2008

Optimal Estimation of Interviewer Effects for Binary Response Variables through Partial Interpenetration

N. von Sanden
Australian Bureau of Statistics

David Steel
University of Wollongong, dsteel@uow.edu.au

Centre for Statistical and Survey Methodology

The University of Wollongong

Working Paper

04-08

Optimal Estimation of Interviewer Effects for Binary Response Variables through Partial Interpenetration

Nicholas von Sanden and David Steel

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Optimal Estimation of Interviewer Effects for Binary Response Variables through Partial Interpenetration

Nicholas von Sanden
Methodology Division, Australian Bureau of Statistics
nick.von.sanden@abs.gov.au

David Steel
Centre for Statistical and Survey Methodology, University of Wollongong
dsteel@uow.edu.au

Estimates of sampling variance underestimate the variance of survey estimates when there are strong interviewer effects. However, interviewer effects are rarely considered in complex field surveys due to the high costs involved with adapting an interpenetrating design to produce estimates of them. This paper demonstrates how interviewer effects can be estimated by taking a sample of interviewers using the technique of partial interpenetration. The partial interpenetration approach is developed to produce cost-optimal survey designs for the estimation of the interviewer effect and population means for non-linear response variables.

Keywords: Interviewer effect; interpenetration; total survey error, non-sampling error, multiple objective survey design.

1 Introduction

Interviewers play a central role in the collection of high quality data in household surveys. They provide initial contact with respondents, elicit and prompt response and collect and enter data. The presence of the interviewer may also have unintended impacts on survey data, for example responses collected by the same interviewer tend to be more similar than if the responses were collected by different interviewers (see Collins, 1980).

Estimates of the contribution of the interviewer effect to the variance of estimates is necessary to produce estimates of the total variance of survey estimates. They are also useful in identifying problems in questionnaire design and interviewer training and performance.

Running large scale surveys is an expensive exercise and there is extensive literature dealing with minimizing sampling errors, e.g. Cochran (1977). Despite a concentrated effort to minimize sampling errors, classical techniques for estimating the contribution of the interviewer to total survey error, such as interpenetration (Mahalanobis, 1946) and re-interviewing (Bailar, 1968) do not occur often in practice due to associated costs and complexity.

In field enumeration surveys, interviewers collect information from respondents clustered in geographical areas. Interviewer effects and geographic effects will then be confounded and cannot be separately estimated if some form of repeated measurement (eg interpenetration or re-interviewing) does not occur. We now consider how partially interpenetrated survey designs can be used to produce optimal estimates of interviewer effects and population means.

The interviewer effect was recognized in the early social surveys of the 20th century. For example Rice (1929) realized that interviewers with different political opinions tended to obtain different results in a survey of destitute men. Early studies concentrated on establishing the existence of interviewer effect (eg Mahalanobis, 1946) while later studies, such as Hansen and Marks (1958), attempted to establish the relative importance of interviewer effects compared with other sources of error. A review of studies estimating the interviewer effect prior to 1980 has been provided by Collins (1980).

Since the 1980s increased computing power and use of statistical models in data analysis has led to the use of multi-level models to directly estimate the interviewer effect. For example Anderson and Aitken (1985) investigated interviewer variability in a survey on consumer spending. The multi-level modelling approach can be applied to estimate the interviewer effect and also cater for the hierarchical structure of datasets. Subsequent work by Hox *et al.* (1991); Pannekoek (1991); Wiggins *et al.* (1992); Hox (1994); Goldstein (1995); Pickery and Loosveldt (2000, 2001, 2004); O'Muircheartaigh and Campanelli (1998, 1999) and Martin and Beerten (2002) have applied and extended this approach.

In the above papers it is assumed that there is either an explicitly interpenetrated survey design in which a minimum of 2 interviewers are allocated to each workload or area or an effectively interpenetrated design in which all of the interviewers are allocated to a single concentrated geographic area. This occurs even in the case of Pickery and Loosveldt (2000, 2001) who consider application of the longitudinal information available in repeated panel surveys and Schnell and Kreuter (2005) who explore separating interviewer and sampling-point effects in a fully interpenetrated survey. This paper will extend the above work to show how estimates of the interviewer effect can be produced optimally, in practice, under budget constraints.

2 Interpenetrated Sampling

Previous studies to estimate the interviewer effect have relied on costly fully interpenetrated designs which are rarely applied in practice for surveys which involve field interviewing because of the costs and complexity of having 2 interviewers in each workload or geographic area. To produce interviewer effect estimates in such surveys we can simply select a sample of workloads to interpenetrate. We call this technique partial interpenetration. This compares with the classical full interpenetration of Mahalanobis (1946) in which at least 2 interviewers are allocated to each geographical area. In the following we have randomly selected interviewers to allocate to interpenetrated areas, however more complex sampling schemes can also be adopted.

Statistically the interviewer effect is the variance of the interviewer level residuals. Specification of this problem as a mixed model enables us compare the properties of fully and partially interpenetrated designs. Mixed models are useful for estimating interviewer effects as they borrow strength from other groups, and as a general rule the higher the number of interpenetrated geographical areas the better our estimate of the interviewer effect.

For example, under the linear mixed model (McCullagh and Nelder, 1989)

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{1}$$

where,

- $\mathbf{y}$ is a $(n \times 1)$ vector of observations.
- $\mathbf{X}$ is a $(n \times q)$ matrix of observed covariate values. $\mathbf{X}$ is also referred to as the fixed effect design matrix.
- β is a $(q \times 1)$ vector of coefficients for the covariates (i.e. there are q covariates included in this model or $q-1$ if a mean response is included).
- $\mathbf{Z}$ is a $(n \times t)$ matrix of known values indicating the presence of random effects. $\mathbf{Z}$ is also referred to as the random effect design matrix. $\mathbf{Z}$ provides an indication of group memberships.

- $\mathbf{u}$ is a $(t \times 1)$ vector of random effects (i.e. there are t random effects included in the model).

- $\mathbf{e}$ is a $(n \times 1)$ vector of residuals.

If we also assume that both $\mathbf{e}$ and $\mathbf{u}$ are independent and normally distributed with expected values of zero and variances matrices of $\mathbf{R}$ and $\mathbf{D}$ respectively, we can see that the response $\mathbf{y}$ will also be normally distributed with variance matrix

$$\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathrm{T}} + \mathbf{R} \tag{2}$$

The variance covariance matrix, $\mathbf{V}$, can be partitioned into components representing

- The survey and workload design. This is captured in the random effect design matrix, $\mathbf{Z}$. Individual columns of $\mathbf{Z}$ can be used to describe different random effects relating to the survey design attributes such as the allocation of interviewers or the geographical clustering of primary sampling units.

