Quality & Quantity **28**: 329–344, 1994.

A comparison of nonresponse in mail, telephone, and face-to-face surveys
Applying multilevel modeling to meta-analysis

JOOP J. HOX[1] & EDITH D. DE LEEUW[2]
[1]*Department of Education, University of Amsterdam, IJsbaanpad 9, 1076 CV Amsterdam, The Netherlands;* [2]*Department of Social Research Methodology, Vrije Universiteit, Amsterdam, Koningslaan 22–24, 1075 AD Amsterdam, The Netherlands.*

Abstract. This article reports a meta-analysis of 45 studies that explicitly compare the response obtained using a mail, telephone or face-to-face survey. The data analysis uses a generalized hierarchical linear model. Sampling procedure (e.g., local convenience sample, random general sample), saliency of topic, and research organization (university, government versus market research) had an effect on the response. On the average, the face-to-face condition achieved the highest completion rate (70.3%), the telephone survey the next highest (67.2%), and the mail survey the lowest (61.3%). There is a significant interaction with the year of publication: The response to face-to-face and telephone surveys is going down in the period covered by this analysis (1947 to 1992), but the response to mail surveys is going up slightly. We attribute this to the large amount of research on nonresponse problems with mail surveys, and recommend more research and development in this direction for face-to-face and telephone methods.

1. Introduction

One of the most important problems in survey research is nonresponse: the failure to obtain measurements from all units in the sample. Nonresponse error is to be distinguished from noncoverage error, which is the absence of certain population elements in the sample because these elements are not present in the sampling frame (e.g., 'address unknown' or 'no telephone'). Nonresponse error occurs when the population elements are present in the sampling frame, but for some reason (e.g., refusal) no data are obtained (Groves & Lyberg, 1988).

The response to a survey is affected by many factors, other than the type of survey, such as the saliency of the topic, the research organization (e.g., government, market research), the type of sample, and the number of reminders in a mail survey or the number of visits in a face-to-face survey (cf. Goyder, 1987). The data collection method is also important. In general,

Correspondence to be addressed to: J. J. Hox, Department of Education, University of Amsterdam, IJsbaanpad 9, 1076 CV Amsterdam. Tel: X-31-20-5703535, fax: X-31-20-5703500, E-mail: A716HOX @ HASARA11.

face-to-face surveys result in a higher response than comparable telephone or mail surveys. For instance, Groves and Kahn (1979, p. 76) report a response of 74% for a face-to-face survey and 70% for a comparable telephone survey. Steeh (1981) reports that a transition from a face-to-face to a telephone survey resulted in a drop in response from 73% to 66%. Finally, Goyder (1987) compares the responses to 385 mail surveys, 112 face-to-face surveys, and 53 telephone surveys. He finds a mean response rate of 67.3% for face-to-face interviews, 60.2% for telephone surveys, and 58.4% for mail surveys.

In most western societies the response to face-to-face surveys is falling. Steeh (1981) reports a lowering response to two ongoing large scale face-to-face surveys in the U.S.A., Goyder (1987) observes that the response to face-to-face surveys is dropping both in the U.S.A. and Canada, and Sugiyama (1992) reports similar experiences in Japan. In the Netherlands, falling response rates have been reported by Bethlehem and Kersten (1981).

Research results can be biased if the nonresponse is nonrandom, and if it is in some way correlated with the variables measured in the survey. Since the process leading to nonresponse is usually unknown, it is often optimistically assumed that when the response is high, there is no serious nonresponse bias. Thus, a high response rate is viewed not only as desirable, but also as an important criterion by which the quality of a survey is judged.

Many studies have been performed to identify methods that raise the response rate. The large amount of research in this area has resulted in several literature reviews. Some of these reviews (e.g., Heberlein & Baumgartner, 1978; Goyder, 1987; Yu & Cooper, 1983; Fox, Crask & Kim, 1988) use meta-analysis (Glass, McGaw & Smith, 1981; Hedges & Olkin, 1985) to summarize the empirical results.

The main findings from these reviews are the following. Multiple contacts and rewards for cooperation have a positive effect on the response, as has the saliency of the survey's topic. The population sampled and the research organization both affect the response: geographically close and nonrandom samples result in higher response rates than random samples from a wide geographical area, and government agencies or universities generally obtain higher response rates than market research agencies.

