Methods for delineating personal networks in surveys contain complex instructions for the interviewers. It is assumed that the interviewers' experience and education influence their ability to follow these instructions. The magnitude of the interviewer effects on the personal network size has been investigated, and differences among interviewers have been explained on the basis of their experience and education. The data are from a survey among 4,059 older adults in the Netherlands interviewed in 1992 by 87 interviewers. A strong interviewer effect was observed. Furthermore, the results of a multilevel regression analysis showed that, controlled for respondent characteristics, well-educated interviewers with minor experience prior to the project and major experience within the project (i.e., the high sequence number of the interview) generated relatively large networks.

Interviewer Effects in the
Measurement of Personal Network Size

A Nonexperimental Study

THEO VAN TILBURG
Vrije Universiteit Amsterdam

*I*nterviewer effects are one of the sources of errors within surveys. According to Groves (1989), interviewer effects differ largely across surveys. He reported that relatively large interviewer effects have been observed in studies where sensitive or emotional topics were under investigation, open questions were used, questions were asked that respondents found difficult to understand, or poorly trained interviewers were used. Interviewer effects on the respondents' answers stem from the different ways interviewers administer a survey. This study focuses on whether there are any differences in the measured personal network size (i.e., the number of persons within the network) of respondents across interviewers and, if so, whether the interviewer effects are related to interviewers' experience and education.

We focused on personal networks because their delineation deviates from the procedures applied to many other questions within interviews. A common characteristic of network delineation methods is that one or more multiple response questions containing the delineation criteria, for example, on closeness or contact frequency are repeated to generate the names of network members. The use of this kind of procedure makes it a complex task for the respondent as well as the interviewer. Network delineation calls for cognitive skills on the part of respondents and interviewers. Respondents have to interpret the question and select the persons who fit the criteria (e.g., to select certain children from all the children available). The interviewer has to check the respondent's answer (Is the name valid? Has the person already been named before?) and to decide after each response whether all the network members who fit the delineation criteria have been named by the respondent. If not, the procedure has to be continued by repeating the delineation question. If all the members have been named and more delineation questions are used, the interview has to be continued by asking the next question. Furthermore, repeating the questions may cause boredom or a lack of motivation among the respondents and interviewers. Given the complexity of delineating networks, measurement errors are to be expected. Test-retest reliability studies of personal networks delineation methods conducted by Broese van Groenou, van Sonderen, and Ormel (1990); Pfennig, Pfennig, and Mohler (1991); and Bass and Stein (1997) have made it clear that the average network size within a sample is relatively stable, but the stability of the network composition and of the individual network size is relatively low. This might be related to interviewer effects, but studies into these effects have not yet been conducted.

Our first research question is whether there are large interviewer effects within the delineation of personal networks. We expected to find a relatively large interviewer effect for two reasons: (a) listing the personal network members is in the category of difficult questions, and (b) there is a parallel between open questions and the delineation of personal networks. For open questions, there is a particular interviewer effect on the number of different answers by a respondent (Hox, de Leeuw, and Kreft 1991). One of the factors relevant in this connection is the number of probes for answers given by interviewers. Feldman, Hyman, and Hart (1951) have shown that differences be-

tween the number of answers to open questions are of specific interest when the distribution between primary categories (concerning answers that are frequently given) and secondary categories is considered. Interviewers who get a small number of responses have a particularly small number within the secondary categories. Feldman et al. attributed this to insufficient probing by interviewers and linked the amount of received information to the quality of the interviewing. The purpose of delineating networks is to make an exhaustive inventory of the actual network members, and to do so, numerous probes have to be given by the interviewers to complete the inventory. The larger the number of nominated network members, the larger the chance the aim of the procedure will be reached. This is even true if network members are nominated who do not fit the delineation criteria. In most studies, in addition to the network delineation, relationship characteristics are collected that cover the delineation criteria. In that case, after the data collection, the investigator can delete relationships that do not fit the criteria. In answering the first research question, it was possible for interviewer effects to be caused by poorly trained or poorly functioning interviewers and to be not specific for a specific set of questions. Therefore, we compared the interviewer effect on the network delineation with that of a less complex set of questions, that is, a standardized item scale on loneliness.

Our second research question, which is only to be posed if the answer to the first one is positive, is whether larger networks are observed among respondents interviewed by well-educated and experienced interviewers than among other respondents. The interviewers' education and experience might influence their ability to follow complex instructions. For education, a positive effect was expected. For experience, Sheatsley (1951) distinguished two effects on interviewer performance. The more experienced interviewer has learned more about shortcuts and how to cheat without detection, resulting in a negative effect. If interviewers are hired for a longer period by a particular survey organization, they become more accustomed to the organizational assignments, whereas interviewers who do not behave as required are dismissed. This results in a positive effect. In many cases, interviewer effects cannot be consistently reduced to specific characteristics of interviewers (Hox et al. 1991). One exception has been noted for racial topics; the racial background of the interviewers

and the answers of the respondents were found to be associated in a number of studies. Mixed results have been found for interviewers' education and experience. In a study by Berk and Bernstein (1988), interviewers' education and experience were not associated with obtaining complete charge data on health expenditures from the respondents or with the validity of these data collected from respondents by comparing it to verification data obtained from their physicians. Booker and David (1952) examined whether, using the same questions, interviewers in three surveys gathered a different number of answers to questions listing several items, such as magazines read regularly. Two of the three groups of interviewers were experienced, one was not. The three groups only differed significantly with respect to one of the five questions, with the smallest number of items listed by respondents interviewed by the least experienced group of interviewers. A study by Stember and Hyman (1949/1950) showed a higher consistency between the interviewers' field classification of responses and the codes on the verbatim answers recorded by the interviewers for experienced than for inexperienced interviewers. Carton (1995) found that poorly educated interviewers obtained significantly less useful information from their respondents than well-educated interviewers, but she did not find an effect for the interviewers' experience. We did not confine ourselves to the experience of interviewers before the start of the project or to their formal education, but we also examined the experience and training obtained during the data collection. Usually, interviewers are trained at the beginning of the data collection. If management devotes more attention to the quality of the interviewing, interviewers are monitored individually during the data collection, receive additional instructions if necessary, and are repeatedly trained.

DESIGN OF THE STUDY

RESPONDENTS

The personal networks of older adults were delineated in a longitudinal study in the Netherlands (Knipscheer et al. 1995). At T_1 (1992), face-to-face interviews were conducted with 4,494 respondents be-

tween the ages of 55 and 89. Respondents were interviewed in their homes, and personal computer assistance was used in the data collection. The sample was stratified according to year of birth and sex and was randomly taken from the registers of 11 municipalities in the east (one city and four rural municipalities), the south (one city and two rural municipalities), and the west (Amsterdam and two rural municipalities). These three regions can be taken to represent differences in culture, religion, urbanization, and aging in the Netherlands.

