (AMSTEL). The AMSTEL project is a community-based follow-up study on cognitive functioning and decline in non-institutionalized elderly people (65 years of age) living in Amsterdam, the Netherlands. Respondents were selected from 30 general practices throughout the city. The sample (n = 4051) was stratified into four 5-year strata (65-69, 70-74, 75-79, and 80-84 years) of equal size. It is described in more detail by Launer and colleagues (1993). In 1994, 1,797 non-institutionalized and non-demented respondents from the original AMSTEL baseline sample participated in an interview on elder mistreatment.

The one-year prevalence of elder mistreatment was found to be 5.6%. The one-year-prevalence of the various categories of elder mistreatment was: chronic verbal aggression 3.2%, physical aggression 1.2%, financial mistreatment 1.4% and neglect 0.2% (Comijs et al., 1998). To investigate the influence of the length of the reference period on the quality of the data, respondents were included if they
reported at least one incident of physical aggression or financial mistreatment since they became 65 years of age, or if they reported at least 10 incidents of neglect or chronic verbal aggression in the year prior to the interview (Pillemer & Finkelhor, 1988; Podnieks et al., 1990). This resulted in a sample of 190 respondents who participated in the present study.

**Interviewer Training**

The interviews were carried out by 31 carefully selected and intensively trained interviewers; all had over 15 years of education, and most of them were female (n = 28). The training of the interviewers took place during five sessions, each lasting six hours. Video-examples illustrated basic interview rules, and role-playing was used to practice interviewer skills. One afternoon was entirely dedicated to questions about elder mistreatment. Each interviewer conducted a test interview that was discussed afterwards.

**Fieldwork**

All interviews were conducted face to-face in the homes of the respondent. The interviews were audiotaped to monitor interviewer behavior. During the interviewing period the interviewers were closely supervised; the audiotapes were randomly checked a few days after the interviews took place to identify mistakes the interviewers had made.

**Questionnaire**

The questionnaire consisted of 23 initial questions (see Appendix), for instance, “Since you became 65, have you ever been shouted at by someone you know well?” If the answer was “yes,” the two following questions asked how often it had happened since the age of 65 and how often it had happened in the year prior to the interview. Only the 23 initial questions were included in the present analyses because they provide the essential information needed to select further questions.

**Coding Procedure**

The audiotapes of interviews with 143 of the victims of elder mistreatment were transcribed into verbatim protocols. The interviews
with the other 47 respondents could not be transcribed for the following reasons: bad recording quality, incompleteness (the interviewer had forgotten to turn the tape on or the tape had run out) or the tape was missing. Each interview was divided into so-called question-answer sequences. A question-answer sequence consists of all verbal acts uttered by the interviewer and the respondent concerning one of the questions on the questionnaire. Because the interview consisted of 23 (initial) questions, this amounted to $23 \times 143 = 3289$ question-answer sequences. The total number of verbal acts was 12,924. Each verbal act was coded according to four coding variables: the speaker, the way in which the information was exchanged, the evaluation of the adequacy of the verbal acts, and the answer to the scripted question (Dijkstra, 1999). Thus, a coded verbal act consisted of four successive codes. For example, if the interviewer asked a question exactly as it was formulated in the questionnaire, this would be coded as IQAX (Interviewer, Question, Adequate and X for not applicable). Table 1 gives an example of a fully coded question-answer sequence.

In general, to gather valid data, the questions must be asked as they are scripted, and an adequate answer to the question must be given. In the present study, the focus was on inadequately obtained scores, namely: (1) a scripted question is skipped (skipped question) or (2) the answer is not based on an explicit answer, but in some way inferred by the interviewer (unusable answer).

**Other Variables**

The length of the reference period was measured as the period between the moment when the respondent became 65 years of age and the moment of the interview. Cognitive functioning of the respondents was measured by means of the Mini Mental State Examination (MMSE) (Folstein et al., 1975), a scale consisting of 21 items to measure cognitive performance. The score of the entire scale ranges from 0-30, with higher scores indicating better cognitive functioning.

