CHAPTER 22

THE EFFECT OF INTERVIEWER AND RESPONDENT CHARACTERISTICS ON THE QUALITY OF SURVEY DATA: A MULTILEVEL MODEL

Joop J. Hox
University of Amsterdam

Edith D. de Leeuw
Vrije Universiteit

Ita G. G. Kreft
University of California at Los Angeles

The impact of interviewers on survey results has been a matter of concern for several decades. As early as 1929 Rice reported on the influence of interviewers' opinions and in 1955 one of the first overviews on interviewers as a source of error was written by Boyd and Westfall. Recently, Groves (1989) in his monograph on survey errors and survey costs gives considerable attention to the interviewer as a source of survey measurement error.

In the literature on interviewer effects three research strategies can be distinguished (cf. Singer and Presser, 1989b). The first research strategy investigates to what extent specific interviewer characteristics influence responses. The second approach concentrates on the estimation of the response variability attributable to the interviewers. The third investigates the technique of questioning by interaction coding (Chapters 19 and 21). In this chapter we concentrate on the first and to a lesser degree on the second research strategy. We will critically discuss previous findings and present an appropriate model for analyzing interviewer effects. The main emphasis is on the nature of the multilevel estimation model we use, which is different from the models used in previous studies on interviewer effects.

22.1 REVIEW OF PREVIOUS RESEARCH

Even in strictly controlled telephone surveys and in face to face interviews with trained interviewers, interviewer effects have been observed. When interviewer variation occurs, the responses of different respondents who are all interviewed by the same interviewer are more alike than those of respondents who are interviewed by different interviewers. Statistically, this is often expressed as the interviewer intraclass correlation (cf. Hansen, et al., 1961; Kish, 1962). In well-conducted face to face interviews the intraclass correlation typically clusters around 0.02; in controlled telephone surveys averages below 0.01 are reported (Kish, 1962; Groves and Magilavy, 1986; Tucker, 1983). A few studies (e.g., O'Muircheartaigh and Marckwardt, 1980) have reported larger effects. As the total effect of the interviewers on the overall variance of a survey statistic is a function of both the intraclass correlation and the interviewer workload, small intraclass correlations are still worrisome for researchers conducting large surveys in which each interviewer conducts many interviews (see Chapters 1 and 24).

Estimates of interviewer variance tell us how much variance in the response is attributable to the interviewer, but they do not reveal which interviewer characteristics are responsible. Three types of interviewer characteristics have been studied, either alone or together with respondent characteristics: 1) socio-demographic variables, 2) personality traits and social skills, and 3) interviewers' opinions and attitudes (cf. Collins, 1980; Hagenaars and Heinen, 1982). Although many interesting results have been reported, as yet no clear and systematic pattern has emerged. A summary of the main findings will be given below.

22.1.1 Interviewer Characteristics

The only consistent findings concerning the sociodemographic characteristics were for interviewer race, especially when race-connected questions were asked (Schuman and Converse, 1971; Hatchet and Schuman, 1975). No systematic effect of interviewers' sex on responses has been detected (Feldman, et al., 1951; Collins, 1980), but Groves and Fultz (1985) noted a tendency of respondents to express more optimism with male interviewers.

The age of the interviewer seems to have some effect on the responses, but findings are contradictory. Reviewing the literature, Sudman and Bradburn (1974) noted a tendency for older interviewers to cause less bias and response variance. Singer, et al. (1983) reported a higher response

rate with older interviewers and Berk and Bernstein (1988) found less item nonresponse when the interviewer was older, but no effect of interviewers' age on response validity. Collins (1980) did not find any clear effects of interviewers' age. In contrast with these findings, Hanson and Marks (1958) reported fewer "no answers" when the interviewer was younger and Freeman and Butler (1976) noted that older interviewers in combination with older respondents caused the highest interviewer variance. Mixed results have also been found concerning interviewer experience. Feldman, et al. (1951) reported that experience had a positive effect on the number of responses to open questions, Singer, et al. (1983) found that a higher response and better data quality were associated with more experience, but Bailar, et al., (1977) reported that experienced interviewers produced more "no answers" to income questions. Boyd and Westfall (1955), Collins (1980), and Berk and Bernstein (1988) found no apparent effect of experience on results obtained. Interviewers' education and socioeconomic status were found to have no discernible effects (Berk and Bernstein, 1988; Collins, 1980; Feldman, et al., 1951; Singer, et al., 1983).

Empirical evidence of which traits make for a successful interviewer is rather scarce. Feldman, et al. (1951) found no influence on response validity of interviewers' performance on a variety of psychological tests. Nevertheless, there is some evidence that effective interviewers (i.e., interviewers who obtained more valid answers) are less socially dependent, have more self-confidence, are more socially skilled and pay more attention to details (Steinkamp, 1966). Dijkstra, et al. (1979) found that socially skilled interviewers produced less interviewer variance. Social skills are associated with knowing and utilizing rules of behavior and communication (cf. Argyle, 1969) and should be clearly distinguished from a general social disposition or social orientation. The latter can have a negative effect. For instance, Hyman, et al. (1954) concluded that characteristics such as agreeableness or cooperativeness are somewhat negatively associated with interviewer performance. Dijkstra, et al. (1979) found that person-oriented (i.e., helping the respondent) interviewers produced more interviewer variance.