Two municipalities (including Amsterdam) forced us to adopt a two-step sampling procedure. In the first step, the municipalities asked the people in the sample permission to pass on their names and addresses to the researchers. In the second step, the researchers sent the remaining people in the sample who lived in the two municipalities and all the people in the sample who lived in the other municipalities ($N = 7,574$) a letter introducing the study, and the interviewers approached the prospective respondents. People who were deceased (5.4%), too ill to be interviewed (7.0%; for some of them proxies were interviewed), or did not speak Dutch (0.2%) were classified as ineligible. Many (27.6%) refused, and a small proportion (0.4%) was not contacted. On the basis of the response results in Step 2, the nonresponse in Step 1 was divided into eligible (i.e., refusals or not contacted) and ineligible. It was estimated that a total of 2,785 people in the sample refused or could not be contacted and 1,416 were ineligible. The response was 61.7 percent, computed as the proportion of the number of face-to-face interviews conducted from the number of eligible sample members. The networks of 4,059 respondents were delineated (90.3%, $N = 4,494$). For the others, 33 (0.7%) refused their cooperation for this section; with 37 (0.8%), the interview was prematurely terminated; with 345 (7.7%), a short version of the questionnaire, not including the network section, was used, due to the respondents' incapacity; and with 20 (0.4%), the network section was skipped due to technical problems.

INTERVIEWERS

Concerning the recruitment of the interviewers, two major decisions were made. First, interviewers were recruited who lived in or near the municipalities in the study. One of the reasons was to save on

traveling costs and time. An additional advantage was that local interviewers could understand or speak the dialect spoken in the region. Particularly in the east and south of the Netherlands, this was expected to increase the participation on the part of the respondents. Another reason was that the respondents were to be approached by the interviewers in person to make an appointment for the interview. This was expected to increase the response of the older adults. Because it was expected that, on average, several attempts to contact the respondent would be necessary before making an appointment, local interviewers were preferred. In addition to living in the region, a few other criteria were used in selecting the interviewers, such as availability and flexibility in spare time during the day, general educational level, and being between 22 and 55 years old. A total of 87 interviewers (10 men and 77 women) were employed, with an average age of 37.3 ($SD = 10.3$, range 21-58). The interviewers received an average of 62 Dutch guilders (about U.S.\$37) per interview, traveling costs included. On the average, the interviewers conducted 47 interviews ($SD = 27$, median $= 40$), with a range from 2 to 157.

Interviewer assignments were geographically defined. Four regions were distinguished: Amsterdam, the rural west, the east, and the south. Within the regions, 33 areas were distinguished on the basis of the first three digits of the postal code. A number of these postal code areas were designated as the employment area of an interviewer. Within that area, respondents were assigned randomly to the interviewer. All but one interviewer were employed in only one of the four regions. The median number of respondents living in a specific postal code area was 125 (range 1-533), and the median number of interviewers employed was 7 (range 1-13). The interviewers were free to conduct the interviews whenever it suited the respondents and themselves. They were encouraged to make appointments for interviews at times that were convenient to the respondents.

The data collection consisted of four periods. For the first, second, and third period, each eight weeks, prospective respondents were sampled from the stock. In principle, an interviewer was employed for one period and had to complete the assigned list (30 to 40 interviews) within that period. Conducting a larger number of interviews was encouraged. If a respondent was eligible but had not refused or was not interviewed (e.g., temporarily hospitalized or out of the region) in

a specific period, the respondent was approached again in the next period. In the fourth period, which lasted five months, only respondents who had been approached earlier were contacted. Of the 87 interviewers, two were project researchers who conducted a small number of interviews. Of the others, 8 did not fulfill their task (they interviewed between 2 and 21 respondents, on the average 10.5), involuntarily (some were fired by the supervisor) or voluntarily (some found a regular job elsewhere); 34 performed their scheduled task (they conducted on the average 37 interviews) within one period; and with 43, the contract was extended after their first period of eight weeks, or they were hired for the next wave of the longitudinal data collection (they conducted on the average 62.7 interviews at T_1 and 50.3 at T_2).

The interviewers were trained for four days by the four regional supervisors and an assistant. On the first and second day (four three-hour sessions), the interviewers were told the general interviewing rules and practiced putting them into effect with role-playing. Videotapes with common interviewing situations and mistakes were shown and discussed. All the sections of the questionnaire were discussed, and particular attention was devoted to the large number of different routings and to deviating question formats. Three sections of the questionnaire were practiced with another interviewer acting as the respondent. These sections contained questions on basic demographics, the network delineation, and an item scale for measuring loneliness. At any time, the interviewing program allowed the interviewer to practice with these three or all the sections. On the third day, a complete pilot interview with someone (ages 55-89) whom the interviewer knew was conducted in the respondent's home. The experiences with these interviews were discussed on the last training day (two three-hour sessions). All the procedures concerning contact with the respondents and with the research team were written down or were part of the computer program and were clarified during the training. During the data collection, the supervisors listened to tape recordings of the interviews. They weekly discussed interview style, suggestive questioning, handling difficult situations, administrative matters, and so forth with the interviewers. A report with evaluations of the interviewers' behavior was regularly drawn up. Three to four weeks after

the start of the data collection within a period, a four-hour training was held, including a meeting to discuss all kinds of interviewing problems.

APPLIED NETWORK DELINEATION METHOD

The main objective was to identify a network that reflected the socially active relationships of the older adults in the core as well as the outer layers of the larger network (van Tilburg 1995). In choosing a method to identify the personal network, several criteria were applied regarding who was to be included in the network. First, the network composition had to be as varied as possible, implying that all types of relationships deserved the same chance to be included in the network. This criterion led to a domain-specific approach in the network identification, using seven formal types of relationships (see below). A second objective was to include all the network members the respondent had regular contact with, thus identifying the socially active relationships. Yet, the aim was not to include everybody who was in contact with the older adult. To avoid eliciting persons who were contacted frequently by definition (such as all the respondents' colleagues), the criterion of importance of the relationship was added. Network members were identified in seven domains of the network: household members (including the spouse, if there was one), children and their partners, other relatives, neighbors, colleagues from work (including voluntary work) or school, members of organizations (e.g., athletic clubs, church, political parties), and others (e.g., friends and acquaintances). With respect to the domains, the question was posed: "Name the people (e.g., in your neighborhood) you have frequent contact with and who are also important to you." Only people above age 18 could be nominated. A limit of 80 was set on the number of names, but no one reached this limit. Information was gathered on all the identified network members with regard to the type of relationship, sex, and frequency of contact. For a subset of network members (i.e., the 12 with the highest frequency of contact), questions were posed on their age, traveling time to reach the network member, their marital or partner status, negative aspects within the relationship, and the receiving and giving of instrumental and emotional support.