**Analyses**

All analyses were performed by means of the SEQUENCE program (Dijkstra, 1994). To assess the reliability of the coded verbal acts, ten randomly selected interviews ($10 \times 23 = 230$ sequences, consisting of 979 verbal acts) were each coded by two independent coders. An act
TABLE 1. An Example of a Fully Coded Question-Answer Sequence

<table>
<thead>
<tr>
<th>Codes</th>
<th>Speech acts</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQAX</td>
<td>I: Since you became 65, have you ever been shouted at by someone you know well?</td>
<td>The interviewer (I), asks a question (C) from the questionnaire, adequately (A) formulated as written, answer code is not applicable (X).</td>
</tr>
<tr>
<td>RRAX</td>
<td>R: Do you mean I shouted at somebody?</td>
<td>The respondent (R) asks for an explanation (R), adequately to the question (A), answer code is not applicable (X).</td>
</tr>
<tr>
<td>IRCX</td>
<td>I: No, has it ever happened that someone shouted at you?</td>
<td>The interviewer (I), repeats the question (Q) from the questionnaire, with a minor change in formulation (C), answer code is not applicable (X).</td>
</tr>
<tr>
<td>RAAN</td>
<td>R: No, we don’t shout at each other.</td>
<td>The respondent (R) gives an answer to the question (A), adequately (A), answer is ‘no’ (N).</td>
</tr>
</tbody>
</table>

was considered to be coded differently if at least one of the four codes was differently coded by both coders. Cohen’s kappa was used as a measurement of reliability. Adequately obtained scores were related to the interviewer, the question, the length of the reference period, and the degree of cognitive functioning of the respondent, either by means of chi-square analyses, using likelihood ratios, or by means of t-test analyses.

RESULTS

The inter-coder reliability in a sample of verbal acts was found to be good ($\kappa = 0.65$). Some descriptive information is presented in Table 2. The mean age of the respondents was 77.6 years, ranging from 69 to 88 years. The mean reference period was 12.8 years (SD 5.3). The mean score on the MMSE was 27.0 (SD 2.6), ranging from 18 to 30. Of the 3,289 anticipated question-answer sequences, 4.2% had been skipped, and in 4.4% of the question-answer sequences no adequate answers had been given.

In Table 3 the skipped questions and unusable answers are related to the question, the interviewer, the length of the reference period, and the cognitive functioning of the respondent. The percentage of skipped questions varied strongly per question (likelihood ratio: 119.0, df 22, p < .001). No statistically significant association was
TABLE 2. Descriptive Information

<table>
<thead>
<tr>
<th>Variables</th>
<th>Descriptives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of question-answer sequences¹</td>
<td></td>
</tr>
<tr>
<td>- neglect</td>
<td>858</td>
</tr>
<tr>
<td>- chronic verbal aggression</td>
<td>429</td>
</tr>
<tr>
<td>- physical aggression</td>
<td>1430</td>
</tr>
<tr>
<td>- financial mistreatment</td>
<td>572</td>
</tr>
<tr>
<td>Total number of sequences</td>
<td>3289</td>
</tr>
<tr>
<td>Number of skipped questions (%)</td>
<td>137 (4.2)</td>
</tr>
<tr>
<td>Number of unusable answers (%)</td>
<td>145 (4.4)</td>
</tr>
<tr>
<td>Reference period, in years, mean (SD)</td>
<td>12.8 (5.3)</td>
</tr>
<tr>
<td>Cognitive functioning, mean score (SD)</td>
<td>27.0 (2.6)</td>
</tr>
</tbody>
</table>

¹based on 143 interviews

TABLE 3. The Association Between the Two Types of Inadequately Obtained Scores and the Question, the Interviewer, the Length of the Reference Period and the Cognitive Functioning of the Respondent (N = 3289)