This article presents a meta-analysis of studies that have explicitly compared the response to face-to-face, telephone, and mail surveys. Two questions are addressed: (1) what is the effect of different data collection methods (face-to-face, telephone, mail) on the response rate, and (2) is there evidence for interaction effects, that is, differential effects of background variables (e.g., saliency) on the response to different methods of data collection.

2. Design and method of analysis

Meta-analysis is a collection of methods to describe and analyze the research results of a number of empirical studies about a specific problem (Glass *et al.*, 1981). The steps of a meta-analytic study are comparable to those of an empirical study, with explicit procedures for defining the problem, collecting data (published and unpublished studies), coding, and conducting the statistical analyses.

While many early meta-analyses (cf Rosenthal, 1984) restricted themselves to determining a combined p-value for the significance of all empirical results, more recently the emphasis has shifted to measuring the effect size and explaining its variation across the different studies (Hunter & Schmidt, 1990). The explanatory variables in these meta-analyses are coded attributes of the studies such as year of publication or type of sample.

The principal difference between meta-analysis and the traditional narrative review is in the application of statistical methods to identify explanatory variables that affect the studies' outcomes (Bangert-Drowns, 1986). Specialized statistical techniques have been developed for meta-analysis; for an overview see Hedges and Olkin (1985).

The next two sections describe the design of our meta-analysis and the analysis model used.

2.1. *Design of the meta-analysis*

To locate the relevant research literature we searched on-line in Psychological Abstracts, Sociological Abstracts, Dialog/SSCI, and the SRM database.[1] In addition, a call for publications was included in three European newsletters. Finally, the reference lists of all publications were used in a snowball search for additional publications. The search was aimed at publications about comparative research related to the effect of data collection methods on the quality of the data collected. When more than one publication described the same research, all publications were taken together and coded as one single study. This search strategy produced an initial sample of 60 studies.

In a comparative analysis of response, it is important to know precisely how this response is defined. Following Goyder (1987, p. 9), we defined the completion rate (CR) as the number of completed interviews or questionnaires divided by the total sample, and the response rate (RR) as the number of completed interviews divided by the effective sample (the total sample minus the ineligibles, such as 'address does not exist' or 'target person deceased').[2] For 12 of the 60 publications it was impossible to determine a completion rate or a response rate (four of these studies do not even report a sample size). Another three studies did not give any codeable information. Thus, a total of 45 studies is available for the meta-analysis. Only nine studies compared all three data collection methods, the remaining 36 compared only two methods. The total number of data collection conditions implemented in the 45 studies is 99 (38 face-to-face, 35 telephone, and 26 mail).

The three data collection methods are coded into two (0/1) dummy variables, that code for the telephone method (0 = ftf, 1 = tel.) and the mail method (0 = ftf, 1 = mail); this makes the face-to-face method the 'reference' condition. The response type (completion rate or response rate) is coded by a binary dummy variable (0 = CR, 1 = RR); this makes the completion rate the reference condition. When both rates were available the completion rate was chosen because it provides a more realistic appraisal of the number of responses expected given the operational sample size. The variables that describe the studies are: Sample type (coded as 1 = convenience sample, 2 = existing panel, 3 = local random, 4 = national random); publication type (1 = unpublished, 2 = published); publication medium (1 = journal, 2 = book, report); year (of publication: 1947 = 0, 1948 = 1, etc.); survey threat (0 = topic not intrusive, . . . , 2 = topic strongly intrusive); saliency (0 = none, . . . , 2 = strong); research quality (z-score based on homogeneity analysis of eight quality indicators); type of research organization (1 = government, 2 = university, 3 = market research, 4 = commercial). Coding was done by two independent coders using a detailed codebook; the inter-coder correlation is 0.93.[3]

Of the 45 studies available for the meta-analysis six studies were unpublished reports or papers, five were books and 34 were journal articles. The articles were dispersed over 20 different scientific journals; the most frequent journal outlet was *Public Opinion Quarterly*. Most studies were American (32) or Dutch (8). The total number of target respondents in the 45 studies was estimated as 106310, of which 79696 participated. Thus, the overall completion rate was 75%.