OTHER RESPONDENT MEASUREMENTS
AND INTERVIEW CHARACTERISTICS

Loneliness

In this study, loneliness was defined as a situation experienced by the individual as exhibiting an unpleasant or unacceptable discrepancy between the quantity and quality of actual social relationships and the desired social relationships. This description included situations with fewer existing relationships than was considered desirable or admissible, as well as situations lacking in the desired intimacy. To measure the degree of loneliness, an instrument was developed that consisted of five positive and six negative items (de Jong Gierveld and Kamphuis 1985). The positive items assessed feelings of belonging, whereas the negative items applied to aspects of missing relationships. An example of a negatively formulated scale item was, "I experience a sense of emptiness around me." An example of a positively formulated item was, "I can rely on my friends whenever I need them." The loneliness scale had a range of 0 (*not lonely*) to 11 (*extremely lonely*). The scale has been used in several surveys and proved to be a reliable and valid instrument (van Tilburg and de Leeuw 1991). In the current study, the items formed a hierarchically homogeneous scale (Loevingers $H = .33$), which was reliably measured ($\rho = .81$). The homogeneity and reliability were computed with the MSP program (Molenaar et al. 1994).

Basic Demographics

By means of a number of questions, for example, on marital status and household composition, the availability of a spouse or partner in or outside the household was assessed. The availability of children and siblings was assessed by having the respondents nominate them by name. The availability of children-in-law was assessed by asking about the partner status of the children. The assessment of church affiliation was based on three questions: Is the respondent a member? If so, of which church? How strong is the affiliation? The answers were divided into four categories: (a) not a member or not in the least affiliated, (b) Roman Catholic, (c) Protestant, (d) a member of some

other church. The respondent's employment status was based on a number of questions and rearranged into employed for more than 10 hours a week and not employed. An index for socioeconomic status was constructed on the basis of education, skill level of the occupation, occupational prestige (based on a coding scheme devised by Sixma and Ultee 1983) and household income. Because the women's status was often derived from their partner's status, the partner's score was taken if one of the first three respondent scores was lower. The four scores had high intercorrelations (average $r = .62$; Cronbach's $\alpha = .87$). The scores were transformed to ranges from 0 to .33, .17, .17, and .33, respectively, and summed to create one index with a range from zero to one. The level of urbanization was measured in five ordinal classes, ranging from (1) *not urban* to (5) *very highly urban*. These data were derived from a database provided by the Netherlands Central Bureau of Statistics on the basis of the mean number of addresses per square kilometer within a circle with a radius of one kilometer (den Dulk, van de Stadt, and Vliegen 1992). The aim was to measure the concentration of human activities.

Health

As regards the health aspect, three instruments were used. The first one consisted of four questions about having difficulties with activities of daily living (ADL): "Can you walk up and down stairs? . . . walk for five minutes outdoors without resting? . . . get up and sit down in a chair? . . . get dressed and undressed (including putting on shoes, doing up zippers, fastening buttons)?" The five answers were the following: not at all, only with help, with a great deal of difficulty, with some difficulty, and without difficulty. The four ADL items constituted a hierarchically homogeneous scale ($H = .68$), which was reliably measured ($\rho = .87$). The scale ranged from 4 (*numerous problems*) to 20 (*no problems*). ADL capacity correlated $-.33$ with age. The second instrument pertained to the respondents' perception of their own health: How is your health in general? Answers could be given on a five-point scale scored from one to five. Subjective health correlated with age ($r = -.12$) and ADL capacity ($r = .42$). The third instrument was an interviewer evaluation of the respondent's cogni-

tive functioning during the interview. The interviewers answered eight questions at their home: Were there problems (e.g., in comprehension) that made it difficult to interview the respondent? How much help (e.g., repeating questions, extra explanations) did the respondent need to answer the questions? In general, how good was the respondent's comprehension of the questions? Did the respondent have difficulty recalling experiences from the past? How good was the respondent at expressing answers? How good was the respondent's ability to concentrate? Did the respondent experience difficulties with the answer categories? How good was the respondent's memory in general? The eight standardized scores had sufficiently high intercorrelations (average $r = .58$; Cronbach's $\alpha = .92$) to construct a scale score by summing the scores; the scale scores were converted to a range from zero to one. The interviewers' evaluation of the respondents' cognitive functioning correlated with age ($r = -.38$), ADL capacity ($r = .30$), and subjective health ($r = .22$).

The duration of each section of the interview was measured by the internal clock of the computer. Breaks of 10 minutes or longer were not counted. The interviews lasted an average of 98 minutes, 10 of which were used to delineate the network, and 45 before and 43 after the network delineation. After the interview, the interviewers recorded whether third parties were present during the interview and, if so, who they were.

INTERVIEWER MEASUREMENTS

In addition to their sex and age, a number of characteristics were collected from all the interviewers. Their general attitude toward older adults was measured with 10 items, based on a semantic differential developed by Osgood, Succi, and Tannenbaum (1957). These items were fast-slow, boring-interesting, clean-dirty, unimportant-important, strong-weak, pleasant-unpleasant, attractive-repulsive, good-bad, happy-unhappy, and active-passive, with values from 1 to 9. A scale (average $r = .40$; Cronbach's $\alpha = .87$) was constructed by summing the scores and, if necessary, the reversed scores (range 10-90). The actual range of the values was between 35 and 89 ($M = 60.4$, $SD = 10.5$).

Experience with interviewing prior to the project was measured on an ordinal scale with four categories: (a) never conducted interviews (37%), (b) conducted some interviews (25%), (c) conducted one or two surveys (23%), and (d) had been involved as interviewer in a number of survey projects (15%). The interviewers' experience within the project was indicated by the interviewers' sequence number of the interview. Education was measured on an ordinal scale with seven categories ranging from lower vocational school to university. The scale was transformed into years of education ($M = 13.2$, $SD = 2.3$). Prior project experience, the number of interviews conducted, and educational level were not significantly correlated ($|r| < .13$).

The supervisors rated the performance of the interviewers during the training and the interviewing up to a maximum of nine times (the average number of ratings was 3.1, $SD = 1.5$; interviewers conducting more interviews were rated more often; $r = .80$). Thirteen aspects were rated, most of them on a nine-point evaluation scale. For the others, three to five answering possibilities were used. For all the aspects, a higher score indicated better interviewing quality. Using an explorative principal components factor analysis with varimax rotation, four factors with an eigenvalue of one or larger were derived (explained variance 68.9%). The first factor consisted of six items with a loading higher than .50, in decreasing sequence: reading questions as they were worded, a global overall evaluation, adequately explaining the questions, adequately repeating questions, being conscientious and meticulous, and probing in a correct, nonsuggestive, and nondirective way. The second factor consisted of four items: being socially confident, degree of extroversion, being pleasant, and being good at ending unpleasant situations. The third factor consisted of two items: adequately providing feedback and posing questions in a natural tone and speaking clearly. The final factor consisted of two items: having a correct interviewing speed and being pleasant. The first and second factors were evaluated as the most relevant and easiest to interpret as technical and social interviewing quality, respectively. Factor scores for these two factors were derived from the analysis and rearranged from 0 to 1. The only significant correlation with the other interviewer characteristics was between social interviewing quality and age ($r = .37$).