Little is known about the effect of interviewers' own opinions and attitudes. Feldman, et al. (1951) and Collins (1980) found no association between interviewers' and respondents' opinions. Hagenaars and Heinen (1982) presented the results of two Dutch studies on the influence of interviewers' opinions. In the first study on the quality of employment, small and hardly significant correlations were found. In the second study, substantial and significant correlations were found between the answers of interviewers and respondents to questions concerning women's liberation. The findings reported on the influence of interviewer expectations were more consistent. Both Sudman, et al. (1977a) and Singer and Kohnke-Aguirre (1979) reported small effects.

22.1.2 Methodological Comment

Several authors have criticized the existing studies on interviewer effects; see Hagenaars and Heinen (1982). A central criticism concerns the adequacy of the statistical models used. The structure of the data to be analyzed is hierarchical, since respondents are nested within interviewers. However, the analysis models which are typically used to relate interviewer and respondent characteristics do not take this into account (Dijkstra, 1983; Hagenaars and Heinen, 1982; Groves and Fultz, 1985). In this chapter we show what goes wrong in a traditional single-level analysis of hierarchical data and present a multilevel regression model which allows us to analyze correctly the joint effects of interviewer and respondent characteristics on data quality. Related statistical models for hierarchical data are discussed in this book by Hill (Chapter 23) and Pannekoek (Chapter 25).

Multilevel models can be applied to make statistically correct inferences about interviewers and respondent effects. We will present several examples of this use of multilevel models. Another application is to statistically control for interviewer effects, in a manner analogous to the way analysis of covariance is used to control for confounding variables. While the statistical analysis in both applications is similar, the substantive emphasis is different, and we will also present examples of this second application.

22.2 MULTILEVEL MODELS FOR INTERVIEWER EFFECTS

The design of interviewer variability studies combines both interviewer and respondent variables. Typically, both interviewer and respondent variables are combined in one ANOVA or multiple regression model. Since there are many more respondents than interviewers, this creates an obvious analysis problem. Traditionally, this problem has been solved by (dis)aggregation. One approach, which is rarely used, is to attach the mean of the respondent variables to the interviewers, and to analyze the resulting interviewer level data (for a good example see Schaeffer, 1980). This procedure will however, overestimate the standard errors, resulting in a severe loss of power (type II error). Another approach which is frequently used is to attach the interviewer variables to the respondent, implying as many interviewers as respondents. This procedure multiplies

tests that show "statistically significant" results far too often (type I error).

The hierarchical data structure also creates a second, more subtle analysis problem. Since the answers of respondents interviewed by the same interviewer tend to be more alike than those given by respondents who are interviewed by different interviewers, observations within the same interviewer are dependent or correlated. This dependency results in a loss of information, which introduces uncertainty in the parameter estimates of an ordinary ANOVA or regression model. Since these traditional techniques assume random sampling, they result in standard errors that are too small (type I error).

22.2.1 The Multilevel Model

An appropriate model to analyze this type of hierarchical data is a multilevel regression model, also known as hierarchical linear model (Mason, et al., 1984; Raudenbush and Bryk, 1986) or random component or variance component model (de Leeuw and Kreft, 1986; Longford, 1990). This is equivalent to performing a multiple regression analysis: one dependent variable is predicted by a number of independent variables. The difference here is that the independent variables can be defined on different levels of the data structure. In our study, we have two levels: the highest level is that of the interviewers, and the lowest level is that of the respondents. The lowest level has the most observations and is nested within the higher level.

In a traditional multiple regression model, we assume random sampling from one level only. Thus, if there were no interviewer effects we could predict a data quality indicator Y from the respondent variable R using the regression equation:

$$Y_i = \beta_0 + \beta_1 R_i + e_i \tag{22.1}$$

where the subscript i refers to respondents, β_0 is the ordinary intercept, and β_1 is the ordinary regression coefficient (slope). The errors e_i are assumed to be independent with a variance $\text{var}(e_i) = \sigma_e^2$.

If there are interviewer effects, equation (22.1) should be changed to:

$$Y_{ij} = \beta_{0j} + \beta_{1j} R_{ij} + e_{ij} \tag{22.2}$$

with the added subscript j referring to the interviewers. In the ordinary regression equation (22.1) there is a single intercept and a single slope for all respondents. In the multilevel regression equation (22.2) it is assumed that the intercepts β_{0j} and the slopes β_{1j} vary between interviewers. This means that the regression parameters β_0 and β_1 are thought of as having a

We may try to account for this variation by introducing explanatory variables at the interviewer level. If we introduce the interviewer variable I we can write the following two regression equations for the intercept β_0 and the slope β_1:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} I_j + U_{0j} \tag{22.3}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} I_j + U_{1j} \tag{22.4}$$

where U_{0j} and U_{1j} are the error terms with variance σ_0^2 and σ_1^2.

If we substitute equations (22.3) and (22.4) in equation (22.2) and rearrange terms we obtain the regression equation:

$$\begin{aligned} Y_{ij} = \gamma_{00} + \gamma_{10} R_{ij} + \gamma_{01} I_j + \gamma_{11} I_j R_{ij} + \ldots \\ + U_{0j} + U_{1j} R_{ij} + e_{ij} \end{aligned} \tag{22.5}$$

Thus, we end up with a multilevel regression equation (22.5) which incorporates respondent, interviewer, and interaction variables and a very complicated error term $[U_{0j} + U_{1j} R_{ij} + e_{ij}]$.