- The magnitude of variance components, such as the interviewer effect, σ^2_{int} and the spatial effect, σ^2_{wk}. This is captured in the variance component matrix, $\mathbf{D}$. The elements of $\mathbf{D}$ will correspond to random effects described in the columns of $\mathbf{Z}$.

- The magnitude of the residual variation, σ^2_{ε}. This is captured in the residual matrix, $\mathbf{R}$.

A similar decomposition can be performed under the Generalized Linear Mixed Model (GLMM) (see McCulloch and Searle, 2001).

Then we can see that the degree of interpenetration relates only to the random effects design matrix, $\mathbf{Z}$, and any change to the degree of interpenetration will also effect the variance covariance matrix, $\mathbf{V}$, from which estimates of $\mathbf{D}$ are isolated. Thus the degree of interpenetration influences our ability to estimate the interviewer effect no matter the magnitude of the interviewer effect. A discussion of methods for estimating variance components in GLMMs can be found in Browne and Draper (2006).

Variance estimates associated with an estimate of the interviewer variance can be obtained from the information matrix based on the distribution of the response variable and given estimates of $\mathbf{D}$ and $\mathbf{R}$. We can calculate a Variance Inflation Factor (vif) for estimates derived under a given survey design compared with the estimates we would have obtained under full interpenetration to assess the impact of interpenetration on the estimation of the interviewer effect. Given a partially interpenetrated survey design and its associated design matrix, $\mathbf{Z}$, let $\mathbf{Z}^*$ be a design matrix describing the same spatial structure in which all observations within a single spatial zone are collected by different interviewers and in which all interviewers collect data from more than one workload. Then $\mathbf{Z}^*$ will be fully interpenetrated and comparable with the partially interpenetrated $\mathbf{Z}$. Then if $Var(\hat{\sigma}^2_{int})_{\mathbf{Z}}$ is the variance of the estimate of the interviewer effect under $\mathbf{Z}$ the vif can be calculated as

$$vif_{\mathbf{Z}} = \frac{Var(\hat{\sigma}^2_{int})_{\mathbf{Z}}}{Var(\hat{\sigma}^2_{int})_{\mathbf{Z}^*}} \tag{3}$$

McCulloch and Searle (2001) show that there is no general expression for the information of the random effects for all possible response distributions and hence variance estimates of the random effect estimates must be considered for each distribution. As an example consider a normally distributed response variable, for which each element in the ith row and jth column of the information matrix for the random effect is

$$I(\sigma^2)_{\{i,j\}} = \frac{1}{2}tr(\mathbf{V^{-1}Z_iZ_i^TV^{-1}Z_jZ_j^T}) \tag{4}$$

where $\mathbf{Z_i}$ and $\mathbf{Z_j}$ correspond to the ith and jth column of $\mathbf{Z}$ respectively and which therefore correspond to different random effects. Calculation of (4) requires an estimate of $\mathbf{V^{-1}}$, which in general requires information regarding both $\mathbf{D}$ and $\mathbf{R}$. In practice this means that we need to know the magnitude of all of the variance components in order to properly assess the effect of partial interpenetration on the variance of our interviewer effect estimates. Although in some cases prior knowledge may give us an approximate idea as to the magnitude of the interviewer effect, in general this information will not be available during the survey design process.

In order to design optimal interpenetrating surveys for the purpose of estimating the interviewer effect we either need approximate prior estimates of the magnitude of all variance components or we need to be able to make general statements regarding the relationship between the degree of interpenetration and the vif associated with the interviewer effect estimates no matter the *true* magnitude of the variance components. Also, because many variables collected in household surveys are categorical we must consider mixed models for non-normal data.

Studies examining multilevel survey design have generally either been simulation based empirical studies, e.g. Mok (1995); Afshartous (1995); Normand and Zou (2002) or theoretical expositions considering the accuracy of the fixed effect parameter estimates conditioning on cost and design, e.g. Snijders and Bosker (1993, 1999); Cohen (1998); Moerbeek *et al.* (2000, 2001a,b); Moerbeek and Wong (2002). An extension of the Cohen (1998) paper is presented by Moerbeek *et al.* (2001a) who derive a linearization for the variance of the fixed effect parameter in multilevel models with logistic response. Moerbeek *et al.* (2001a, p 18) state that

'*... optimal designs cannot be derived analytically for PQL* (Penalized Quasi-Likelihood) *and numerical integration*'

As Marginal Quasi-Likelihood (MQL) estimates are generally biased when considering non-normal response variables (see Rodriguez and Goldman, 1995, 2001; Breslow, 2003) they present a general methodology for empirically determining the sampling variance of parameters in the multilevel logistic model. Moerbeek *et al.* (2001a) conclude that design decisions based upon biased MQL linearization of the variance of the fixed effect parameter estimates will generally be similar to those that would have been determined empirically using unbiased estimation methods. However, simulation techniques should be applied to explore the implications of design scenarios on variance component parameters such as the interviewer effect.

We now explore the relationship between the survey design used to allocate interviewers to workloads and estimates of the interviewer effect. Our initial focus will be to minimize the variance of estimates of the interviewer effect for a given cost function. For binary data items this entails producing estimates of a vif comparing interviewer effect estimates under competing survey designs. We present a general empirical methodology for exploring the impact of survey design on estimates of the interviewer effect and thereby establish a relationship between the degree of interpenetration and the variance that can be associated with interviewer effect estimates with particular focus on non-linear response variables. We have already seen that numerical integration techniques are to be preferred when faced with a non-linear response variable. Consequently, extending the work of Moerbeek *et al.* (2001a), design scenarios for the optimal estimation of the interviewer effect will be assessed through MCMC simulation techniques.

2.1 Variance Inflation Factors for Logistic Response

In the following we consider a simple binary response multilevel model.

$$\Pr(y_{ijk} = 1|\pi_{jk}) = \frac{\exp(\pi_{jk})}{1 + \exp(\pi_{jk})} \tag{5}$$

where π can be decomposed into variance components corresponding to the levels in the dataset.

$$\pi_{jk} = \mu + \phi_k + \theta_j$$

and

- i, j and k are indices referring to the person/individual level, the interviewer and workload levels respectively.
- μ is a fixed effect.
- The random effects are independent and normally distributed, i.e. $\phi_k \sim N(0, \sigma^2_{wk})$ and $\theta_j \sim N(0, \sigma^2_{int})$.