PROCEDURE

To assess the magnitude of the interviewer effect, the intraclass correlation ρ_{int} developed by Kish was used, computed on the basis of analysis of variance (Groves 1989). We also used multilevel analysis (ML3) (Prosser, Rasbash, and Goldstein 1991) to compute ρ_{int}. In his review of a number of studies, Groves (1989) reported an overall mean of .03 for ρ_{int}. He evaluated interviewer effects as large if ρ_{int} was about .20; interviewer effects could be considered small if ρ_{int} was smaller than .02.

A hierarchical multilevel regression analysis with forced entry of explanatory variables in four steps was conducted to examine whether differences in network size could be explained by differences in the interviewers' experience and level of education. The respondents were on the lower and the interviewers on the higher level. Interviewers were not assigned to equivalent respondent categories. Therefore, respondent characteristics were first entered into the equation as control variables. They included sex, age, availability of a partner, the number of children and siblings, socioeconomic status, church membership and affiliation (dummy variables with not a member and not affiliated as category of reference), being employed, ADL capacity, subjective health, and cognitive functioning. For all of them, there was evidence from earlier studies of associations with network size. Females, younger old people, and people with a partner relationship had more relationships. For the latter, not only the partner relationship itself but also a broader network became available via the partner. If children and their partners and siblings were available, they were likely to be in the network. By a matching procedure, it was assessed that, on average, the respondents nominated 86 percent of their children ($n = 3,454$; respondents with children only) and 43 percent of their siblings ($n = 3,378$; respondents with siblings only) in their networks. Furthermore, the respondents nominated 67 percent of their children-in-law ($n = 3,229$; respondents with children-in-law only). The number of children was strongly associated with the number of children-in-law ($r = .91$), so there was no need to enter the latter into the equation. There were more chances for people with a higher socioeconomic status to engage in personal relationships, and they tended to be more attractive to other people. The church was an

important source of social contacts. Employed people had opportunities to engage in work-related personal relationships, although they might have less spare time to meet people outside their work. People with poor physical capacities or cognitive functioning had fewer opportunities to meet other people, although informal helpers might enter the network. In addition, people with poor cognitive functioning might have problems with being interviewed and with remembering network members. The geographical region (dummy variables with Amsterdam as category of reference) and urbanization were entered as controls because the sampling procedure and the data collection were region based.

To answer the second research question, in the second step, the prior experience, the sequence number of the interview, and the education level of the interviewer were entered into the equation. In an explorative third step, other control and explanatory variables were entered to evaluate whether they would modify the effects of interviewers' experience and education. Characteristics of the interview were entered, including being interviewed in the fourth period of the data collection (respondents, interviewers, and procedures in that period deviated from the other periods), the duration of the interview before the network delineation, and whether third parties were present during the interview. Furthermore, the interviewer's sex, age, and attitude toward older adults; the number of interviews conducted (covering the differences between the interviewers who quit early, conducted the scheduled number of interviews, and had an extended contract); and the technical and social quality of the interviewing were entered. Finally, the significance of a nonlinear effect and a number of interaction effects were explored. A nonlinear (quadratic) effect of the interview sequence number combines the positive effect of the increasing experience within the project and a negative effect. It was known from studies by Cannell, Marquis, and Laurent (1977) and van der Zouwen and Dijkstra (1988) that the quality of interviewing could decrease if an interviewer conducted numerous interviews. When there is a positive linear and a negative quadratic effect, it indicates that the increasing experience of interviewers has the strongest (positive) effect at earlier stages of the data collection and that this positive effect decreases at later stages. The interaction effects included were of (a) respondent and interviewer characteristics (sex, age, and edu-

cation) as control variables, (b) the sequence number of the interview and education of the interviewer to assess whether poorly educated interviewers learned more from the interviewing than well-educated interviewers, and (c) the sequence number and (d) prior experience, on one hand, and the duration of the interview, on the other, to assess whether more experienced interviewers tried to be efficient with their time. To avoid multicollinearity, all the variables for these effects were centered before interaction and quadratic terms were computed. In total, 37 explanatory variables were entered into the equation. For the significance of the regression coefficients, t values were computed by dividing the coefficient by its standard error. For variables at the respondent and the interviewer level, there were an indefinite number and 86 degrees of freedom, respectively. For computing standardized regression coefficients, the variables were standardized.

Within multilevel analysis, there are two methods to evaluate the fit of models. The first one focuses on the significance of the model change. Each model is characterized by the -2 log likelihood (deviance, i.e., the lack of fit between the model and the data). We applied the forward modeling approach using an empty model (only containing a constant) at the start and adding the explanatory variables one by one (dummy variables were entered as a block). The difference between the deviance of the successive models is χ^2 distributed with the number of added parameters as degrees of freedom. The second method uses the decrease of the unexplained variance (Snijders and Bosker 1994). In each model, the variability of the dependent variable is estimated at both levels. The sum of these variance components in the empty model equals the variance of the variable. By adding explanatory variables to the model, the variance will decrease for either one or both levels. The amount of decrease provides insight into the explanatory power of the model. Unlike ordinary regression analysis, the added explained variances might be negative. If they are strongly negative, the specification of the model should be doubted.

RESULTS

The average network size was 13.4 ($SD = 9.4$, median = 11) with a minimum of 0 and a maximum of 77. The average loneliness score

was 2.3 (SD = 2.7, median = 1, range 0-11). Both variables had highly skewed distributions, and the decision was made to use original scores and scores transformed to approximately a normal distribution in parallel analyses. For network size, ρ_{int} was computed as .200 (based on analysis of variance), .214 (based on multilevel analysis with original scores), and .252 (based on multilevel analysis with normalized scores). For the loneliness scores, ρ_{int} was .021, .020, and .022, respectively. These figures indicated that the interviewer effects were large for the measurement of network size, not for the measurement of loneliness.

Table 1 shows the respondent and interviewer characteristics, broken down by the region. For most of the respondent characteristics, differences between the regions were observed. Interviewers employed in Amsterdam were younger and interviewed fewer respondents than the other interviewers. Both of these differences might be related to the fact that there are two universities in Amsterdam, so a number of interviewers were recruited who had recently completed their education and were doing the interviews while they were looking for a regular job. Interviewers in the rural west were rated by their supervisor as relatively highly socially skilled. We do not have an explanation for this.