<table>
<thead>
<tr>
<th></th>
<th>Questions skipped</th>
<th>Unusable answers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 137 (4.2%)</td>
<td>n = 145 (4.4%)</td>
</tr>
<tr>
<td>Question-number¹</td>
<td>.18 **</td>
<td>.10 ns</td>
</tr>
<tr>
<td>Interviewer¹</td>
<td>.34 **</td>
<td>.24 **</td>
</tr>
<tr>
<td>Reference period²</td>
<td>.52 ns</td>
<td>1.34 ns</td>
</tr>
<tr>
<td>Cognitive functioning (MMSE)²</td>
<td>.53 ns</td>
<td>.25 ns</td>
</tr>
</tbody>
</table>

¹ association measured by Cramér by means of chi-square statistics
² association measured by t-test for independent samples
** p < 0.001
# not significant

found between the number of the question and the percentage of unusable answers.

The percentage of skipped questions was highest for the questions on neglect (5.7%), followed by physical aggression (5.0%), and financial mistreatment (2.8%). Hardly any of the questions on chronic verbal aggression were skipped (0.2%). In the series of questions about neglect, physical aggression, and financial mistreatment, more
questions were skipped towards the end of the series. Among the questions on physical aggression, the last questions concerned severe violence, such as being shot at and being attacked with a knife (14.0% questions skipped). The questions on neglect do not increase in severity; the most frequently skipped questions were those concerning lack of help with incontinence problems and getting in or out of a chair or bed. The question, which was most often skipped in the series on financial mistreatment, was that concerning mismanagement of the financial affairs of the respondent. The highest percentages of unusable answers were found in connection with physical aggression (5.4%), followed by financial mistreatment (4.4%), and chronic verbal aggression (3.7%). The percentage of unusable answers was lowest in connection with neglect (3.1%).

The overall percentage of inadequately obtained scores, including all skipped questions and all unusable answers, is strongly associated with the question (likelihood ratio: 73.3, df 22, p < .001) as well as with the interviewer (likelihood ratio: 284.1, df 30, p < .001). The percentage of inadequately obtained scores was highest for physical aggression (10.3%) and lowest for chronic verbal aggression (4.0%).

Whether certain interviewers registered more incidents of elder mistreatment than other interviewers was investigated. The results showed no association between the reporting of elder mistreatment and the interviewers. However, as is shown in Table 3, significant differences between interviewers were found in the percentage of skipped questions (likelihood ratio: 277.86, df 30, p < .001) and the percentage of unusable answers (likelihood ratio: 115.07, df 30, p < .001). Six out of 31 interviewers skipped over 10% of the questions they should have asked, with a maximum of 30.4% skipped questions. Two of the 31 interviewers had over 10% unusable answers; one of these two interviewers also skipped more than 10% of the questions.

Whether the occurrence of inadequately obtained scores was associated with the length of the reference period and the cognitive functioning of the respondent was also analyzed. As Table 3 shows, no such associations were found. The percentage of skipped questions and the percentage of unusable answers were neither significantly related to the length of the reference period nor to the cognitive functioning of the respondent.

A total of 8.6% of the 3,289 scores were inadequately obtained. This implies that for 3,007 of the 3,289 (91.4%) question-answer
sequences the score registered by the interviewer corresponded with the answer given by the respondent. In 282 question-answer sequences (8.6%) the score registered by the interviewer was, in fact, not based on an explicit answer, but in some way inferred by the interviewer.

**DISCUSSION**

The results of the present study show that 8.6% of the scores in interviews on elder mistreatment were inadequately obtained. Despite the fact that all interviewers were carefully selected, intensively trained, and supervised during the interview period, 4.2% of the questions were skipped, and in 4.4% of the interactions no adequate answer was given. However, in all of these cases the interviewer did register an answer, but the validity is questionable. Considerable differences were found in the individual performances of the 31 interviewers. Whether bad performance was related to the gender or level of education of the interviewer was not investigated because of low variability. All interviewers were highly educated, and most of them were female. Moreover, in previous studies, background variables of interviewers had no effect (Hox et al., 1991, van der Zouwen et al., 1995). The differences in performance between interviewers may be related to their ability to handle unforeseen behavior of a respondent, since respondents differ greatly in the way they react to an interview. Interviewers did not differ with respect to the mean of the registered scores on elder mistreatment.