2.2. *Method of analysis*

The basic idea of meta-analysis is to conduct a statistical analysis on the results from previous studies. This strategy effectively treats the set of available results as if they came from one comprehensive study. This presents several distinct methodological problems (Bangert-Drowns, 1986).

The first problem is conceptual. It may be difficult to determine the identical dependent variable in all studies, because different studies may employ different operationalizations of the same construct. In our case, the outcome variable 'response' is not identical in all studies, since some studies report the completion rate (number completed/total sample) and others the response rate (number completed/effective sample). We decided to use the completion rate, because it is both more realistic and easier to interpret. If a study reports the response rate, but contains sufficient information to calculate the completion rate, the completion rate is calculated and used. Otherwise the response rate is used, and this is indicated by coding the explanatory variable type of response as '1.' Possible differences in the effect of other explanatory variables on the completion rate and the response rate are investigated by including interaction terms between these explanatory variables and the dummy variable 'type of response' in the model.

The second problem concerns the statistical model used to integrate the results from the separate studies. In a meta-analysis the available results may not be independent. In our case we have 99 different data collection conditions, contained in 45 studies. Conditions that are part of the same study will share many characteristics (such as the topic, population, and fieldwork organization) that may influence the response. Consequently, conditions within the same study will be more alike than conditions from different studies. An extreme form of lack of independence presents itself in the study variables; by definition all conditions within the same study will have identical values for explanatory variables such as publication medium or publication year.

The third problem is also statistical, and concerns the homogeneity of the results from the separate studies. The goal of meta-analysis is to summarize the results of different studies. Only if all differences between the studies can be explained as the mere effect of sampling variation it is allowed to combine the separate results into one final statistic, which estimates the outcome of all studies taken together. In this case the results are called homogeneous. If the results from the separate studies differ more than can be explained by sampling variation, they are called heterogeneous, and it is dangerous to combine them to reach an overall average. Instead, the main focus of the meta-analysis shifts to explaining the systematic variation between the separate studies.

Several different statistical models have been proposed to test whether the results of a set of studies are homogeneous (cf Hedges & Olkin, 1985) or to analyze non-independent results (Rosenthal & Rubin, 1986). As Bryk and Raudenbush (1992) pointed out, multilevel analysis serves both purposes (analyzing non-independent data and testing for homogeneity) within one comprehensive statistical model. A multilevel model is a statistical model for data that are hierarchical, with variables defined at each level of the hierarchy. In our case, there are two hierarchical levels: we have 45 studies containing 99 conditions. The variables at the condition level are the dependent variable 'response' and the two explanatory (dummy) variables coding for the telephone and the mail condition. We have several explanatory variables at the study level, such as saliency and year of publication. Our model is a multilevel regression model, also known as the hierarchical linear model or the random coefficient model; for an overview see Goldstein (1987) and Bryk and Raudenbush (1992). The multilevel regression model accommodates the dependencies in the data by introducing residual error terms at the study level. The question whether the results from the 45 studies are homogeneous or heterogeneous translates to the question whether or not the regression coefficients at the condition level (the intercept and the slopes for the telephone and mail dummy) vary across the studies. If a regression coefficient is invariant across studies, it is called a fixed coefficient that describes a homogeneous effect. If a regression coefficient varies, it is a random coefficient that describes a heterogeneous effect. The variation of such a random regression coefficient can be modelled by including study level explanatory variables in the analysis. This serves one of the main purposes in modern meta-analyses: explaining variation in effect sizes between studies.

The dependent variable 'response' is a proportion. An appropriate model for the analysis of proportions is the generalized linear model (GLM, cf. McCullagh & Nelder, 1989). Generalized linear models consist of three components: (1) a linear regression equation; (2) a specific error distribution; (3) a link function that transforms the predicted values to the scale of the observed values. For the analysis of proportions, the standard link function is the logit function, with a binomial error distribution. The interpretation of the regression coefficients is in terms of the underlying logit scale defined by the logit link function. For interpretation of our results, we will transform the predicted logits back to proportions.