The results of the multilevel regression analysis based on the normalized scores were very similar to those based on the original scores. Therefore, only the latter are shown in Table 2 and will be discussed. Equation (1) contains the respondents' characteristics (including the region) and the interviewers' experience and education. For all the respondents' characteristics, being employed excepted, effects in the expected direction were observed. The significance of ADL capacity disappeared as soon as cognitive functioning entered the equation. On the respondent level, about 15 percent of the variance was explained, indicating that there were still large differences within the categories of respondents with the same characteristics. On the interviewer level, about 11 percent of the variance was explained. This indicates that respondents assigned to different interviewers had different characteristics that influenced the average network size across interviewers. In addition, the geographical region produced differences in the average network size of respondents interviewed by specific interviewers (10% added explained variance on the inter-

TABLE 1: Respondent and Interviewer Characteristics by Region

	All	Amsterdam	West (Rural)	East	South	Significance Level
Respondent characteristics	$N = 4,059$	$N = 1,133$	$N = 676$	$N = 1,293$	$N = 957$	
Network size (*M*, range 0-77)	13.4 (9.4)	11.6 (8.8)	12.9 (8.6)	15.9 (10.0)	12.7 (9.3)	**
Loneliness ($N = 4,044$; *M*, range 0-11)	2.3 (2.7)	2.8 (2.9)	2.3 (2.5)	2.1 (2.5)	2.1 (2.7)	**
Sex (% female)	51	50	49	51	54	
Age (*M*, range 55-89)	72.2 (9.9)	73.2 (9.9)	71.4 (9.3)	71.3 (10.1)	72.6 (9.9)	**
Percentage with partner	64	59	66	67	62	**
Number of children (*M*, range 0-14)	2.8 (2.2)	2.2 (1.8)	2.8 (2.1)	3.0 (2.1)	3.3 (2.5)	**
Number of siblings (*M*, range 0-15)	2.8 (2.5)	2.4 (2.5)	2.6 (2.3)	2.8 (2.3)	3.4 (2.7)	**
Socioeconomic status (*M*, range 0-1)	.33 (.19)	.35 (.20)	.31 (.18)	.33 (.19)	.32 (.19)	**
Church, membership and affiliated (%)						**
None	42	70	50	28	22	
Roman Catholic	45	21	42	45	75	
Protestant	12	6	7	26	2	
Other	2	3	1	1	1	
Employed (%)	9	9	9	11	7	
ADL capacity (*M*, range 4-20)	18.6 (2.7)	18.6 (2.5)	18.8 (2.4)	18.9 (2.3)	18.2 (3.3)	**
Subjective health (*M*, range 1-5)	3.7 (.9)	3.5 (.9)	3.7 (.9)	3.7 (.8)	3.7 (.9)	**
Cognitive functioning (*M*, range 0-1)	.75 (.16)	.73 (.16)	.75 (.15)	.77 (.16)	.76 (.16)	**
Urbanization (*M*, range 1-5)	3.0 (1.5)	4.8 (.5)	2.0 (.9)	2.2 (1.0)	2.7 (1.1)	**

Interviewer characteristics	$N = 87$	$N = 31$	$N = 15$	$N = 24$	$N = 17$	
Sex (% female)	89	77	87	96	100	
Age (M, range 21-58)	36.6 (10.4)	27.1 (5.4)	41.2 (10.4)	41.4 (7.3)	43.3 (8.9)	**
Attitude toward older adults (M, range 35-89)	60.0 (10.2)	57.9 (11.7)	59.3 (9.0)	63.0 (9.8)	60.2 (8.6)	
N of interviews conducted (M, range 2-157)	46.7 (27.1)	36.7 (27.7)	44.8 (20.1)	53.9 (24.4)	56.3 (30.2)	*
Prior experience (M, range 1-4)	2.2 (1.1)	2.5 (.9)	2.2 (1.3)	1.9 (1.1)	1.9 (1.1)	
Educational level (M, range 9-18)	13.2 (2.3)	13.6 (2.1)	12.3 (2.4)	13.0 (2.5)	13.3 (2.2)	
Technical quality interviewing (M, range 0-1)	.70 (.21)	.70 (.19)	.73 (.20)	.65 (.26)	.72 (.18)	
Social quality interviewing (M, range 0-1)	.48 (.17)	.41 (.15)	.65 (.17)	.46 (.11)	.49 (.16)	**

NOTE: Significance for nominal variables based on the χ^2, for others, on the F statistic. *SDs* in parentheses. ADL-activities of daily living.
*$p < .05$. **$p < .001$.

viewer level). In particular, the interviewers employed in the east "produced" large networks. The added explained variance by region and urbanization on the respondent level was zero, which indicates that the differences between respondents across the regions were effectively controlled for. The partial ρ_{int} after the introduction of all these variables was still high (.205).

The three variables on the interviewers' experience and education added substantially to the explained variance on the interviewer level (15%) and were all significant. These effects were controlled for region and indicate that interviewers employed within a specific region who had different characteristics "produced" different network sizes. The partial ρ_{int} was reduced to .173. The effect of prior experience indicated that the least experienced interviewers generated, on average, 3.4 more network members than the most experienced interviewers. The interviewers with the highest education level generated, on average, 4.1 more network members than the interviewers with the lowest education level. The sequence number of the interview, indicating the experience and training of the interviewers during the project, had a very skewed distribution and was consequently normalized. The difference between the first and the median last interview (with sequence number 40) was estimated as 1.4, in favor of the latter.

Other control and explanatory effects were added to equation (1). Of the variables entered into Equation 1, particularly the effects of having a partner and of the region, were modified. All the effects for the interviewers' experience and education became stronger. Four added variables either had a significant effect in the final equation (2) or improved the model significantly. The respondents interviewed in the fourth period of the data collection who had specific characteristics, such as being hard to reach or recovering from an illness, had smaller networks than the others. There was no effect for the number of interviews conducted by the interviewer. The duration of the interview prior to the network delineation, which was normalized, had no main effect, and there were no significant interaction effects observed for this variable. Two contrary effects might counterbalance each other here. The respondents who took more time answering the questions also took more time recalling significant network members. Otherwise, if it took them a long time to answer the preceding questions, during the network delineation, time-efficient interviewers

TABLE 2: Multilevel Regression of the Network Size on Experience and Education of the Interviewer, Controlled for Respondent, Interview, Interviewing, and Other Interviewer Characteristics