A strong relationship was found to exist between characteristics of the questions and the number of questions skipped. Within the series of questions in a particular category of mistreatment, the pattern of the skipped questions is striking. Especially towards the end of the series of questions on neglect, physical aggression, and financial mistreatment, there was an increase in the number of skipped questions. This might be associated with repetition of the same type of questions. For instance, in the series of questions on neglect, questions are asked about help with performing certain activities of daily living, such as eating. After two or three questions of this type, respondents often report that they have never needed any help.

Some interviewers find it hard to continue asking these questions once respondents have suggested that they are completely indepen-
dent. The same applies to physical aggression; if respondents report that they have never experienced any violence, interviewers may stop asking the remaining questions about physical aggression. The omission of questions may also depend on their content. Regarding neglect, the most frequently skipped question concerned incontinence problems. This question often needed some explanation because respondents did not understand the word. Despite the training, this may have embarrassed interviewers, who therefore skipped this particular question. With respect to physical aggression, the most frequently skipped questions concerned the most severe acts of violence, which interviewers may have found difficult to ask. In the case of financial mistreatment, the most frequently skipped question was the mismanagement of the respondents’ financial affairs. Some interviewers first asked the respondents if they took care of their own financial affairs. If the answer was affirmative, they skipped the question because they probably thought it was no longer relevant. It is important to note here that skipping questions may lead to serious errors, for example, financial mistreatment. At first sight it does not seem wrong to skip questions about financial mistreatment if respondents said that they take care of their financial affairs themselves. However, if the following questions were asked and not skipped, it might be that respondents are taking care of their own financial affairs because they experienced financial mistreatment by someone.

In this study, no statistically significant associations were found between the percentage of inadequately obtained scores and the length of the reference period or the cognitive functioning of the respondent. However, this does not mean that the reference period or the cognitive functioning does not influence the quality of the data at all, because it was not possible to check the validity of the answers.

The results show that nearly 9% of all scores were inadequately obtained, and therefore might have biased the findings regarding the prevalence of elder mistreatment. The skipping of questions has probably resulted in under-reporting of incidents of mistreatment in the sample under study. The magnitude of this under-reporting is difficult to estimate because not all questions that were skipped would have provided an affirmative answer. The percentage of unusable answers may have resulted in both under-estimation and over-estimation of the prevalence. However, interviewers may also have adequately inferred
some of the scores from previous answers or from the preceding conversation.

Certain limitations should be considered in the design of this study. Firstly, only the audiotapes of victims of elder mistreatment were used. Audiotapes of respondents not reporting elder mistreatment were not included. In these latter interviews, questions may have been skipped as well, which may also have resulted in under-reporting of elder mistreatment. Secondly, the number of inadequately obtained scores may be under-estimated; scores were only considered to be inadequately obtained if they produced an unusable answer. However, it is possible that other mistakes were made, for instance, when the formulation of a question was changed in such a way that the content of the question was changed, or when the interviewer suggested an answer to obtain an adequate answer. A global inspection of these possible mistakes in the study sample made it clear that they are difficult to objectify because they were mainly made at a later stage in the question-answer sequence. These interactions often started with adequate questions, but because the respondent did not answer the question correctly, the interviewers often reformulated the question inadequately or suggested an answer.

The results of the present study make it clear that it is necessary to reconsider the content of the training and supervision of interviewers. Firstly, next to the basic training on interviewing techniques and the topic of the interview, attention should be given to questioning elderly people about an emotionally charged subject and to the prevention of the type of mistakes, which occurred in the present study.

Interviewers should never skip any questions. Admittedly, posing questions when the answer already seems clear may disturb the relationship between the interviewer and the respondent. It might suggest that the interviewer is not really listening to the respondent. Hence, the interviewer must explain that he or she does listen (e.g., by repeating answers given by the respondent or by referring to earlier answers: “We already talked about . . .”). Furthermore, the interviewer has to explain that every question has to be posed. There are various subtle ways to accomplish this. Interviewers should practice such techniques during the training.