Asymptotic standard errors can be calculated for the regression coefficients γ, which allow us to conduct significance tests similar to those in ordinary multiple regression analysis. There are also asymptotic standard errors for the various variance components, which allow us to test whether the variation of the regression coefficients (intercept and slopes) between the interviewers is significant.

A multilevel analysis with no explanatory (respondent and interviewer) variables at all, the so called intercept-only model, produces σ_e^2 and σ_0^2, which can be used to estimate the familiar intraclass correlation for the interviewer effect, rho, as $\sigma_0^2/(\sigma_0^2 + \sigma_e^2)$. The intercept-only model is a useful starting point for an analysis, because it indicates how much variance there is at the interviewer level. If this variance is small, the attempts to explain it should not be expected to produce very dramatic results. Note that the error variances σ_0^2 and σ_e^2 are conditional upon the explanatory variables already included in the model. If explanatory variables are included in the model, the formula $\sigma_0^2/(\sigma_0^2 + \sigma_e^2)$ no longer estimates the intraclass correlation. Nevertheless, it is still useful to compare the variance components of the regression coefficients in different models to see how well the explanatory variables explain the between-interviewer variance of these coefficients.[1]

In the general multilevel model, the regression coefficients for the respondent level variables are random, which means that they are allowed

[1] If there are random slopes the situation is more complicated (cf. Aitkin and Longford, 1986). In practice this means one should not directly compare the intercept variance of models with different random slopes.

to vary across interviewers. In other words, there are interactions between that particular respondent level variable and one or more interviewer level variables. Subsequently, we may define interaction terms which specify which of the available interviewer variables are hypothesized to interact with the respondent variable in question.

The intercept is always assumed to be random. However, we may designate some or all lower level regression coefficients as fixed, instead of random. Thus, if a statistical test shows that the between-interviewer variance of a specific regression coefficient is not significant, we may decide to model that coefficient as fixed. This statistical decision carries a substantive meaning: if we define a regression coefficient as fixed, we assume that the effect of that respondent variable on the dependent variable is identical for all interviewers. In other words: we assume no interactions between that specific respondent variable and interviewer variables. If we define all regression coefficients as fixed, this results in a model which closely resembles the ordinary regression model. The only difference is the random intercept, which reflects the between-interviewer variance.

22.2.2 Model Search

Including all interviewer and respondent variables, and adding all possible interviewer-method and interviewer-respondent interactions, would lead to statistically unwieldy and unstable multilevel models. Consequently, some heuristic is needed for the selection of variables. Longford (1990) advises to start the model construction by determining the appropriate fixed effects, followed by a procedure to test for random effects. This leads to a search procedure, in which lower level (i.e., respondent level) variables are selected before higher (interviewer) level variables and random effects are selected.

Following Longford (1990), we select the variables and build a multilevel model in several steps. First of all, for each dependent variable, a smaller set of promising respondent variables, including interactions of interviewing method with interviewer characteristics, is selected by ordinary least squares (OLS) regression (using $p = 0.10$ as a criterion).

In the multilevel analysis, we start with a fixed effects model which includes all these independent variables. A backward selection procedure is used to arrive at a model which includes only significant fixed effects. In the next step, the regression coefficients for the respondent variables are examined for possible random effects. Finally, if a random effect is found, we try to model it with the available interviewer variables.

The respondent-level variables in this study are: interviewing method, age, sex, education, evaluation of respondent by interviewer (e.g., honest, cooperative), evaluation of interview by respondent (e.g., pleasant, threatening), the attitudinal distance between interviewer and respondent, and whether the respondent had ever refused an interview before. From all the possible interactions, two were designated on theoretical grounds as potentially interesting. The first is the (Euclidean) distance between responses of the interviewer and the respondent on four multi-item scales measuring (un)happiness, self-esteem and loneliness: this variable is the attitudinal distance between interviewer and respondent. The second is the set of interactions between interviewer characteristics and method: these reflect the hypothesis that the effect of the interviewer characteristics might be different in different modes of interview. (cf. Sykes and Collins, 1988). From this set, a smaller number of variables was selected using OLS (see above).

All potentially relevant interviewer-level variables are always included in the first stage of the multilevel model exploration (i.e., age, experience, previous training, preference for telephone vs. face to face interviews, extroversion, conscientiousness, social anxiety, ability to initiate social contacts, and ability to efficiently terminate undesirable social situations; for a description see Section 22.3). Also, four respondent-level variables are always included (age, sex, education, and interviewing method used).

22.3 DATA COLLECTION

In the autumn of 1989 a controlled field experiment was conducted in which the effects of interviewer and respondent characteristics were investigated.

The subject of the questionnaire was well-being. The questionnaire included questions on general satisfaction, on loneliness, and on (un)happiness. Furthermore a large number of background questions including questions on income and work situation were asked.

Twenty interviewers were selected, varying in previous interviewer experience and training. Important selection criteria were clarity of voice over the telephone, legible handwriting and higher education. Prior to the data collection all interviewers received a standardized interviewer training based on the Survey Research Center manual (1976). Additional training was given in telephone interviewing techniques. All interviewers conducted both face to face and telephone interviews. Ten randomly assigned interviewers started with telephone interviews and then conducted face to face interviews; the other ten started with face to face interviews.