Generalized linear models for hierarchical data are described by Wong and Mason (1985), Longford (1988, 1990), and Goldstein (1991). We used Longford's VARCL program, which estimates the parameters of the generalized hierarchical linear model by a quasilikelihood procedure (Longford, 1988), and also supplies standard errors for the parameter estimates and an overall deviance for the model. The statistical model is given in Appendix A.

Table 1. Mean completion rate (CR) and response rate (RR) by condition (percentages)

	F-t-F	Tel.	Mail	Total
CR	70.9	63.1	60.9	65.4
RR	73.5	70.3	68.2	71.0

3. Results

3.1. *Observed response by condition*

If all studies compared all three data collection methods and reported both a completion rate and a response rate, the outcomes could be summarized simply by giving the mean response by condition, as we do in Table 1.

However, most studies compared only two conditions, and while some studies report both a completion rate and a response rate, most studies report only one of these. Thus, on the one hand, the percentages in Table 1 are not independent, and on the other hand, they are partly based on dissimilar studies. As a consequence, a direct comparison of the percentages in Table 1 is misleading. The relatively low response rate, say, for the mail survey, could well be caused by differences between the studies that contain a mail condition and those that do not. We need a statistical model that incorporates both the dependencies within and the differences between the studies. The next section describes the results of a multilevel analysis that accomplishes both objectives.

3.2. *Multilevel modeling*

The dependent variable 'response' is the completion rate (CR). When the completion rate cannot be determined, the response rate (RR) is used. A dummy variable (CR = 0, RR = 1) indicates which response type is used. Differences in the response process for completion rate and response rate can be modeled incorporating interaction terms between this (explanatory) dummy variable and the other explanatory variables.

To assess possible differences in the response process between the completion rate and the response rate, a multilevel model was fitted separately for both the completion rate and the response rate. Three explanatory variables had a substantially different regression weight in the two analyses: sample type, saliency, and type of research organization. For these variables, we computed the interaction with the dummy variable 'response type,' to be included in the subsequent analyses. In all analyses, estimates are considered statistically significant if they exceed twice the corresponding standard error.

Table 2. Results of multilevel modeling

Step:	1 null model	2 condition model	3 study model	4 random model	5 cross-level model
Fixed Eff.					
intercept	0.81	1.10	1.16	1.06	1.05
tel. dummy		−0.32	−0.34	-0.19^{ns}	-0.14^{ns}
mail dummy		−0.59	−0.62	−0.39	−0.39
saliency			0.50	0.53	0.47
sample type			−0.45	−0.40	−0.23
research organization			−0.49	−0.44	−0.45
response type (RR)			0.61	0.59	0.56
year					−0.02
interaction					
resp. type × sample type			−0.48	−0.47	
cross-level interactions					
tel. × sample type					−0.27
mail × year					0.02
Random Eff.					
σ^2	1.00	1.00	1.00	1.00	1.00
$\sigma^2_{\text{intercept}}$	0.61	0.57	0.30	0.21	0.26
σ^2_{tel}				0.39	0.35
σ^2_{mail}				0.44	0.31
deviance	105625	103087	99835	98248	98602

The multilevel model is built up by a stepwise procedure:

(1) Intercept-only or Null-model: no explanatory variables.
(2) Condition-model: add both explanatory variables (telephone and mail dummy) at the condition level with fixed slopes.
(3) Study-model: add all explanatory variables at the study level that show a significant effect.
(4) Random-slope model: specify random slopes for both condition level variables, if the error variance is significant.
(5) Cross-level model: explain random slope variation (step 4) by adding significant interactions between the random variables at the condition level and explanatory variables at the study-level.

Table 2 presents the results for each step.

All effects in Table 2 exceed twice the corresponding standard error, except those marked ‘ns’. The figure of 1.00 for the condition level variance σ^2 is a fixed value, not an estimate, and hence will not be interpreted (cf. Appendix A). The explanatory variables ‘mail’, ‘telephone’ and ‘response types’ are dummy (0, 1) variables, therefore their slopes can be directly interpreted as the expected difference between the condition indicated and the reference (zero) condition. For instance, the value of 0.56 for the 'response type' dummy in the last model indicates that, controlling for the other explanatory variables in the model, the response rate (RR) is on average 0.56 higher than the completion rate (CR).