Based on model (5) a simple 3 level logistic response multilevel model was simulated, with a design matrix based on a known degree of inter-workload interpenetration and equal workload sizes. The following parameter settings were applied, $\mu = 2.5$, $\sigma^2_{int} = 0.5^2$, $\sigma^2_{wk} = 1.5^2$, $n = 5000$, $n_{int} = 100$ and $n_{wk} = 50$ and full intragroup interpenetration was required for all interpenetrated groups, ie all responses in an interpenetrated group were collected by different interviewers. The degree of interpenetration was controlled by allocating a single interviewer to only enumerate each non-interpenetrated workload and all respondents in interpenetrated groups were randomly allocated one of the remaining interviewers. Note that under this scheme there was no scenario in which only one workload was interpenetrated (as interviewers in interpenetrated workloads were required to also collect data from other workloads). When there were more respondents in a workload than available interviewers, all available interviewers were allocated as close to an equal number of respondents as possible. Consequently for $d_Z = 0.5$ there would be

- 25 interviewers who fully enumerate 25 workloads and do not collect data from any other workload.

- 75 interviewers who collect at least 1 response from each of the remaining 25 workloads. 25 of these 75 interviewers will collect 2 responses in any of these interpenetrated workloads, while 50 interviewers will only collect one response.

The empirical variance of the interviewer effect estimate based on a number of different degrees of inter-workload interpenetration was then calculated via equation (6), using $R = 250$ simulations for each degree of interpenetration. MCMC estimation in MLwiN was used to produce each estimate of $\sigma^2_{int,r}$.

$$\hat{\text{var}}(\hat{\sigma}^2_{int}) = \frac{1}{R-1} \sum_r \left(\hat{\sigma}^2_{int,r} - \frac{\sum_r \hat{\sigma}^2_{int,r}}{R} \right)^2 \tag{6}$$

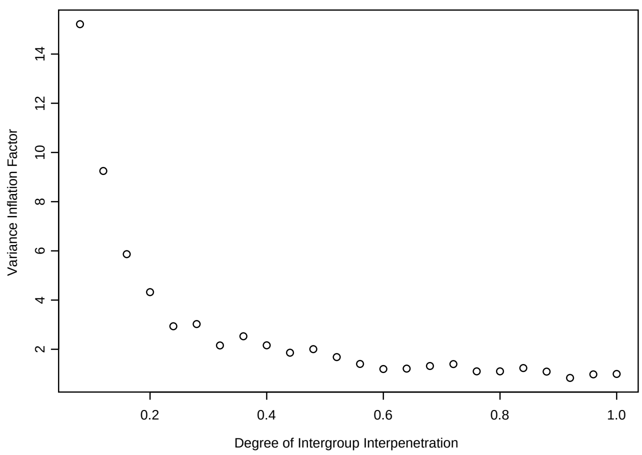

Figure 1: Variance Inflation Factor for Interviewer Effect Estimate by Degree of Intergroup Interpenetration: Logistic Response Model

The variance inflation factors for the interviewer effect estimate against the degree of intergroup interpenetration can be seen in Figure 1

In Figure 1 the variance inflation factor on the interviewer effect estimate increases as the degree of intergroup interpenetration decreases. However the variance inflation factor does not exceed 1.5 until less than 50% of the workloads are interpenetrated, i.e. the degree of intergroup interpenetration falls below 0.5. Moreover we can see that the variance inflation factor increases rapidly for lower degrees of intergroup interpenetration, asymptoting to infinity when there is full confounding. Consequently, under this simple model, we would be able to make a reasonable estimate of the interviewer effect without full interpenetration.

Based on this empirical technique we can now compare any two competing survey designs and through this determine the most appropriate survey design for the estimation of the interviewer effect. As increased degrees of interpenetration are generally associated with increased travel costs this suggests a cost optimal partially interpenetrated design.

3 Optimal Design Based on Travel Cost Functions

Application of empirical techniques for optimal design purposes would require an estimate of the magnitude of the interviewer effect. However in practice we generally cannot obtain this estimate until after the survey has been conducted. We therefore need to establish a general relationship that can be applied in practice to determine an optimal interpenetrating design for the estimation of the interviewer effect without prior knowledge of the magnitude of the interviewer effect. We have seen in Figure 1 that there appears to be a relationship between the degree of intergroup interpenetration and the variance of the interviewer effect estimate. Next we will explore the general relationship between the variance of interviewer effect estimates, the sample size and the degree of intergroup interpenetration in order to produce an approximate relationship that can be applied in practice to produce an optimal partially interpenetrated survey design based on a specified cost function.

3.1 Travel Cost Function

Although full interpenetration will generally lead to a more reliable estimate of the interviewer effect this will require interviewers to travel between workloads to interview respondents, leading to increased travel costs. A fully confounded design will minimize travel costs as all of the respondents enumerated by a single interviewer will reside in closer proximity to one another.

By way of example assume the cost of an interviewer travelling to an interview in a different workload is four times that of the interviewer travelling to meet any new respondent within the same workload, i.e. let $c_1 = 1$ cost unit and $c_2 = 4$ cost units. This is a strong simplification, roughly equivalent to assuming all workloads are adjacent. The total travel costs are related to how many workloads an interviewer travels to and how many observations the interviewer collects in each workload rather than just the total number of workloads. For a fixed sample size, n, increased levels of interpenetration will generally lead to higher travel costs. With a fixed budget the methodologist can therefore either choose to design a survey with an increased total sample size or with a higher degree of interpenetration. If the aim is to produce as accurate estimates of the interviewer effect as possible based on a given budget, this will imply an optimal degree of interpenetration. We can demonstrate this with a simple cost function

$$C = (c_1 + c_3)\, n + (c_2 - c_1) \sum\nolimits_i b_i \tag{7}$$

where

- C is the total cost.
- c_1 is the total cost associated with enumerating different respondents in the same workload.

- c_2 is cost of travelling to interview the first respondent in a different workload.
- c_3 is the total cost associated with including an extra respondent in the sample.
- b_i is total number of workloads in which interviewer i conducts interviews.

$\sum_i b_i$ is related to the proportion of intergroup interpenetration, d_Z, implied by the random effects design matrix, $\mathbf{Z}$. If we assume our fixed body of interviewers is greater than the given number of workloads, i.e. $n_{int} > n_{wk}$, and that we have full intragroup interpenetration in any interpenetrated groups then we can say that

$$\begin{aligned} \sum_i b_i &= n_{wk} + n_{wk} d_Z \left(n_{int} - n_{wk} - 1 + n_{wk} d_Z \right) \\ &= n_{wk} \left\{ 1 + d_Z \left(n_{int} - 1 + n_{wk} \left[d_Z - 1 \right] \right) \right\} \end{aligned} \quad (8)$$

This result recognizes that there must be at least one interviewer allocated to each of the n_{wk} areas. In the $n_{wk} d_z$ interpenetrated areas full intragroup interpenetration is assumed so that all remaining available interviewers (i.e. take the total number of available interviewers minus the number who are already enumerating non-interpenetrated areas; $n_{int} - n_{wk} + n_{wk} d_z$) collect the data. Result (8) allows us to restrict consideration to only designs containing full intragroup interpenetration. Other forms of intragroup interpenetration can also be specified in a similar way, for example random allocation of two interviewers to each workload would lead to $\sum_i b_i = n_{wk}(1 + d_Z)$ provided $n_{int} \geq n_{wk}$ and full confounding implies $\sum_i b_i = n_{wk}$ when $n_{int} = n_{wk}$.