As part of the selection procedure all interviewers completed a biographical question sheet and a specially designed self-administered form of the interview schedule. Before the interviewer training started we asked each interviewer to complete a Dutch personality inventory, which measures extroversion, friendly disposition, conscientiousness, social anxiety, and ability to efficiently terminate undesirable social situations (the reliability of the personality scales ranges from 0.69 to 0.85; Akkerman, 1977; Hox, 1978).

The respondents were a random stratified sample from the general population in the Netherlands. For the complete sample (respondents and nonrespondents) detailed background information was available based on the Dutch zip code system (Geo-marktprofiel). The Dutch zip codes form an extremely fine grid; as a consequence aggregated information was available on, for instance, socioeconomic status (SES), income, type of household, and type of community for clusters of, on average, 15 households each.

The sample was randomly split in two. One half was interviewed by telephone; the other half, face to face. In the telephone situation the respondents were randomly assigned to interviewers. In the face to face situation respondents were randomly assigned to interviewers within four geographical regions in order to avoid excessive travelling. All sample units received an advance letter. The interviewers used a standardized script in asking for respondent cooperation. In all cases (telephone and face to face situations) the request for cooperation was made by telephone. At least seven callbacks were made. No attempt was made to convert definite refusals; refusers were not called back by special interviewers.

22.4 RESULTS

In the next sections we employ a multilevel model of respondent and interviewer characteristics as described in Section 22.2, with special attention to random (i.e., between-interviewer) effects. We try to answer the following questions:

1. How do interviewers affect the nonresponse?
2. How do interviewers affect the measurement of variables?
3. How do interviewers affect the respondents' feelings about the interview?
4. How do interviewers affect substantive models?

The results are discussed in separate sections. A summary table is provided in Appendix A. To help interpret the results, all regression coefficients are reported as standardized coefficients. For comparison with previous research we also give the intraclass correlation rho (cf. Groves, 1989, chapter 8).

22.4.1 Nonresponse

In the face to face situation 243 interviews were completed and 266 were done in the telephone situation. This resulted in a response rate of 51% for the face to face situation and of 66% for the telephone situation. The difference in response rates is almost entirely due to a difference in explicit refusals: in the face to face situation 191 respondents (40%) refused cooperation, while in the telephone condition only 114 respondents (28%) refused.

Nonresponse, especially the large nonresponse of the face to face interview, is a potential source of error. Fortunately, external (zip-code based) information on both respondents and nonrespondents was available. We investigated the possibility of selective nonresponse using this auxiliary zip-code information. No differences were found between respondents and nonrespondents. Also no interaction effects of respondent characteristics and mode of data collection were observed: although the response rates differ, characteristics of respondents and nonrespondents do not differ across the modes.

When the different types of nonresponse are investigated separately, a very interesting pattern is discovered. Respondents and refusers do not differ from each other, but they do differ significantly ($p < 0.05$) from the unreachables (noncontact, ill, senile, language problem). In general the unreachables were less affluent, did not own a house and were more often found in urban areas.

Cooperation Rate

Interviewers can differ in the proportions of refused and completed interviews they had from potential respondents. The cooperation rate was estimated to assess how successful individual interviewers were in persuading those able to do the interview to comply with the request (Groves, 1989). The cooperation rate is the ratio of completed interviews to all contacted cases capable of being interviewed.

The between-interviewer variance in cooperation rate was small, resulting in an intraclass correlation for the interviewers of rho = 0.02. In the multilevel model, the cooperation rate did not depend on any of the auxiliary zip-code variables. As can be deduced from the overall response rates above, the specific request (for a telephone or a face to face interview) had a large influence (beta = 0.58; $p < 0.01$). There were no significant contributions of interviewer variables. Also, the variance component for interviewers was not significant, which indicates that interviewers did in fact not differ in their ability to persuade respondents to participate in the survey.

22.4.2 Interviewer Influence on the Measurement of Variables

Three different aspects were investigated: completeness of answer, response bias caused by acquiescence, and the psychometric reliability of a measurement. Indicators for completeness were item nonresponse and the number of different statements to open questions.

Completeness: Item Nonresponse

Missing data can pose serious problems in social research. Interviewers are therefore instructed to reinforce adequate respondent behavior and probe for an answer whenever necessary. To investigate interviewer effects on missing data, we constructed a global indicator measuring the proportion of item nonresponse over all questions.

Overall, the incidence of missing data was low (1.3 percent). The intraclass correlation for interviewers was 0.01. As expected from the research literature, older respondents had more missing data (beta = 0.26, $p < 0.01$). A respondent who had missing data was negatively evaluated by the interviewer (beta = −0.32, $p < 0.01$). The data collection method used did not influence the amount of missing data (beta = 0.07, $p = 0.13$).

Two interviewer variables were significant: a preference for face to face interviews by the interviewer (beta = 0.13, $p = 0.03$) and the extroversion of the interviewer (beta = −0.13, $p = 0.02$). Introverted interviewers and interviewers who, when asked, expressed a preference for the face to face method produced interviews with more missing data in both telephone and face to face interviews.

The regression coefficients for the age of the respondents and the respondent evaluation by the interviewers showed significant random components, indicating an interaction with certain interviewer characteristics. However, an exploratory analysis employing the available interviewer variables did not reveal any significant interactions.