The null-model estimates the variance of the intercepts ($\sigma^2_{\text{intercept}}$) across studies as 0.61. This can be viewed as the baseline estimate of the unexplained between-study variance (cf. Bryk & Raudenbush, 1992).

In the next model, the condition model, the explanatory dummy variables 'telephone' and 'mail' are added. Both show a significant effect in the expected direction: in the telephone condition the response is lower than in the face-to-face (reference) condition, and in the mail condition the response is even lower. The variance of the intercepts drops to 0.57.

Adding explanatory variables at the study level (the study model in Table 2) reduces the variance of the intercepts across studies to almost half its original value. The regression slopes for these variables are all in the expected direction. A salient topic increases the response. Samples that are national and random receive a lower response than local and nonrandom samples, and studies with a market research or commercial background have a lower response than government or university sponsored studies. The significant effect of the response type represents the difference between the completion rate and the response rate. The significant interaction between the response type and the sample type shows that this difference is larger when local nonrandom samples are used. Presumably, in these studies more information is available about the nonrespondents, which increases the number of known ineligibles and thus the response rate, but does not change the completion rate. After including the explanatory variables in the model, the slope estimates for the telephone and mail dummies are virtually unchanged, which indicates that the differences between the conditions cannot be explained by differences between the studies in the available background variables.

The next model is the random model, which assumes that the differences between the conditions (as reflected in the variances σ^2_{tel} and σ^2_{mail} of the slopes of the dummy variables 'telephone' and 'mail') vary across studies. The random model reveals large and significant variances for these slopes, while the estimates of these slopes are much lower than in the previous (fixed) models. The regression slope for the 'telephone' variable is no longer significant, although still in the expected direction. We will consider the interpretation of these regression slopes in the discussion section.

In the final analysis step we model the variance of both regression slopes by introducing cross-level interactions. Only two interactions turn out to be significant: the interaction between the mail dummy and the publication year,

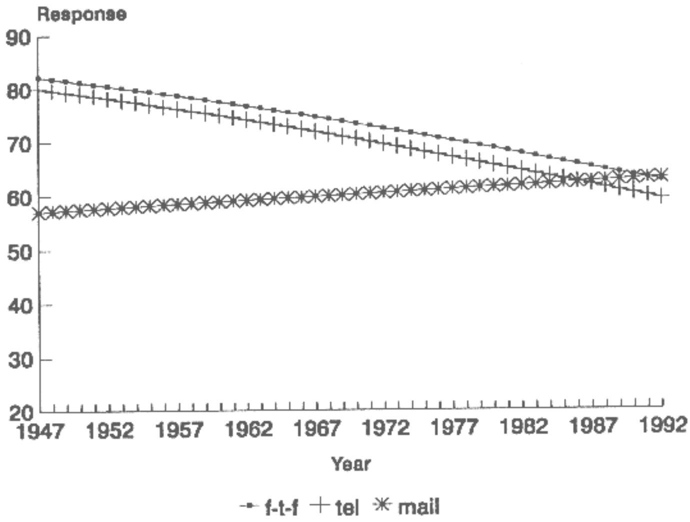

Fig. 1.

and the interaction between the telephone dummy and the sample type. Over the years, the difference between the mail condition and the other two conditions becomes smaller, and the telephone condition performs better with local samples. When these interactions are added to the model, the response × sample interaction from step four is no longer significant, and it is dropped. Since the year of publication is part of an interaction, this variable is added to the model (Jaccard *et al.*, 1990).[4] It shows a significant effect in the expected direction (in previous models it showed an effect in the same direction that fell just short of being significant at the 5% level). Finally, we note that the variance of the slopes in the last model is substantially lower than in the previous model; adding the two cross-level interactions explains part of the slope variation. The remaining variances σ^2_{tel} and σ^2_{mail} are still significant, meaning that we have not succeeded in explaining *all* slope variation.

The interaction between the telephone dummy and the response type shows that the telephone method gets a relatively better response in local and selected samples. The interaction between the mail dummy and the year of publication shows that the difference between the mail survey and the other two methods has become smaller over the years. To aid the interpretation of this interaction, it is useful to plot the regression slopes for 'year' for all three data collection methods in one figure.