	Equation 1			Equation 2			Change Deviance (χ^2)	Explained Variance (% added) at the Level of	
	B	β	t	B	β	t		Interviewers	Respondents
Constant	2.28		1.4	2.13		.6			
Characteristics respondent									
Sex (male-female)	.83	.04	3.0**	1.04	.06	3.7***	.3	−.2	.0
Age (55-89 years)	−.09	−.10	−5.6***	−.11	−.11	−6.2***	229.7***	2.8	5.6
With partner (no-yes)	1.52	.08	5.0***	.58	.03	1.5	63.0***	1.7	1.4
Number of children (0-14)	.69	.16	11.5***	.65	.15	10.1***	132.3***	.8	3.0
Number of siblings (0-15)	.22	.06	4.1***	.21	.06	3.9***	12.6***	−1.0	.3
Socioeconomic status (0-1)	5.56	.11	7.3***	5.85	.12	7.5***	108.9***	3.5	2.3
Church membership and affiliation[a]							54.5***	4.1	1.1
Roman Catholic	1.88	.10	6.3***	1.86	.10	6.2***			
Protestant	2.11	.07	4.7***	2.10	.07	4.7***			
Other	1.76	.02	1.7	1.69	.02	1.7			
Employed (no-yes)	−.11	−.00	−.2	−.18	−.01	−.4	.0	.0	.0
ADL capacity (4-20)	.05	.02	1.0	.06	.02	1.1	9.2**	.6	.2
Subjective health (1-5)	.18	.02	1.2	.18	.02	1.2	3.3	.1	.1
Cognitive functioning (0-1)	7.24	.12	7.5***	7.88	.13	8.1***	55.5***	−1.6	1.2
Geographic region[b]							9.8*	10.3	.0
Rural west	−.48	−.02	−.4	2.45	.10	1.6			
East	1.63	.08	1.4	3.41	.17	2.5*			
South	−2.43	−.11	−1.9	−.21	−.01	−.1			
Urbanization (1-5)	−.16	−.02	−.8	−.17	−.03	−.8	.5	−.8	.0
Experience and education interviewer									
Prior project experience (1-4)	−1.12	−.13	−2.8**	−1.28	−.15	−3.4**	9.8**	10.1	−.0
Education (9-18 years)	.46	.11	2.5*	.49	.12	2.7**	5.9*	5.2	.0
Sequence number interview (0-1)[c]	2.02	.06	4.1***	2.50	.08	4.7***	16.7***	−.1	.4
Characteristics interview, interviewing, and interviewer									
Interviewed in the fourth period of the data collection (no-yes)				−1.37	−.03	−2.0*	3.7	−.2	.1

TABLE 2 Continued

	Equation 1			Equation 2			Change Deviance (χ^2)	Explained Variance (% added) at the Level of	
	B	β	t	B	β	t		Interviewers	Respondents
Number of interviews conducted by interviewer (2-157)				.00	.01	.2	.0	.0	.0
Duration interview (0-1)[c]				.73	.02	1.3	1.1	.4	.0
Partner present during the interview (no-yes)				1.52	.08	4.5***	20.7***	.3	.4
Other person present during the interview (no-yes)				.16	.00	.3	.2	.1	.0
Sex interviewer (male-female)				−1.54	−.05	−1.1	1.4	1.3	−.0
Age interviewer (21-58 years)				−.11	−.12	−1.9	6.6*	6.3	−.0
Interviewer's attitude toward older adults (35-89)				.01	.01	.3	.2	.1	.0
Technical quality interviewing (0-1)				.45	.01	.2	.1	−.1	.0
Social quality interviewing (0-1)				−5.94	−.10	−2.0*	4.2*	3.8	−.0
Nonlinear and interaction effects									
Sequence number interview (quadratic)				−.61	−.00	−.3	.0	.0	.0
Sex interviewer × Sex respondent				−.50	−.01	−.6	.4	.1	.0
Age interviewer × Age respondent				.0028	.03	2.3*	4.9*	−.4	.1
Education interviewer × SES respondent				.01	.00	.0	.0	.0	.0
Sequence number interview × Education interviewer				−.07	−.00	−.3	.2	.0	.0
Sequence number interview × Duration interview				−2.88	−.03	−1.9	3.5	.1	.1
Prior experience interviewer × Duration interview				−.20	−.01	−.5	.2	.1	.0
Total								47.3	16.3

NOTE: N interviewers = 87, N respondents = 4,059. Range of the independent variables in parentheses (dichotomous variables have values 0 and 1). Constant, Bs and βs of final equations. ADL=activities of daily living; SES=socioeconomic status.
a. Dummy variables with not a member or not affiliated as category of reference.
b. Dummy variables with Amsterdam as category of reference.
c. Normalized scores.
*p < .05. **p < .01. ***p < .001.

cut off the respondent as early as possible or gave fewer probes. The presence of a partner during the interview resulted, on average, in 1.5 more network members. For 62 percent of the respondents with a partner, the partner was present during the interview. Therefore, with the introduction of this effect, the effect of having a partner decreased by 1. It is not clear from the data whether the partner made suggestions about nominating certain network members or whether the presence of a partner served as a memory aid to nominate network members connected to the partner, such as siblings or colleagues. The presence of other people did not influence the network size.

The sex of the interviewer did not influence the network size. There was a main effect for age: The respondents interviewed by younger interviewers had larger networks than those interviewed by older interviewers. When it was entered, the effect was significant, but as soon as the social interviewing quality entered the equation, the significance was lost. In the final equation, the estimated deviations for the extremes of the interviewers' age were +1.9 and −1.9, respectively. Simultaneously, by entering the interviewers' age into the equation, the effects of the dummy variables for region changed dramatically for the rural west from −.2 to 1.6, for the east from 2.0 to 3.8, and for the south from −2.1 to .0. This might indicate that the relatively large network size of the respondents living in Amsterdam in equation (1) (they were ranked as second) was biased by their being interviewed by young interviewers. However, it is also possible that the effect of interviewers' age is caused by the employment of young interviewers in Amsterdam. In addition to the main effect of the interviewers' age, the interaction effect of the respondents' and interviewers' age increased the differences among the interviewers: Large networks were delineated by the youngest interviewers interviewing the youngest respondents (a deviation of +4.7 from the grand mean), and small networks were delineated by the oldest interviewers interviewing the oldest respondents (−2.9), with the combination of youngest interviewers and oldest respondents and the reversed combination in the middle (−.8 and −.9, respectively). No differences in network size were observed for interviewers with a different attitude toward older adults. The findings with respect to the interviewers' age and attitude were contrary to the findings of Freitag and Barry (1974). They argued that an interviewer's attitude toward aged people would