Completeness: Open Questions

Three standard open questions were used. In these questions the respondents were asked to elaborate their previous answers. For each respondent, the mean number of answers to these three open questions was computed. For completeness, the intraclass correlation for the interviewers was substantial: rho $= 0.13$. The number of answers depended on two respondent variables: age (beta $= 0.15$, $p < 0.01$) and enjoyment of the interview (beta $= 0.14$, $p = 0.02$). Older respondents and respondents who reported that they had enjoyed the interview produced more answers to the open questions.

Even after introducing these two respondent variables, the between-interviewer variance component remains significant ($p < 0.01$). None of the available interviewer variables we analyzed made a significant contribution to this interviewer effect. The regression coefficient for interview enjoyment by the respondent had a significant random component ($p < 0.01$), which means that this respondent variable interacted with some interviewer characteristics. However, an exploratory analysis employing the available interviewer variables did not reveal any significant interactions. We should note that in this example all interviewer variables available for analysis concerned socio-demographic interviewer characteristics and personality traits. Process variables, such as the degree of probing, were not available for further analysis.

Response Bias: Acquiescence

In the Dutch version of the "Affect Balance Scale" all positively formulated items are exactly balanced by corresponding negatively formulated items (this is not true for the American version, see Bradburn, 1969; Hox, 1986). In a balanced scale acquiescence or yea-saying can be estimated by counting the number of "agree"-responses to all items.

The intraclass correlation for the interviewers on this scale was fairly substantial (0.06). Three respondent variables were associated with acquiescence: age (beta $= -0.24$, $p < 0.01$), sex (beta $= 0.19$, $p < 0.01$), and refusal of earlier surveys (beta $= 0.14$, $p < 0.01$). Younger respondents, women, and respondents who had refused other surveys but were willing to participate in this one all had a greater tendency to give an affirmative response to the questions. The regression slopes for these variables did not differ between interviewers, meaning that there were no interactions with interviewer variables.

After including the respondent variables in the model, the between-interviewer variance component still remains significant ($p = 0.03$). However, none of the interviewer variables which were available for analysis made a significant contribution to this interviewer effect.

Psychometric Reliability

To gauge the psychometric reliability, we computed for each respondent the inter-item variance for four multi-item scales (Positive affect or

happiness, Negative affect, Loneliness, and Self-esteem). The mean inter-item variance is an overall indicator of the scalability of the individual responses; a high variance indicates an unstable response pattern.

The intraclass correlation for the interviewers was estimated to be zero. Nevertheless, a multilevel model revealed significant interviewer effects, including random regression coefficients for the variables Attitudinal Distance and Evaluation of Respondent by Interviewer.

A respondent with variable responses or in other words, a respondent with an unstable response pattern, was negatively evaluated by the interviewer (beta $= -0.19$, $p < 0.01$). Furthermore, interviewer extroversion is negatively related to response variability (beta $= -0.09$, $p = 0.03$), indicating that interviewers who were extroverted produced less variability and a more stable response pattern. Although the mean regression coefficient for attitudinal distance between the interviewer and the respondent was not significant (beta $= 0.10$, $p = 0.19$), the regression slopes for this variable differed significantly between interviewers, indicating an interaction with certain interviewer characteristics. An exploratory analysis employing the available interviewer variables did not reveal any significant interactions.

Since the inter-item variance was computed for four distinct multi-item scales, it is possible to analyze this data with a three-level model, with "interviewers," "respondents," and "scales" as the three levels. This enabled us to analyze the between-respondent variance. The intraclass correlation for respondents was estimated as 0.15 ($p < 0.01$), which indicated that respondents do in fact differ systematically in the degree of response variability. The intraclass correlation for interviewers was again estimated as zero. The three-level analysis selected the same explanatory variables as the two-level analysis. The between-respondent variance component remained significant ($p < 0.01$).

22.4.3 Interview Evaluation

The interview was evaluated by the respondents in two different ways. First, the respondents were asked to assess the interview on a five-item questionnaire threat scale. Second, they were asked whether they enjoyed the interview.

Questionnaire Threat

Questionnaire threat was measured by a five-item scale which asked for the respondent's assessment of the interview as a whole. The intraclass correlation was small (0.01). Perceived questionnaire threat was related to the attitudinal distance between interviewer and respondent (beta $= 0.10$, $p = 0.05$) and to the interviewer's conscientiousness (beta = -0.14, $p < 0.01$). Respondents assessed the questions as more threatening, and as more personal and intruding on their privacy, when there was a greater attitudinal difference between interviewer and respondent, and when interviewers were less conscientious according to their answers on a personality inventory.

The regression slope for attitudinal distance did not differ between the interviewers. After including these variables, the interviewer variance was not significant.

Interview Enjoyment

The intraclass correlation for interview enjoyment was 0.05. Respondents who were interviewed face to face enjoyed the interview more than respondents in the telephone situation (beta = -0.26, $p < 0.01$). Furthermore, the following respondent variables were related to enjoyment: age (beta = 0.19, $p < 0.01$), education (beta = -0.12, $p = 0.00$), and respondent evaluation by the interviewer (beta = 0.29, $p < 0.01$). Older respondents and respondents who were positively evaluated by the interviewer reported more enjoyment. More highly educated respondents, however, reported less interview enjoyment. The regression slopes did not differ between the interviewers, indicating that there were no interactions of these respondent variables with interviewer variables.