Figure 1 shows that in the cross-level model the general trend over the

Table 3. Estimated completion rate (CR) and response rate (RR) by condition (percentages)

	F-t-F	Tel.	Mail	Total
CR	70.3	67.2	61.3	66.9
RR	80.5	78.2	73.5	78.0

years is a lowering response, but that the mail survey appears to be an exception. In the last year of observation (1992), the predicted differences have all but disappeared.

3.3. *Model-based response by condition*

The completion rates and response rates for the three data collection methods, reported earlier in Table 1, did not correct for systematic differences between the studies. The multilevel analyses reported in the previous section show that a number of explanatory variables have substantial effects upon the response, indicating substantial differences between the studies. In the next table the completion rates and response rates for the three data collection methods are reported, corrected for these variables.

The estimates in Table 3 are based on the final model presented in Table 2 of Section 3.2. To calculate the estimated values, all study level variables were set at their mean value over all studies. This removes all differences between the data collection methods that are the result of imbalances in the empirical comparisons. To facilitate interpretation, the predicted logits have been transformed into percentages.

Compared to the uncorrected results in Table 1, the pattern is the same, but there are two marked differences. Overall, the differences between the three methods have become smaller. Apparently the response in the telephone and mail conditions is lower in part because these conditions are more often implemented in studies that report an overall lower response for other reasons, such as using a nationwide random sample. Next, the difference between the completion rate and the response rate has become much larger. This can be explained by researchers' preference to report the higher response rate in those studies that have to report a low overall response rate, and the lower completion rate in studies that can boast a high overall response. Thus Table 1, in contrast to Table 3, underestimates the difference between the completion rate and the response rate and overestimates the differences between data collection methods.

4. Discussion

The direct effects found in the multilevel model are consistent with the results previously reported in the research literature (e.g., Heberlein and Baumgartner, 1978; Goyder, 1987; Groves, 1989). A salient topic increases the response. National and random samples receive a lower response than local and selected, nonrandom samples, and studies with a market research or commercial background receive a lower response than government or university sponsored studies.

As for to the central problem formulated in the introduction, we conclude that there are significant differences in the response to face-to-face, telephone, and mail surveys. There are also some interaction effects: the sample type and year of publication have a differential effect on the response in the three methods.

The large and significant variation of the regression coefficients for the telephone and mail condition across studies has important implications. On the underlying logit scale, the regression slope for 'mail' is -0.39 in model 4 of Table 2). This value is significant, which means that *on the average* face-to-face surveys perform better than mail surveys. However, σ^2_{mail}, the variance of the distribution of this regression slope across the studies, is estimated as 0.44, which corresponds to a standard deviation of 0.63. Using the standard normal distribution we calculate that in 27 percent of similarly conducted studies this regression coefficient may actually be expected to be larger than zero! It is instructive to see that, even if there is little doubt that *on the average* the mail survey has a lower response rate than the face-to-face interview, there still is an appreciable chance that *in a specific study* the relationship found may actually be the opposite. The interaction term between 'mail' and 'publication year' in model 5 explains part of this variation: in model 5 the regression slope for mail remains -0.39, but σ^2_{mail} decreases to 0.31. This residual variation is still statistically significant; it implies that in 24 percent of similarly conducted studies we may expect to find a higher response to mail surveys.

In the introduction, we referred to the declining willingness of the general public to take part in surveys. In our last model (model 5), the interaction term implies that the mail survey goes against this trend; Figure 1 shows this in a vivid graphic. The conclusion that in recent years the face-to-face method does no longer guarantee the highest response has been reported before. Goyder (1987, p. 67) reports that the response to face-to-face interviews is declining, while the response to mail surveys proves to be stable. A similar pattern has been noted by De Leeuw (1992) in the Netherlands, and Lyberg and Lyberg (1990) in Sweden. The recent findings of relatively high responses to mail surveys give survey researchers some reason for optimism, if we interpret them as the outcome of the large amount of research effort put into improving the response rates to mail surveys (Dillman, 1978, 1991). The amount of research into improving the response to the other data collection methods is substantially less (cf. Goyder, 1987). It is plausible that a similar research effort to improve the response to interview methods would also have a sizeable effect.