In large scale surveys the number of workloads is generally determined geographically and can therefore be considered as fixed. The hiring and training of interviewers is a slow and costly process and hence for design purposes we will also consider the body of available interviewers to be fixed. Note that with longer lead-in periods it will be possible to prepare further interviewers, however this scenario has not been considered here. Given cost coefficient estimates, c_1, c_2 and c_3 the total cost is a simple function of both the degree of intergroup interpenetration and the sample size. Combining (7) and (8) then gives

$$C = (c_1 + c_3)\, n + (c_2 - c_1)\, n_{wk} \left\{ 1 + d_Z \left(n_{int} - 1 + n_{wk} \left[d_Z - 1 \right] \right) \right\} \quad (9)$$

We have already seen in Figure 1 that the variance of the interviewer effect estimate is also a function of the degree of intergroup interpenetration, and so we can minimize this variance subject to the cost constraint (9) to determine the optimal degree of interpenetration for the estimation of the interviewer effect. To do this we first need to establish the relationship between the vif, which we are trying to minimize, and the remaining variables in the cost function (9). As c_1, c_2, c_3, n_{wk} and n_{int} are generally all fixed this means we need to establish the relationship between the vif, the sample size, n and the degree of intergroup interpenetration, d_Z. The following section will begin by examining the relationship between the variance of the interviewer effect estimate and the degree of intergroup interpenetration.

3.2 Relationship Between vif and d_Z

The more interviewers we observe the lower the variance that will be associated with the interviewer effect estimate. Also as the degree of interpenetration increases we would expect the variance of the interviewer effect estimate to fall and there would be an approximate inverse relationship between vif and d_Z.

Figure 2 shows the estimated relationship between the inverse of the variance inflation factor and the degree of intergroup interpenetration and can be used to estimate the approximate relationship. This data was simulated based on model (5) and with a design matrix based on a known degree of inter-workload interpenetration and equal workload sizes. The following parameter settings were applied, $\mu = 2.5$, $\sigma^2_{int} = 0.5^2$, $\sigma^2_{wk} = 1.5^2$, $n = 5000$, $n_{int} = 100$ while $n_{wk} = 50$, Full intragroup interpenetration was required for all interpenetrated groups.

Figure 2 suggests that the relationship between the inverse of the variance inflation factor and the degree of intergroup interpenetration is approximately linear. The fitted OLS regression line has an R-Squared of 0.94 with an estimated intercept that is not significantly different from 0 and a slope that is not significantly different from 1.

There is an inverse relationship between the variance inflation factor, vif, on the interviewer effect estimate and the degree of intergroup interpenetration, d_Z, conditional on a fixed degree of intragroup interpenetration (ie the level of repeated use of interviewers in workloads) in any interpenetrated groups. Let v_{dZ} be the variance of the interviewer effect estimate associated with a specific degree of intergroup interpenetration, d_Z. Under our assumptions, for a number of fixed design parameters, C, c_1, c_2, c_3, n, n_{int}, n_{wk} and the degree of intragroup interpenetration, we can see that

$$vif_{dZ} = \frac{v_{dZ}}{v_{dZ|dZ=1}} \simeq \frac{1}{d_Z} \tag{10}$$

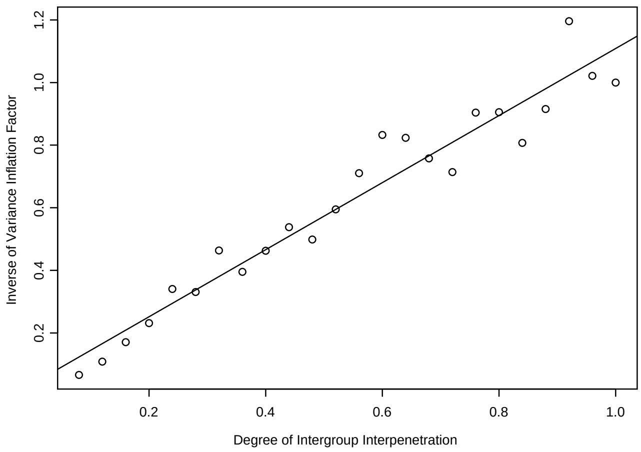

Figure 2: Inverse of Variance Inflation Factor for Interviewer Effect Estimate by Degree of Intergroup Interpenetration: Logistic Response Model

Based on this relationship we can now make some statements regarding optimal design of partially interpenetrated surveys for the estimation of the interviewer effect. Recall, however, that larger sample sizes will generally lead to more accurate estimates of the interviewer effect under full interpenetration. In other words $v_{dZ|dZ=1}$ is also a function of the sample size, n, conditional on the population parameters, σ^2_{int}, σ^2_{wk}, σ^2_{ε} and μ, the fixed design parameters, C, c_1, c_2, c_3, n_i, n_{wk} and the degree of intragroup interpenetration as specified by $\sum_i b_i$. The following section will explore the relationship between the variance of the interviewer effect estimate under full interpenetration and the sample size and assess the implications of this relationship for optimal interpenetrating survey designs.

3.3 Relationship Between $v_{dZ|dZ=1}$ and the Sample Size

We can develop an initial analytic understanding of the relationship between the variance of the interviewer effect estimate under full interpenetration and the sample size, by considering the simple case of a balanced 2 level (respondent at level 1 and interviewer at level 2) Hierarchical Linear Model (HLM) for a normally distributed response variable. This can be done by adapting an asymptotic expression for the second level variance component. For example, if we take the asymptotic expression for the second level variance component in the 2 level HLM (Longford, 1993, p 58) and consider the case of equal size workloads, ie all interviewers collect data from $\bar{n}$ respondents so that $\bar{n} = \frac{n}{n_{int}}$, we can write the asymptotic variance of the second level variance component as

$$\begin{aligned} \mathrm{var}(\sigma^2_{\mathrm{int}}) &= \frac{2\left(\sigma^2_\varepsilon\right)^2}{n}\left(\frac{1}{\bar{n}-1} + 2\omega + \bar{n}\omega^2\right) \\ &= \frac{2\left(\sigma^2_\varepsilon\right)^2 n_{int}}{n\left(n - n_{int}\right)} + \frac{4\left(\sigma^2_\varepsilon\right)^2\omega}{n} + \frac{2\left(\sigma^2_\varepsilon\omega\right)^2}{n_{int}} \end{aligned} \quad (11)$$

Where $\omega = \frac{\sigma^2_{int}}{\sigma^2_\varepsilon}$. By definition $n \geq n_{int} \geq 1$ and holding constant n_{int} (as recruiting and training interviewers is generally much more costly than altering the sample size) we can then see in (11) that as the sample size, n, increases, the variance of the interviewer effect will approach $\frac{2\left(\sigma^2_\varepsilon\omega\right)^2}{n_{int}}$. However, generally we are worried about the impact of interviewer effects when cost constraints limit both the number of interviewers and the sample size. In general we cannot say anything about the actual magnitude of σ^2_ε and σ^2_{int}, although we would generally expect that $\sigma^2_\varepsilon > \sigma^2_{\mathrm{int}}$ for carefully run surveys. In this case we would expect that (11) will be dominated by the second term within the range of interest for the methodologist. Consequently we would expect that for the purposes of designing cost effective partially interpenetrated surveys there will be an approximately inverse relationship between $v_{dZ|dZ=1}$ and the sample size.