After including the respondent variables, the between-interviewer variance component remains significant ($p < 0.01$). None of the interviewer variables we analyzed made a significant contribution to this interviewer effect.

22.4.4 Controlling Interviewer Influence on a Substantive Model

While the intraclass correlations for most methodological quality indicators are small, their combined effect on the quality of the data as a whole may still be substantial, especially if the interviewer workload is high (cf. Groves, 1989). In our case, the mean number of interviews per interviewer was moderate (25). To illustrate both the impact of interviewer variance on a substantive model and ways to control for this effect statistically, we looked more closely at two models. The first was a loneliness model, and the second was a model for well-being.

The interview survey can be viewed as a procedure for cluster sampling with interviewers defining clusters of respondents. To control for interviewer effects one can then employ an analysis model which corrects for the effects of this "sampling" design (cf. Chapter 31; Lee, et al., 1989). However, if researchers are interested in multiple regression

models, the most elegant approach is to use multilevel regression models to correct for the hierarchical nature of the data. The easiest and simplest approach is then to model only fixed regression coefficients: this is similar to an ordinary multiple regression analysis, but adjusts for the between-interviewer variance. This approach is referred to as the Hierarchical Linear Model (HLM). A more general and theoretically more interesting approach is to allow the regression coefficients for the respondents to be random at the interviewer level, and to introduce explanatory interviewer variables in the model. We will refer to this as the Random Coefficient Model (RCM).

While the HLM only gives us an estimate of a regression coefficient and its corresponding standard error, the RCM also produces an estimate of the variation due to interviewers. An intuitive way to interpret this variation is to view it as an indication of how much a researcher could influence the substantive results of a study by a judicious or malicious choice of the interviewers employed (see Wiggins, et al., 1989 for a thorough discussion). Furthermore, the RCM allows us to model interviewer effects. This, too, is of interest when analyzing a substantive model, because it effectively controls for these variables.

Loneliness Model

For our comparison, we used a loneliness model derived from De Jong-Gierveld (1987). In our model, loneliness is (negatively) determined by presence of a partner, extension of the social network, satisfaction with the social network, and self-esteem. We included the respondent's sex and age in the analysis. These variables were included in the path model used by De Jong-Gierveld, but did not directly affect loneliness. In addition to these variables, we included auxiliary respondent and interviewer variables in the random coefficient model. Two of these variables were significant: attitudinal distance between interviewer and respondent, and respondent evaluation by interviewer (see Table 22.1). The coefficients in the analyses are virtually the same. The amount of variance explained hardly differs between the three models (35%, 35%, and 37%).

In the Random Coefficient Model, none of the regression coefficients has a significant random component, which means that the regression coefficients may be considered equal for all interviewers.

Well-being Model

For our comparison, we used a model for well-being derived from Burt, et al. (1978, 1979). In this model, well-being (general satisfaction) is determined by positive affect, negative affect, and satisfaction with specific life domains (housing, health, income, and social network).

Table 22.1. Loneliness Model Regression Coefficients

	OLS		HLM		RCM	
Variable	beta	p	beta	p	beta	p
Age	0.03	0.52	0.03	0.65	0.00	0.98
Sex	0.00	0.99	0.00	0.99	0.00	0.95
Ext. soc. netw.	−0.13	0.00	−0.13	0.00	−0.11	0.00
Sat. soc. netw.	−0.40	0.00	−0.40	0.00	−0.38	0.00
Self-esteem	−0.26	0.00	−0.25	0.00	−0.22	0.00
No partner	0.16	0.00	0.16	0.00	0.17	0.00
Att. distance[a]					0.12	0.01
Respondent eval. by interviewer					−0.09	0.04

[a] In this analysis, attitudinal distance was computed with exclusion of the loneliness score.
OLS, Ordinary Least Squares Model; HLM, Hierarchical Linear Model; RCM, Random Coefficient Model.
Source: De Leeuw, Social Cultural Sciences Foundation (NWO) grant no. 500278008.

Again, in addition to these variables, we included auxiliary respondent and interviewer variables in the random coefficient model. Five of these variables were significant: attitudinal distance, respondent evaluation by interviewer, interviewer extroversion, interviewer social anxiety, and interviewing method used (see Table 22.2). The coefficients in the Ordinary Least Squares model and the Hierarchical Linear Model analyses are virtually the same. The Random Coefficient Model yields coefficients which are generally somewhat lower. No difference was found in the amount of explained variance between the OLS and HLM models (both were estimated as 0.26); the RCM model explains a much larger amount of variance (0.34).

The coefficient for "satisfaction with income" is larger (0.19), and random over interviewers. This means that the strength of this particular effect varies with different interviewers. To decide how much, we have to look at the variance component for the regression coefficient. In this case the variance is 0.02, which corresponds to a standard deviation of 0.14. Assuming a normal distribution of the regression coefficient, this means that a large majority of all interviewers (91%) would produce a positive relationship between "satisfaction with income" and general satisfaction. In sum, the size (0.19) and the significance ($p = 0.00$) of the random regression coefficient tells us that the average regression coefficient across interviewers is positive and statistically significant. The size of the standard deviation (the square root of the variance component) tells us that selecting different interviewers is unlikely to change this picture.