Finally, it is interesting to note that survey threat, type or medium of publication, and research quality are all unrelated to the response. The lack of significance of type and medium of publication is reassuring, since it indicates that there is probably little publication bias in the meta-analysis. We have not included the *country of origin* in the multilevel analysis, because this would introduce a categorical variable with a large number of sparsely filled categories. However, if we analyze the residuals from the last model by country of origin, we find no differences. It appears that our results hold for all regions present in the meta-analysis (Northern American and Western European countries).

Acknowledgements

We thank Nick Longford, Bill Yeaton and Godfried van den Wittenboer for their comments on an earlier draft.

The data have been partly funded by the Netherlands Organization for Scientific Research (NWO) under grant number 500278008. The second author has contributed in the context of the NESTOR research program 'Living arrangements and social networks of older adults', conducted by the Vrije Universiteit Amsterdam and the Netherlands Interdisciplinary Demographic Institute in The Hague.

Appendix A: The hierarchical logit model for proportions

Let $\underline{P}_{ij}$ represent the observed proportion of respondents in condition i of study j, and π_{ij} the corresponding population proportion.[5] Then,

$$\underline{P}_{ij} = \pi_{ij} \tag{1}$$

$\underline{P}_{ij}$ given π_{ij} has a binomial distribution with variance $\sigma^2(\pi_{ij}/\ (1 - \pi_{ij}))$.

The parameter σ^2 is a scaling factor that is fixed at 1.00 and left uninterpreted.

The logit link function is defined by $\text{logit}(x) = \ln(x/(1-x))$. Using a logit regression model for the proportion P leads to a linear regression model for the logits of the proportions. The simplest multilevel model for $\text{logit}(\underline{P}_{ij})$ is:

$$\text{logit}(\underline{P}_{ij}) = \underline{\beta}_{0j} \tag{2}$$

and

$$\underline{\beta}_{0j} = \gamma_{00} + \underline{\delta}_{0j}. \tag{3}$$

Substituting (3) into (2) gives

$$\text{logit}(\underline{P}_{ij}) = \gamma_{00} + \underline{\delta}_{0j}. \tag{4}$$

In equation (4) known as the null or 'intercept only' model, the overall proportion is given by the intercept γ_{00}, and the residual error terms $\underline{\delta}_{0j}$ are assumed to have a normal distribution on the logit scale.

The complete multilevel model for one explanatory variable X_{ij} at the condition level and one explanatory variable Z_j at the study level can be written as:

$$\text{logit}(\underline{P}_{ij}) = \underline{\beta}_{0j} + \underline{\beta}_{1j} X_{ij} \tag{5}$$

and

$$\underline{\beta}_{0j} = \gamma_{00} + \gamma_{01} Z_j + \underline{\delta}_{0j} \tag{6}$$

$$\underline{\beta}_{1j} = \gamma_{10} + \gamma_{11} Z_j + \underline{\delta}_{1j} \tag{7}$$

which gives

$$\text{logit}(\underline{P}_{ij}) = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} X_{ij} Z_j + \underline{\delta}_{0j} + \underline{\delta}_{1j} X_{ij}. \tag{8}$$

Equation (5) states that the (logit of the) response is predicted by the condition-level variable X_{ij} (for instance the dummy denoting the mail condition), and that both the intercepts $\underline{\beta}_{0j}$ and the slopes $\underline{\beta}_{1j}$ are assumed to vary across the studies. Equation (6) is a regression equation which predicts the values of the intercepts $\underline{\beta}_{0j}$ by the study level variable Z_j, and Equation

(7) is a regression equation which predicts the values of the slopes $\underline{\beta}_{1j}$ by the study level variable Z_j. By substitution we obtain the single equation version of the model in (8). Model (8) states that the proportion of respondents in condition i of study j is explained by a linear regression model which includes an explanatory variable at both the condition and the study level. Variation in the regression slope $\underline{\beta}_{1j}$ for the condition-level explanatory variable X_{ij} can be explained by the interaction term $X_{ij}Z_j$ in (8). The residual error terms $\underline{\delta}_{0j}$ and $\underline{\delta}_{1j}$ are assumed to have a normal distribution on the logit scale with variances σ^2_{00} and σ^2_{11}.