In the logistic response multilevel model (5) the relationship between the sample size and the variance of the interviewer effect estimate is not immediately clear as the expansion under MQL is biased and as Moerbeek *et al.* (2001a) point out expressions for the variance of the interviewer effect cannot be derived analytically for PQL and numerical integration. The relationship between the variance of the interviewer effect estimate and the sample size can be determined empirically and this relationship is presented in Figure 3 following.

In Figure 3 the data is simulated according to Model (5) with the same parameter settings as before and full intragroup interpenetration for all interpenetrated groups. The variance of the interviewer effect estimate decreases as the sample size increases, eventually asymptoting to a level determined by the given design and population parameters. There appears to be an approximately inverse relationship between the variance of the interviewer effect estimate and the sample size and this relationship is presented in Figure 4 following

Figure 4 indicates that the relationship between the logarithm of the empirical variance of the interviewer effect estimate and the logarithm of the inverse of the total sample size is approximately linear. The fitted OLS

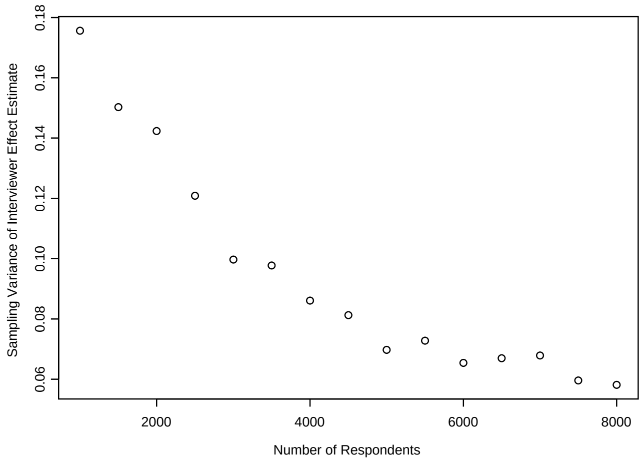

Figure 3: Variance of Interviewer Effect Estimate by Sample Size: Logistic Response Model under Full Interpenetration

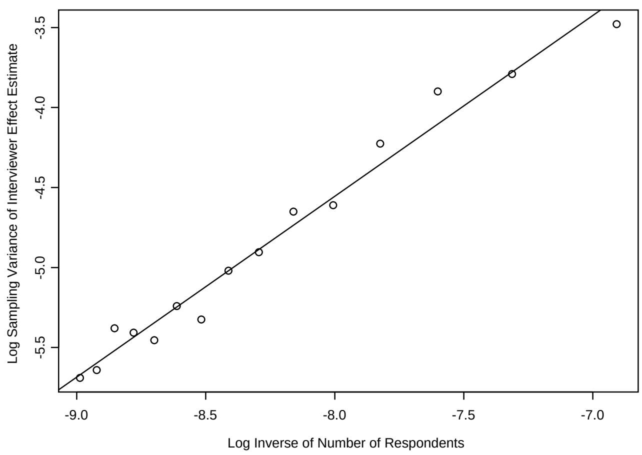

Figure 4: Log of Variance of Interviewer Effect Estimate by Log of Inverse of Sample Size: Logistic Response Model under Full Interpenetration

regression line has an R-Squared of 0.98 and the estimated coefficients can be found in Table 1.

There is an inverse relationship between the variance of the interviewer effect estimate under full interpenetration and the total sample size. Thus

$$v_{dZ|dZ=1} \simeq \frac{a_z}{n} \tag{12}$$

	Value	Std. Error	t value	P value
Intercept	4.4868	0.3983	11.2648	0.0000
Slope	1.1102	0.0481	23.4981	0.0000

Table 1: OLS Estimates: Log of Empirical Variance as the Dependent Variable and Log of the Inverse of Total Sample Size as the Explanatory Variable

where a_Z is a constant for any given set of population parameters, σ^2_{int}, σ^2_{wk}, σ^2_{ε}, μ, the fixed design parameters, n_{int}, n_{wk} and the degree of intragroup interpenetration as specified by $\sum_i b_i$. So in this case based on our parameter settings and assuming full interpenetration for each design, we can see in Table 1 that a_Z will be approximately equal to exp(4.4868) or 88.8.

3.4 Optimal Design for Estimation of Interviewer Variance

Putting together (12) and (10) we get the following relationship between the variance of the interviewer effect estimate, the degree of intergroup interpenetration and the sample size

$$v_{dZ} \simeq \frac{a_z}{n.d_Z} \tag{13}$$

Then combining (13) and (9) we can express the variance of the interviewer effect estimate for a fixed total cost C in terms of the degree of intergroup interpenetration

$$v_{dZ} \simeq \frac{a_z(c_1 + c_3)}{d_z \left\{C - (c_2 - c_1)[n_{wk} + n_{wk} d_z(n_{int} - n_{wk} - 1 + n_{wk} d_z)]\right\}} \tag{14}$$

The corresponding sample size is

$$n \simeq \frac{C - (c_2 - c_1)[n_{wk} + n_{wk} d_z(n_{int} - n_{wk} - 1 + n_{wk} d_z)]}{(c_1 + c_3)} \tag{15}$$

To find the optimal degree of interpenetration we minimize the variance of the interviewer effect as expressed in Equation (14), for a given total budget, C. As the numerator of (14) is a constant this is equivalent to maximizing the denominator of (14) over the entire range of possible degrees of intergroup interpenetration, i.e. $0 < d_Z \leq 1$. As the denominator is a cubic expression in d_Z, its derivative will be quadratic and the maximum value within this range will occur either when we have full interpenetration, i.e. $d_Z = 1$ or at a local maximum which can be determined by one of the quadratic roots in Equation (16) following

$$\frac{(c_1 - c_2)n_{wk}[n_{int} - n_{wk} - 1] \pm \Delta_C}{3(c_2 - c_1)n^2_{wk}} \tag{16}$$

where $\Delta_C = \sqrt{((c_1 - c_2)n_{wk}[n_{int} - n_{wk} - 1])^2 - 3((c_2 - c_1)n_{wk}^2)((c_2 - c_1)n_{wk} - C)}$.