affect the findings in a survey of older adults. Their findings showed that interviewers retired from the labor force and interviewers who viewed older adults favorably elicited significantly more responses than the others. They suggested that the validity of interview data would be greater if differences between the interviewer and the respondent were minimized. However, our findings did not support this view. As regards the age of the interviewer, older adults might be more open to interviewers with an age difference of about two generations or might want to show young interviewers that they still have a large network despite their age. The difference with respect to the interviewers' attitude toward older adults might be caused by the changes in society in the two decades since the study by Freitag and Barry: A larger proportion of the population is aged, and it is now generally recognized that older adults do not constitute a homogeneous category. The technical quality of the interviewing did not influence the network size, but the social quality had a significant negative effect. We should bear in mind that the supervisors evaluated the whole interview, and not the network delineation, separately. The findings were quite unlike those of Dijkstra (1983), who observed no effect of interviewer style on the amount of personal information elicited. Dijkstra (1987) had two contrary hypotheses on the effect of a personal interviewing style. First, by showing interest and understanding, the interviewer motivated the respondent to fulfill the task requirements. The second hypothesis, which was supported by our results, was that respondents were more inclined to try to ingratiate themselves with the interviewer, leading to more irrelevant and less relevant information. The negative effect of the quadratic term of the sequence number of the interview indicated that, controlled for the linear effect, the largest networks were delineated in the middle of the interviewers' employment period. We should also bear in mind that interviews with a high sequence number were conducted by interviewers whose contracts were extended, whereas the contracts of others were not extended. However, the effect was small and not significant and contrary to the results of earlier studies. The other interaction effects in the equation were not significant. Within the final equation (2), 47 percent of the variance at the interviewer and 16 percent at the respondent level was explained. The partial ρ_{int} was reduced to .147.

DISCUSSION

Interviewer effects were studied using data from a large-scale national survey. There are disadvantages to using data from a nonexperimental study, such as the regional ties of interviewers and the unequal number of interviews conducted across interviewers. However, we corrected for these errors within the analysis by taking into account respondent and region characteristics and the number of interviews conducted by an interviewer (Hox 1994). Furthermore, the advantage of conducting a secondary analysis is that the particular data collection covers a large number of respondents and a large number of interviewers, which facilitates a reliable estimate of the interviewer effects (Hanson and Marks 1958). Due to the costs of delineating personal networks, this type of approach would be impossible if an experimental design had been used.

There were large interviewer effects in the measurement of personal network size: The respondents of some interviewers had small networks on average, and the respondents of other interviewers had large networks on average. We hypothesized that the experience of interviewers in conducting survey interviews and their education would influence their performance. For education, the positive effect was in accordance with our hypothesis and with Carton's (1995) study. The results of our study support the significance of the interviewers' experience. However, the results show that we should distinguish between experience within the project and prior to the project: Interviewers with less prior experience and also interviewers with a great deal of project experience generated relatively large networks. A possible explanation might be that the interviewers without prior experience were more open to following instructions than the more experienced ones, probably because the latter adhered more strictly to their perception of the instructions given within projects they had participated in before. Furthermore, the finding that major project experience produced larger networks can be interpreted to mean it took time to become experienced with a complex interviewing task such as delineating personal networks. A second explanation is that only effective interviewers conducted a large number of interviews,

whereas others quit or did not have their contracts extended. However, in the analysis, we controlled for that factor.

Our findings give rise to questions about the validity of studies on personal networks. One of the consequences of generating only a few network members and not having the exhaustive network size is that it limits the options for studying change over time in personal network size and composition. For example, a longitudinal study by Bowling, Grundy, and Farquhar (1995) showed that the network size changed in all directions: Three years later, 42 percent of the respondents had a smaller network than at the first observation, for 16 percent the network size increased, and for 42 percent no change was observed. Our longitudinal study also showed a large variation in the changes in the individual network size. It is not clear whether these findings were substantive or biased: A person who was nominated at only one observation may have entered or left the network (natural circulation in the network membership) (Starker, Morgan, and March 1993), or the appearance in one observation may have been caused by poor interviewing at the other observations or by the unreliability of the instrument applied.

Awareness of interviewer effects in the measurement of personal network size might enable researchers starting a data collection to avoid the large and systematic ones found in the current study. One solution could be to omit the interviewer and collect the data by mailed questionnaires or computer-assisted self-interviewing. However, the complexity of the questionnaire on personal network delineation makes these alternatives unfeasible. Furthermore, we doubt whether the oldest respondents would be apt to fill in a questionnaire (Kaldenberg, Koenig, and Becker 1994). For face-to-face interviewing, one way to diminish the interviewer effects is to select the interviewers more carefully and thus introduce more uniformity. This might work with regard to prior project experience. However, because our results showed that interviewers with major within-project experience generated large networks, and because we prefer to homogenize toward large networks, this solution could not be applied because it would be impossible to only employ interviewers with extended within-project experience. Another solution would be to accept the interviewer effect and randomize it. The number of interviews to be conducted by each

interviewer could be very strictly limited. However, this solution would not be cost-effective because more interviewers would have to be trained. More important, it would diminish the chance of obtaining the exhaustive network size.

Of course, training the interviewers should receive the utmost possible attention. Although we think this was already the case in the current project, the training was not very specific on the avoidance of interviewer errors related to this type of questioning. In a study using other types of difficult questions (Berk and Bernstein 1988), no effects of prior experience and education were observed, and Billiet and Loosveldt (1988) observed that respondents interviewed by trained interviewers enumerated more items than those interviewed by interviewers who were trained minimally. Furthermore, the evaluation of the social interviewing quality by the supervisors had a negative effect on the measured network size, which suggests that this aspect should be given more attention during the training. We conclude that there is not enough knowledge available to design a specific training for personal network delineation.

The interviewer effects found in this study might reflect differences in the ability of interviewers to perform the difficult task of delineating personal networks. As criterion for the performance of interviewers, we used the amount of information elicited. Because validating information was missing, one might doubt whether the interviewers who generated larger networks also performed better than the interviewers who generated smaller networks, although Billiet and Loosveldt (1988) assumed a positive association between the amount of information gathered by interviewers and the quality of the response. This study only provided insight into differences in the results of different interviewers, that is, the measured network size. It is not clear whether the interviewers behaved differently—in other words, whether they formulated questions or interpreted the respondents' answers in a different way. It is possible that interviewers with prior experience tried to be time efficient and cut off the respondents as early as possible or gave fewer probes. In the near future, we are planning to study the behavior of interviewers using tape recordings of the interviews to try to connect specific interviewer strategies (e.g., probing, cutting, inadequate or suggestive questioning) with the measured network size.

We hope this future study will reveal whether specific interviewer behavior is connected to differences in network size and will result in specific interviewer guidelines to diminish interviewer effects in the measurement of personal network size.

REFERENCES

Bass, Lee Ann and Catherine H. Stein. 1997. "Comparing the Structure and Stability of Network Ties Using the Social Support Questionnaire and the Social Network List." *Journal of Social and Personal Relationships* 14:123-32.

Berk, Mark L. and Amy B. Bernstein. 1988. "Interviewer Characteristics and Performance on a Complex Health Survey." *Social Science Research* 17:239-51.

Billiet, Jacques and Geert Loosveldt. 1988. "Improvement of the Quality of Responses to Factual Survey Questions by Interviewer Training." *Public Opinion Quarterly* 52:190-211.

Booker, H. S. and S. T. David. 1952. "Differences in Results Obtained by Experienced and Inexperienced Interviewers." *Journal of the Royal Statistical Society* 115:232-57.