Table 22.2. Well-being Model Regression Coefficients

	OLS		HLM		RCM	
Variable	beta	p	beta	p	beta	p
Neg. affect	−0.25	0.00	−0.24	0.00	−0.21	0.00
Pos. affect	0.18	0.00	0.18	0.00	0.12	0.01
Sat. house	0.05	0.23	0.05	0.22	0.04	0.32
Sat. health	0.13	0.00	0.13	0.00	0.11	0.01
Sat. income	0.14	0.00	0.14	0.00	**0.19**	0.00
Sat. soc. netw.	0.19	0.00	0.20	0.00	0.18	0.00
Att. distance					−0.12	0.01
Respondent eval. by interviewer					0.13	0.00
Extroversion					0.12	0.01
Soc. anxiety					−0.13	0.00
Method					0.12	0.00

Bold and underscored: random coefficient.
OLS, Ordinary Least Squares Model; HLM, Hierarchical Linear Model; RCM, Random Coefficient Model.
Source: De Leeuw, Social Cultural Sciences Foundation (NWO) grant no. 500278008.

22.5 SUMMARY AND DISCUSSION

22.5.1 The Multilevel Model

Interview survey data have a hierarchical structure, and multilevel analysis methods provide efficient and flexible techniques to analyze such data. The model we present is essentially a (multilevel) multiple regression model, which has been described in a number of publications (e.g. Mason, Wong, and Entwhisle, 1984; Raudenbush and Bryk, 1986; Goldstein, 1987), and for which several computer packages are available (for a critical review see Kreft, et al., 1990; three PC-based programs are listed in Appendix B to this chapter.) The models described in this volume by Hill (Chapter 23) and Pannekoek (Chapter 25) are very similar, but some of the statistical details differ. Hill's model does not assume a normal error structure, as our model does, but allows for testing and fitting more complicated error distributions, while Pannekoek's model provides a direct analysis of the interviewer intraclass correlation. These and other developments such as extensions to binomial data (Longford, 1988; Goldstein, 1991) and covariance structure models (e.g., Muthén, 1989) are important. Nevertheless, the flexibility of the multiple regression model (cf. Cohen and Cohen, 1983) and the availability of easy-to-use software make the multilevel regression model a powerful tool for the analysis of hierarchical survey data.

It is often stressed that in interviewer effect studies, respondents should be assigned to interviewers at random (cf. Hagenaars and Heinen, 1982; Groves, 1989). In large scale face to face surveys this is expensive and difficult to organize. It is difficult to use such studies for methodological research into interviewer effects, because without interpenetrating designs the interviewer and respondent explanatory variables are correlated and the interviewer intraclass correlation no longer estimates interviewer effects only. Multilevel analysis methods offer some remedies to overcome this difficulty. If the relevant respondent variables are known, they can be put into the regression model to equate the respondent "input" between interviewers by statistical means. If, after statistical control of the respondent variables, interviewer variables are still statistically significant, we may conclude that this reflects real differences between interviewers. Thus, the appropriate analysis procedure is to test whether interviewer variables explain significant variance in addition to the relevant respondent variables. Again, this is similar to analysis of covariance, with the interviewers as the independent variable and the respondent variables as the covariates to be controlled for, but the assumptions of the multilevel model are much more realistic than those of analysis of covariance. Additionally, even researchers who are not interested in interviewer effects may find it useful to include interviewer variables in the analysis, because it shows how large their potentially disturbing effect is, and offers some means of control (cf. Goldstein, 1990).

22.5.2 Interviewer Effects

Four questions were asked about the way interviewers can influence survey results.

1. The first question concerned the nonresponse. Two remarkable results were obtained: telephone interviews resulted in a higher response rate than face to face interviews, and interviewers did not differ in their ability to persuade contacts to participate in the interview. The difference in response rates between face to face and telephone interviews probably results from the specific procedures used to persuade respondents. The telephone interview involves two steps: after asking for the designated respondent in the household, the interviewer asks her/him to answer some

questions. In the face to face interview a three-step procedure is used: the interviewer telephones and after asking for the designated respondent, tries to make an appointment for an interview at the respondent's home.

As stated in Dillman's application of social exchange theory (Dillman, 1978), the perceived costs of responding may be higher in the face to face interview, while in this case the cost of refusing is the same for the telephone and the face to face interview. In a more traditional approach, interviewers are sent directly to the respondent's home, asking for cooperation on the doorstep. This increases the social cost of refusing cooperation, resulting in a higher response rate. However, for the researcher, this also increases the financial costs of callbacks.

The absence of interviewer effects on cooperation rate may be caused by two factors: the interviewers were well-trained and used a detailed script to persuade contacts to participate. This is likely to result in little variation between interviewers at the beginning of the interview, when the decision to participate is made.

2. The second question concerned the measurement of variables. Overall, the literature reports low intraclass correlations and mixed results concerning the influence of most interviewer characteristics. We find very few interviewer variables that explain significant interviewer variance (see the Summary Table in Appendix A). This difference could very well be the result of biased significance tests in earlier research, caused by disaggregating interviewer variables.

In our present study, in two cases a relatively high intraclass correlation was observed, in combination with a significant residual interviewer variance. For acquiescence, no significant interviewer variables were found. For the number of answers to open questions, again no significant interviewer variables were found. This is more surprising, since the intraclass correlation of this variable is 0.13, the highest correlation coefficient in our results. We suspect that in this case process variables may be more important than socio-demographic interviewer characteristics or personality variables. The number of answers on open questions may be more dependent on differences in the degree of paralinguistic behavior, such as saying "mhmm-hmm"after an answer. Also interviewers may differ in the manner and perseverance of additional probing after the prescribed standard probe on open questions.