By way of example consider a binary response variable simulated according to Model (5) as if it was collected via a survey with a total budget of $C = 10000$ and design parameters $n_{int} = 100$, $n_{wk} = 50$, $c_1 = 1$, $c_2 = 4$, $c_3 = 2$, design matrices, full intragroup interpenetration as specified by the form of $\sum_i b_i$ in (8) and finally $a_Z = 88.8$ (see Table 1). A plot of the variance of the interviewer effect estimate can be seen in Figure 5 following

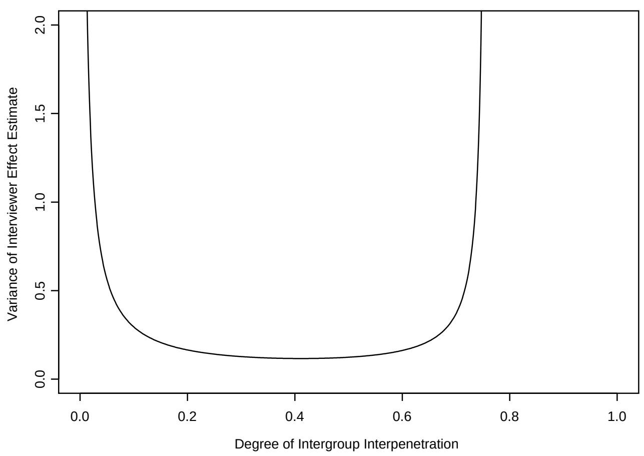

Figure 5: Variance of Interviewer Effect Estimate by Degree of Intergroup Interpenetration: Total Fixed Cost 10000

From Figure 5 we see that the total budget is not high enough for full interpenetration to be considered in this case as the highest degree of intergroup interpenetration affordable under this budget is $d_Z = 0.756$. Note also that as we approach full confounding, i.e. $d_Z \to 0$ it becomes harder to estimate the interviewer effect and the variance of the estimate approaches infinity. In comparison as we increase the degree of interpenetration we are forced to reduce our sample size, n, accordingly due to our total budget constraint. Consequently, as $d_Z \to 0.756$, $n \to 0$ the impact of this small sample size is that the variance of the interviewer effect estimate again approaches infinity.

When the sample size is high, for a fixed budget we can only afford a low degree of interpenetration, leading to an unreliable estimate of the interviewer effect. When the degree of interpenetration is high, however, we can only afford a small sample size, which again leads to unreliable estimates of the interviewer effect. Consequently we can see in Figure 5 that there is an optimal degree of intergroup interpenetration associated with the minimum possible variance of the interviewer effect estimate given our total budget constraint. In this case the optimal degree of interpenetration is not at the end points of the range $0 < d_Z \leq 1$ and the optimal degree of interpenetration is at a local minimum which can be determined by Equation (16). This corresponds to a degree of intergroup interpenetration of $d_z = 0.411$ with an implied sample size that can be determined by Equation (15) i.e. $n = 1854$ and an optimal variance of the interviewer effect estimate of $v(\hat{\sigma}^2_{int}) = 0.12$. We can also see in Figure 5 that the variance of the interviewer effect does not increase rapidly as we move away from the optimal degree of interpenetration and hence degrees of interpenetration near the optimal may still be applied to produce reliable estimates of the interviewer effect.

We can utilize Equation (16) to determine the minimum total budget required before the optimal degree of interpenetration occurs when the survey is fully interpenetrated. In other words to estimate the interviewer effect with unlimited finances we would require a total budget of at least

$$C \geq c_{12}n_{wk} + \frac{(3n^2_{wk}c_{12} + c_{12}n_{wk}[n_{int} - n_{wk} - 1])^2 - (c_{12}n_{wk}[n_{int} - n_{wk} - 1])^2}{3n^2_{wk}c_{12}} \tag{17}$$

where $c_{12} = c_2 - c_1$, for full interpenetration to be optimal. This reflects the point at which the positive root of (16) becomes greater than one. In the case of our example this means that full interpenetration is optimal under our cost function, design and population parameters when the total budget is greater than $C = 42,152$.

3.5 Effect of Optimal Interpenetrating Design on Sampling Variance of Mean

We have already seen that there is a cost associated with increasing the degree of interpenetration when designing a survey and hence under a fixed budget this will lead to a reduced sample size. In isolation we can use this information to determine an optimal degree of interpenetration for the estimation of the interviewer effect. In practice, however, any reduction in the sample size will have an impact on the sampling variance of other estimates, in particular estimates of means.

Consider the Sampling Variance (SV) component of the Total Variance (TV) associated with the sample mean, $\bar{y}_s$. Under a Simple Random Sampling WithOut Replacement (SRSWOR) sampling scheme, the sampling variance will be

$$SV\left(\bar{y}_s\right)=\left(1-\frac{n}{N}\right)\frac{S_Y^2}{n}$$

where the population size is N, n is the sample size and S_Y^2 is the adjusted population variance and therefore constant for a given population. Then given a fixed total budget C and information regarding the fixed design parameters, c_1, c_2, c_3, n_{int}, n_{wk} and the degree of intragroup interpenetration as specified by $\sum_i b_i$ we can calculate the optimal degree of intergroup interpenetration for estimating the interviewer effect. We can see from the cost function (7) that d_z determines n for a fixed C and so the degree of interpenetration determines the sample size and hence also impacts on the magnitude of the sampling variance.

We can then work out the variance inflation factors against what would be achieved at the optimal level of interpenetration for both the variance of the interviewer effect and the sampling variance of the mean. Figure 6 following compares the variance inflation factors for estimates of both the sampling variance and the variance of the interviewer effect, compared with the variance at the optimal degree of interpenetration for a binary response variable simulated according to Model (5), as if it was collected via a survey with a total cost of $C = 10000$ and design parameters $n_{int} = 100$, $n_{wk} = 50$, $c_1 = 1$, $c_2 = 4$, $c_3 = 2$, design matrices, full intragroup interpenetration as specified by the form of $\sum_i b_i$ in (8), $a_Z = 88.8$ and a large population size of $N = 1000000$. For our total budget of $C = 10000$ the optimal degree of intergroup interpenetration in this example is $d_Z = 0.411$ and the maximum degree of intergroup interpenetration affordable is $d_Z = 0.756$.