Interviewers should learn how to obtain adequate answers. Firstly, they should learn to distinguish adequate answers from inadequate ones. In this study adequate answers only consist of “yes” or “no” or
any other answer that can be interpreted unequivocally as such. In all other cases the interviewer has to probe in a non-suggestive way to obtain an adequate response. Secondly, retraining periods should be included during the interviewing period in which the experiences on questioning elderly people about an emotional subject should be a major issue. Furthermore, individual supervision as a consequence of reviewing the audiotapes should be included in the retraining program.

REFERENCES


**APPENDIX**

**Questionnaire on Elder Mistreatment**

*Finally, I am going to ask you some questions about negative experiences in your private life. They are about aggressive behavior towards you, by your partner, children, or other people you know well. The first questions are about lack of assistance when you needed it.*

(NEGLECT$^1$)

1. Since you became 65, has it ever happened that a person you know well did not help you to wash yourself or have a bath, although he or she knew you needed help?

2. Since you became 65, has it ever happened that a person you know well did not help you to dress, although he or she knew you needed help?

3. Since you became 65, has it ever happened that a person you know well did not help you to go to the toilet, although he or she knew you needed help?
4. Since you became 65, has it ever happened that a person you know well did not help you getting in or out of your chair or bed, although he or she knew you needed help?

5. Since you became 65, has it ever happened that a person you know well did not help you with incontinence problems, although he or she knew you needed help?

6. Since you became 65, has it ever happened that a person you know well did not help you with your food, although he or she knew you needed help?

The next questions are about rude and violent behavior that you might have experienced.

(CHRONIC VERBAL AGGRESSION\textsuperscript{2})

7. Since you became 65, have you ever been shouted at by someone you know well?

8. Since you became 65, have you ever been insulted or sworn at by someone you know well?

9. Since you became 65, have you ever been threatened by someone you know well? (For instance, threatened to hit you or throw something at you.)

(PHYSICAL AGGRESSION\textsuperscript{3})

10. Since you became 65, has anyone you know well ever thrown something at you?

11. Since you became 65, has anyone you know well ever pushed or grabbed you?

12. Since you became 65, has anyone you know well ever hit or slapped you?

13. Since you became 65, has anyone you know well ever pinched or scratched you, or pulled your hair?

14. Since you became 65, has anyone you know well ever kicked or bitten or punched you?
15. Since you became 65, has anyone you know well ever hit you or tried to hit you with something?

16. Since you became 65, has anyone you know well ever beaten you up?

17. Since you became 65, has anyone you know well ever threatened you with a knife or a gun?

18. Since you became 65, has anyone you know well ever stabbed you or tried to stab you with a knife or shot at you with a gun?

19. Since you became 65, has anyone you know well ever done or said any other violent things to you that have not been mentioned before?

(FINANCIAL MISTREATMENT)

Finally, I am going to ask you some questions about financial mistreatment.

20. Since you became 65, has anyone you know well ever taken money from you without your permission?

21. Since you became 65, has anyone you know well ever taken any of your possessions without your permission? For instance, jewelry, valuables, the deeds of a house, etc.

22. Since you became 65, has anyone you know well ever purchased things at your expense without your permission? For instance, groceries or durable goods such as a TV, a video-recorder, etc.

23. Since you became 65, has the person who manages your financial affairs ever taken advantage of you?

NOTES

1. Neglect (Q1-Q6) was evaluated by means of six questions, based on items from a modified version of an ADL questionnaire (Katz, et al., 1963).
2. Chronic verbal aggression (Q7-Q9) was measured according to three items from a revised and translated version of the Conflict Tactics Scale (Straus 1979, Yin 1985) and certain items from the Measure of Wife Abuse (Rodenburg &
3. Physical aggression (Q10-Q19) was assessed by means of certain items from a revised and translated version of the Conflict Tactics Scale, and items from the Measure of Wife Abuse and the Violence Against Man Scales (Marshall, 1992).

4. Financial mistreatment (Q20-Q23) was measured by means of two questions of the Measure of Wife Abuse and two newly developed questions.