Bowling, Ann, Emily Grundy, and Morag Farquhar. 1995. "Changes in Network Composition Among the Very Old Living in Inner London." *Journal of Cross-Cultural Gerontology* 10:331-47.

Broese van Groenou, Marjolein I., Eric F.L.P. van Sonderen, and Johan Ormel. 1990. "Test-Retest of Personal Network Delineation." Pp. 121-36 in *Social Network Research: Substantive Issues and Methodological Questions*, edited by C.P.M. Knipscheer and T. C. Antonucci. Lisse, the Netherlands: Swets & Zeitlinger.

Cannell, Charles F., Kent H. Marquis, and André Laurent. 1977. "A Summary of Research Studies of Interviewing Methodology, 1959-1970." *Vital and Health Statistics* 69:1-78.

Carton, Ann. 1995. "Towards an Effective Evaluation of Interviewers: Working Within the Context of a Permanent Network of Interviewers." Pp. 145-50 in *Proceedings of the International Conference on Survey Measurement and Process Quality*.

de Jong Gierveld, Jenny and Frans H. Kamphuis. 1985. "The Development of a Rasch-Type Loneliness-Scale." *Applied Psychological Measurement* 9:289-99.

den Dulk, C. J., H. van de Stadt, and J. M. Vliegen. 1992. "Een Nieuwe Maatstaf voor Stedelijkheid: De Omgevingsadressendichtheid" (A New Measure for Degree of Urbanization: The Address Density of the Surrounding Area). *Maandstatistiek van de Bevolking* 40 (July): 14-27.

Dijkstra, Wil. 1983. "How Interviewer Variance Can Bias the Results of Research on Interviewer Effects." *Quality and Quantity* 17:179-87.

———. 1987. "Interviewing Style and Respondent Behavior: An Experimental Study of the Survey-Interview." *Sociological Methods and Research* 16:309-34.

Feldman, Jack J., Herbert Hyman, and C. W. Hart. 1951. "A Field Study of Interviewer Effects on the Quality of Survey Data." *Public Opinion Quarterly* 15:734-61.

Freitag, Carl B. and John R. Barry. 1974. "Interaction and Interviewer Bias in a Survey of the Aged." *Psychological Reports* 34:771-74.

Groves, Robert M. 1989. *Survey Error and Survey Costs*. New York: John Wiley.

Hanson, Robert H. and Eli S. Marks. 1958. "Influence of the Interviewer on the Accuracy of Survey Results." *Journal of the American Statistical Association* 53:635-55.

Hox, Joop J. 1994. "Hierarchical Regression Models for Interviewer and Respondent Effects." *Sociological Methods & Research* 22:300-18.

Hox, Joop J., Edith D. de Leeuw, and Ita G. G. Kreft. 1991. "The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data: A Multilevel Model." Pp. 439-61 in *Measurement Errors in Surveys*, edited by P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman. New York: John Wiley.

Kaldenberg, Dennis O., Harold F. Koenig, and Boris W. Becker. 1994. "Mail Survey Response Rate Patterns in a Population of the Elderly. Does Response Deteriorate With Age?" *Public Opinion Quarterly* 58:68-76.

Knipscheer, Kees C.P.M., Jenny de Jong Gierveld, Theo van Tilburg, and Pearl A. Dykstra., eds. 1995. *Living Arrangements and Social Networks of Older Adults*. Amsterdam: VU University Press.

Molenaar, Ivo W., P. Debets, Klaas Sijtsma, and Bas T. Hemker. 1994. *MSP: A Program for Mokken Scale Analysis for Polychotomous Items (Version 3.0)*. Groningen: ProGAMMA.

Osgood, Charles E., George J. Succi, and Percy H. Tannenbaum. 1957. *The Measurement of Meaning*. Urbana: University of Illinois Press.

Pfennig, Astrid, Uwe Pfennig, and Peter Ph. Mohler. 1991. "Zur Realiabilität von Egozentrierten Netzwerken in Massenumfragen" (Reliability of Egocentric Networks Within Surveys). *ZUMA Nachrichten* 15:92-93.

Prosser, Robert, Jon Rasbash, and Harvey Goldstein. 1991. *ML3: Software for Three-Level Analysis*. London: Institute of Education, University of London.

Sixma, Herman and Wout C. Ultee. 1983. "Een Beroepsprestigeschaal voor Nederland in de Jaren Tachtig" (An Occupational Prestige Scale for the Netherlands in the 1980s). *Mens en Maatschappij* 58:360-82.

Sheatsley, P. B. 1951. "An Analysis of Interviewer Characteristics and Their Relationship to Performance, Part III." *International Journal of Opinion and Attitude Research* 5:191-220.

Snijders, Tom A. B. and Roel J. Bosker. 1994. "Modeled Variance in Two-Level Models." *Sociological Methods & Research* 22:342-63.

Starker, Joan E., David L. Morgan, and Stephen March. 1993. "Analyzing Change in Networks of Personal Relationships." Pp. 229-60 in *Advances in Personal Relationships*, vol. 4, *A Research Annual*, edited by D. Perlman and W. H. Jones. London: Kingsley.

Stember, Herbert and Herbert Hyman. 1949/1950. "Interviewer Effects in the Classification of Responses." *Public Opinion Quarterly* 13:669-82.

van der Zouwen, Johannes and Wil Dijkstra. 1988. "Types of Inadequate Interviewer Behaviour in Survey Interviews: Their Causes and Effects." *Bulletin de Méthodologie Sociologique* 18:5-20.

van Tilburg, Theo. 1995. "Delineation of the Social Network and Differences in Network Size." Pp. 83-96 in *Living Arrangements and Social Networks of Older Adults*, edited by C.P.M. Knipscheer, J. de Jong Gierveld, T. G. van Tilburg, and P. A. Dykstra. Amsterdam: VU University Press.

van Tilburg, Theo and Edith D. de Leeuw. 1991. "Stability of Scale Quality Under Different Data Collection Procedures: A Mode Comparison on the 'de Jong Gierveld Loneliness Scale.' " *International Journal of Public Opinion Research* 3:69-85.

Theo van Tilburg is an assistant professor in the Department of Social Research Methodology, Vrije Universiteit Amsterdam, the Netherlands. He is affiliated with the "Research on Methods of Data Collection," "Living Arrangements and Social Networks of Older Adults," and "Longitudinal Aging Study Amsterdam" research programs. His research interests include the development and evaluation of survey-measuring instruments, the effects of personal network characteristics and social support on well-being, loneliness and depression, and change in personal networks. His recent publications are "Network Analysis" in Encyclopedia of Gerontology. Age: Aging, and the Aged *(1996) and "Direct and Buffer Effects of Social Support and Personal Coping Resources in Individuals With Arthritis" in* Social Science and Medicine *(1997).*