We can see in Figure 6 that as the sample size, n, increases, the sampling variance decreases. Consequently we get a lower sampling variance component of the total survey error if we decrease the degree of intergroup interpenetration for a fixed budget constraint. However, if we decrease the degree of intergroup interpenetration past the optimal level, in this case $d_Z = 0.411$ then the reduction in sampling variance comes at the expense of reduced accuracy of the interviewer effect estimate. We can also see the vif for the variance of the interviewer effect and the sampling variance are equal at a vif of 1 and a degree of intergroup interpenetration of $d_Z = 0.411$. This occurs because we have calculated the vifs with respect to the optimal degree of interpenetration.

When the sample size in Figure 6 is maximized this corresponds to a degree of interpenetration of $d_z = 0$, as for a fixed budget we can only afford to increase the degree of interpenetration by reducing the sample size. Thus when $n = 3283$ we cannot afford any interpenetrated areas and we cannot

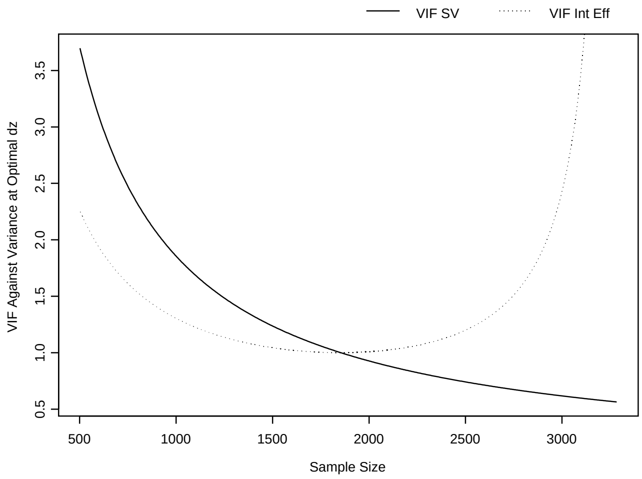

Figure 6: vifs Against Optimal Degree of Interpenetration for Sampling Variance and Variance of Interviewer Effect under SRSWOR for Total Cost 10000

produce an estimate of the interviewer effect. We can see in Figure 6 that this will lead to the lowest possible sampling variance. On the other hand if we reduced the sample size by a small margin, such as 270, we could afford a degree of interpenetration of $d_z = 0.1$, with an associated 9.2% increase in the sampling variance. With $d_z = 0.1$ the interviewer effect would be estimable, though with a vif of 2.53, so for a minor increase in the sampling variance we can greatly improve the reliability of estimates of the interviewer effect. If we wanted to produce the same estimate of the interviewer effect without altering the sampling variance (i.e. hold the sample size fixed) we can apply Equation (14) to show this could also be achieved by increasing the total budget by 9.6%

We can see in Figure 6 that as we use sample sizes less than that corresponding to the optimal degree of intergroup interpenetration, $n = 1854$, and therefore with a higher degree of interpenetration, the variance inflation factor for both the sampling variance and the variance of the interviewer effect increase, indicating these points are sub-optimal for minimizing either the sampling variance or the variance of the interviewer effect estimate. Consequently if we are interested in both the sampling variance and the variance of the interviewer effect estimate, we would never design a survey with a degree of intergroup interpenetration higher than the optimal degree of interpenetration chosen for the sole purpose of estimating the interviewer effect. On the other hand we can see that points to the right of the single-objective optimal degree of intergroup interpenetration lead to an increased variance for the interviewer effect estimate, but a decreased sampling variance. This implies that a degree of intergroup interpenetration less than the single-objective optimal may be preferred by the survey designer as it will lead to a lower sampling variance, even though the interviewer effect estimate will be less accurate than could have been achieved with the optimal degree of intergroup interpenetration. From our example above we can see that when faced with a fixed budget of $C = 10000$ we could achieve a 20% reduction in the sampling variance by accepting a 9% increase in the variance of the interviewer effect estimate compared with the single-objective optimal position, ie when $d_Z = 0.3$ and $n = 2323$.

Multiple objective designs can also be prepared which aim to simultaneously minimize both the total variance of the mean (TV) and the variance of the interviewer effect estimate. However vifs associated with the TV depend on the relative magnitude of the sampling variance and the interviewer effect as $TV(\bar{y}_s) = \left(1 - \frac{n}{N}\right)\frac{S_Y^2}{n} + \frac{\sigma_\varepsilon^2}{n} + \frac{\sigma_{\text{int}}^2}{n_{\text{int}}}$. Thus, although the sampling variance and individual level measurement error reduce as the sample size increases, the magnitude of the interviewer effect term in the total variance, $\frac{\sigma_{\text{int}}^2}{n_{\text{int}}}$, is fixed for a constant body of interviewers. Consequently, for a fixed sample size, to reduce the contribution of the interviewer effect to the total variance we would need to increase the number of interviewers, n_{int}, collecting data in the survey. For a fixed body of interviewers this suggests that Figure 6 presents a conservative relationship as the vif plot for the TV will be flatter than the vif plot for the SV. Hence it would generally be expected that varying the degree of interpenetration, and therefore the sample size, will have less of an effect on the TV. We must therefore make assumptions regarding the relative magnitude of the interviewer effect to the sampling and measurement variance to prepare optimal multiple objective designs which minimize both the total variance and the variance of interviewer effect estimates.

In Figure 7 vifs are presented for the total variance against the sample size when the interviewer effect is both a high and a low proportion of the total variance. When the interviewer effect comprises the majority of the total variance, then increasing the sample size, without altering the number of interviewers conducting the survey, does not greatly affect the total variance. In this case it is important that the interviewer effect is estimated as it will be the major source of uncertainty in the survey. On the other hand, if the interviewer effect is a relatively small component of the total variance then the sampling variance dominates and increasing the sample size has a strong impact on the total variance. It is therefore

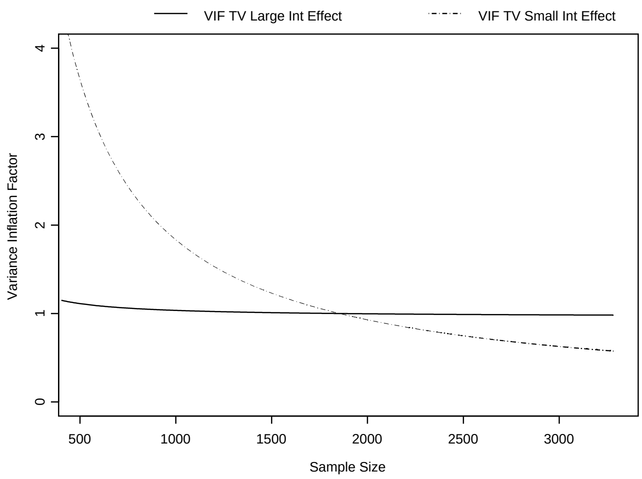

Figure 7: vifs for Total Variance with Interviewer Effect as Large or Small Proportion of Total Variance

important that interviewer effect estimates are prepared so that appropriate survey designs for minimizing the total variance in surveys can be made.