Several models showed significant random coefficients, which indicate interactions between respondent and interviewer characteristics. However, further analysis using the variables available in this study did not explain this variability. Again, we suspect that process variables may be more important than the substantial number of interviewer characteristics available in this study.

3. The third question concerned respondents' feelings about the interview. Concerning experienced questionnaire threat, a small interviewer effect is observed which is explained by interviewer conscientiousness and attitudinal distance. Conscientious interviewers and interviewers who have a small attitudinal distance to the respondents lead to less questionnaire threat. Enjoyment of the interview is influenced by interviewers (rho = 0.05), but none of the interviewer variables in this study could explain this effect. The interviewer's evaluation of the interview did concur with the respondent's judgment on what was an enjoyable interview (beta = 0.29). Furthermore, older respondents, less educated respondents and respondents who were interviewed face to face stated that they enjoyed the interview more.

4. The fourth question concerned the way interviewers affect substantive models. For the loneliness model the interviewer effects were negligible. For the well-being model small effects were found, but it is unlikely that they would have substantially influenced the model interpretation. That this is not always the case is shown by Wiggins, et al., 1989 who in an earlier analysis of a survey on physical handicaps found a large random effect. In their case, a clever choice of the "right" interviewers could even have determined the sign of the regression coefficient for the respondents' age.

22.5.3 Implications

Our results suggest some implications for research utilizing face to face or telephone interviews. First of all, while the literature suggests small method effects (De Leeuw and van der Zouwen, 1988), in our case there was virtually no direct method effect. Furthermore, with the exception of the number of answers to open questions, the interviewer effects were generally small, and did not differ between the two methods. Of course, our data come from a relatively small-scale methodological study employing a thoroughly pilot-tested questionnaire, and interviewers who were well trained in both face to face and telephone interviewing, closely supervised, and provided with scripts for difficult situations. In large-scale surveys the field conditions may be less optimal, and differences between methods or interviewers may be larger. Furthermore, since multilevel methods control type I error much better than ordinary single level analyses of multilevel data, they are expected to produce fewer significant findings. This is not a disadvantage; many of the "significant" effects found in an ordinary single level analysis of multilevel data are bound to be spurious.

At any rate, our results provide few grounds for further selection of interviewers on socio-demographic or psychological characteristics (cf. Collins, 1980). Kish (1962) suggests that interviewer effects are likely to be the result of many small differences in interviewer characteristics, respondent populations, study designs and resources. Our results point in the same direction. The implication is that controlling interviewer effects by suppressing them completely or keeping them constant will be difficult. One radical solution is to get rid of the interviewer completely, either by using mail surveys (Dillman, 1978), or by using completely computerized data collection procedures (Saris, 1989), with the risk of facing an explosion in respondent errors. When there are good reasons to employ interviewers, one feasible solution is to control for interviewer effects statistically.

ACKNOWLEDGMENTS

This study was conducted while the first two authors were at the UCLA as Fulbright and visiting scholars. They are indebted to the Department of Psychology and the Program for Social Statistics for their hospitality. The data collection was made possible by grant no. 500278008 from the Social Cultural Sciences Foundation, which is subsidized by the Netherlands Organization for the Advancement of Scientific Research (NWO).

Special thanks are due to Nick Longford for his many constructive suggestions and to Gerard Kurvers and GEO-MARKTPROFIEL for their zip-code information.

PENDIX A

mary Table of Standardized Regression Coefficients

pendent variable	Cooperation rate	Item nonresponse	Open questions	Acquiescence	Psychometric reliability	Questionnaire threat	Enjoyment of interview
	Dependent Variable						
rviewer Level							
ge							
xperience							
evious training							
efer face-to-face		0.13					
xtroversion		−0.13			−0.09		
onscientious						−0.14	
iendly							
nxiety							
erminate contacts							
ondent level							
Iethod	0.58						−0.26
ge	na	**0.26**	0.15	−0.24			0.19
ex	na			0.19			
ducation	na						−0.12
valuation of respondent by interviewer	na	**−0.32**			−0.19		0.29
njoyment by respondent	na		**0.14**				na
ttitudinal distance	na				**0.10**ns	0.10	
Iaving refused earlier surveys	na			0.14			
ractions							
ethod × A/H							
aclass Correlation	0.02	0.01	0.13	0.06	0.00	0.01	0.05
iance between interviewers	ns	ns	$p=0.00$	$p=0.03$	ns	ns	$p=0.00$

l and underscored = random coefficient; ns = not significant at 0.05 level; na = not applicable.
rce: De Leeuw, Social Cultural Sciences Foundation (NWO) grant no. 500278008.

APPENDIX B

Program information on the multilevel model

HLM (Hierarchical Linear Modeling):
Scientific Software, Inc.
1369 Neitzel Rd.
Mooresville, IN 46185-9312, USA

ML3 (three-level analysis)
Multilevel Models Project
Institute of Education, University of London
20 Bedford Way
London, WC1H 0AL, United Kingdom

VARCL (Variance Component Analysis)
N.T. Longford
21T Educational Testing Service
Rosedale Rd.
Princeton, NJ 08541, USA

All programs require at least a PC/AT with 640k memory and hard disk; numeric coprocessor is strongly recommended.