Notes

1. SRM is a Dutch inter-university service that catalogs and indexes international publications on Social Science Methodology.
2. There are more comprehensive definitions of response (cf. Groves, 1989, par. 4.2), but the available publications do not give sufficient information to calculate them.
3. This is the combined correlation for only the judgmental variables survey threat and saliency. All other variables can be assessed objectively, and the inter-coder agreement for these variables was 100%. For the final coding, differences were discussed until consensus was reached.
4. To ease interpretation, the explanatory variables making up the interaction terms have been centered around their overall mean (Jaccard *et al.*, 1990); the interaction terms have not been centered.
5. In the following sections, random variables and coefficients are underscored.

References

Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method, *Psychological Bulletin* 99: 388–399.

Betlehem, J. G. & Kersten, H. M. P. (1981). The nonresponse problem, *Survey Methodology* 7: 130–156.

Bryk, A. S. & Raudenbusch, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park: Sage.

De Leeuw, E. D. (1992). *Data Quality in Mail, Telephone and Face-to-Face Surveys*. Amsterdam: Vrije Universiteit (Doctoral dissertation).

Dillman, D. A. (1978). *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley.

Dillman, D. A. (1991). The design and administration of mail surveys, *Annual Review of Sociology* 17: 225–249.

Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail survey response rate; A meta-analysis of selected techniques for inducing response, *Public Opinion* 52: 467–491.

Goldstein, H. (1987). *Multilevel Methods in Educational and Social Research*. New York: Oxford University Press.

Goldstein, H. (1991). Nonlinear multilevel models, with an application to discrete response data, *Biometrika* 78: 45–51.

Glass, G. V., McGaw, B. & Smith, M. L. (1981). *Meta-Analysis in Social Research*. Beverly Hills: Sage.

Goyder, J. (1987). *The Silent Minority*. Cambridge: Blackwell.

Groves, R. M. & Kahn, R. L. (1979). *Surveys by Telephone, A National Comparison with Personal Interviews*. New York: Academic Press.

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

Groves, R. M. & Lyberg, L. E. (1988). An overview of nonresponse issues in telephone surveys, pp. 191–212 in R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (eds), *Telephone Survey Methodology*. New York: Wiley.

Heberlein, T. A. & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature, *American Sociological Review* 43: 447–462.

Hedges, L. V. & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando: Academic Press.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of Meta-Analysis*. Beverly Hills: Sage.

Jaccard, J., Turrisi, R. & Wan, C. K. (1990). *Interaction Effects in Multiple Regression*. Newbury Park: Sage.

Longford, N. T. (1988). *A Quasi-Likelihood Adaptation for Variance Component Analysis*. Princeton, NJ: Educational Testing Service.

Longford, N. T. (1990). VARCL. *Software for Variance Component Analysis of Data with Nested Random Effects (Maximum Likelihood)*. Princeton, NJ: Educational Testing Service.

Lyberg, I. & Lyberg, L. (1990). Nonresponse Research at Statistics Sweden. Paper presented at the First Workshop on Household Survey Nonresponse. Stockholm, Oct. 15–17, 1990.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman & Hall.

Rosenthal, R. (1984). *Meta-Analytic Procedures for Social Research*. Beverly Hills: Sage.

Rosenthal, R. & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes, *Psychological Bulletin* 99: 400–406.

Steeh, C. G. (1981). Trends in nonresponse rates, 1952–1979, *Public Opinion Quarterly*, 45. Reprinted as pp. 32–49 in E. Singer & S. Presser (1989), *Survey Research Methods*, A Reader. Chicago: University of Chicago Press.

Sugiyama, M. (1992). Response and non-response, pp. 227–239 in L. Lebart (Ed.), *Quality of Information in Sample Surveys*. Paris: Dunod.

Wong, G. Y. & Mason, W. M. (1985). The hierarchical logistic regression model for multilevel analysis. Extensions of the hierarchical normal linear model for multilevel analysis, *Journal of the American Statistical Association* 80: 513–524.

Yu, J., & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires, *Journal of Marketing Research* 23: 36–44.