4 Discussion

Partial interpenetration combined with modern estimation techniques allows us to estimate the interviewer effect even in surveys with low degrees of interpenetration. In practice almost all surveys contain some degree of interpenetration even if they have not been designed for the purpose of estimating the interviewer effect and this opens up the possibility for widespread application of these techniques.

Provided there is no selection bias determining the interviewer allocation, the cost of estimating the interviewer effect under partial rather than full interpenetration is one of accuracy. We have demonstrated how to produce a valid estimate of the interviewer effect under partial interpenetration and the variance inflation factors that can be associated with these designs. Using these techniques it will generally be possible to produce an appropriate estimate of the interviewer effect with only a minor change to current survey designs and for a small increase in costs. Alternatively we can explicitly evaluate the increase in budget needed to provide estimates of interview variance.

Further extensions to this work would be to consider more complex travel cost functions and to explore the effect of practical considerations which will also influence workload formulation decisions. We have demonstrated the potential gain from utilizing partial interpenetration, but the actual gain will depend on the structure of the survey to which it is applied. Fully exploring the implications of a design for a non-linear response is still somewhat computationally intensive, however the approximate relationships allow rapid calculation of optimal design parameters. Multiple objective optimal designs can be considered in more detail and the potential benefits of explicitly incorporating available spatial or longitudinal information remains to be explored. Optimal partial intragroup interpenetration can also be considered which should lead to larger budget savings for a given sample size with smaller gains in the vif of interviewer effect estimates, as compared with the partial intergroup interpenetration considered here.

References

Afshartous, D. (1995) Determination of sample size for multilevel model design. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Anderson, D. A. and Aitken, M. (1985) Variance component models with binary response: interviewer's variability. *Journal of the Royal Statistical Society, B*, **47**, 203–210.

Bailar, B. (1968) Recent research in reinterview procedures. *Journal of the American Statistical Association*, **63**, 41–63.

Breslow, N. (2003) Whither pql? UW Biostatistics Working Paper Series.

Browne, W. and Draper, D. (2006) A comparison of bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, **1**, 473–514.

Cochran, W. (1977) *Sampling Techniques*. New York: Wiley, 3 edn.

Cohen, M. (1998) Determining sample sizes for surveys with data analyzed by hierarchical linear models. *Journal of Official Statistics*, **14**, 267–275.

Collins, M. (1980) Interviewer variability: A review of a problem. *Journal of the Market Research Society*, **22**, 77–95.

Goldstein, H. (1995) *Multilevel Statistical Models*. London: Edward Arnold.

Hansen, M. and Marks, E. (1958) Influence of the interviewer on the accuracy of survey results. *Journal of the American Statistical Association*, **53**, 635–655.

Hox, J., de Leeuw, E. and Kreft, I. (1991) The effect of interviewer and respondent characteristics on the quality of survey data: a multilevel model. In *Measurement*

Errors in Surveys (eds. P. Beimer, R. Groves, L. Lyberg, N. Mathiowetz and S. Sudman), chap. 22, 439–461. New York: John Wiley and Sons.

Hox, J. J. (1994) Hierarchical regression models for interviewer and respondent effects. *Sociological Methods and Research*, **28**, 300–318.

Longford, N. (1993) *Random Coefficient Models*. Oxford Science Publications. Oxford.

Mahalanobis, P. (1946) Recent experiments in statistical sampling in the indian statistical institute. *Journal of the Royal Statistical Society*, **109**, 325–370.

Martin, J. and Beerten, R. (2002) The effect of interviewer characteristics on survey response rates. Available on the web. Office of National Statistics.

McCullagh, P. and Nelder, J. (1989) *Generalized Linear Models*. London: Chapman and Hall, 2 edn.

McCulloch, C. and Searle, S. (2001) *Generalized, Linear, and Mixed Models*. New York: John Wiley and Sons.

Moerbeek, M., Gerard, J. and Berger, M. (2000) Design issues for experiments in multilevel populations. *Journal of Educational and Behavioural Statistics*, **25**, 271–284.

— (2001a) Optimal experimental designs for multilevel logistic models. *The Statistician*, **50**, 17–30.

— (2001b) Optimal experimental designs for multilevel models with covariates. *Communications in Statistics - Theory and Methods*, **30**, 2683–2698.

Moerbeek, M. and Wong, W. (2002) Multiple-objective optimal designs for the hierarchical linear model. *Journal of Official Statistics*, **18**, 291–303.

Mok, M. (1995) Sample size requirements for 2-level designs in educational research. *Multilevel Modelling Newsletter*, **7**, 11–15.

Normand, S. and Zou, K. (2002) Sample size considerations in observational health care quality studies. *Statistics in Medicine*, **21**, 331–345.

O'Muircheartaigh, C. and Campanelli, P. (1998) The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society A*, **161**, 63–77.

— (1999) A multilevel exploration of the role of interviewers in survey non-response. *Journal of the Royal Statistical Society A*, **162**, 437–446.

Pannekoek, J. (1991) A mixed model for analyzing measurement errors for dichotomous variables. In *Measurement Errors in Surveys* (eds. P. Beimer, R. Groves, L. Lyberg, N. Mathiowetz and S. Sudman), chap. 25, 517–530. New York: John Wiley and Sons.

Pickery, J. and Loosveldt, G. (2000) Modeling interviewer effects in panel surveys: an application. *Survey Methodology*, **26**, 189–198.

— (2001) An exploration of question characteristics that mediate interviewer effects on item nonresponse. *Journal of Official Statistics*, **17**, 337–350.

— (2004) A simultaneous analysis of interviewer effects on various data quality indicators with identification of exceptional interviewers. *Journal of Official Statistics*, **20**, 77–89.

Rice, S. (1929) Contagious bias in the interview. *American Journal of Sociology*, **35**, 420–423.

Rodriguez, G. and Goldman, N. (1995) An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society A*, **158**, 73–89.

— (2001) Improved estimation procedures for multilevel models with binary response: a case study. *Journal of the Royal Statistical Society*, **164**, 339–355.

Schnell, R. and Kreuter, F. (2005) Separating interviewer and sampling-point effects. *Journal of Official Statistics*, **21**, 389–410.

Snijders, T. and Bosker, R. (1993) Standard errors and sample sizes for two level research. *Journal of Educational Statistics*, **18**, 237–259.

— (1999) *Multilevel Analysis. An introduction to basic and advanced multilevel modelling*. London: Sage.

Wiggins, R., Longford, N. and O'Muircheartaigh, C. (1992) A variance components approach to interviewer effects. In *Survey and Statisical Computing* (eds. A. Westlake, R. Banks, C. Payne and T. Orchard), 243–254. Elsevier Science Publishers B.V.