



Catalogue no. 12-001-XIE

Survey Methodology

June 2004



How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.



Statistics Canada
Business Survey Methods Division

Survey Methodology

June 2004

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2006

All rights reserved. The content of this electronic publication may be reproduced, in whole or in part, and by any means, without further permission from Statistics Canada, subject to the following conditions: that it be done solely for the purposes of private study, research, criticism, review or newspaper summary, and/or for non-commercial purposes; and that Statistics Canada be fully acknowledged as follows: Source (or "Adapted from", if appropriate): Statistics Canada, year of publication, name of product, catalogue number, volume and issue numbers, reference period and page(s). Otherwise, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form, by any means—electronic, mechanical or photocopy—or for any purposes without prior written permission of Licensing Services, Client Services Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

April 2006

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (n° 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Weighted Estimation in Multilevel Ordinal and Binary Models in the Presence of Informative Sampling Designs

Leonard Grilli and Monica Pratesi¹

Abstract

Multilevel models are often fitted to survey data gathered with a complex multistage sampling design. However, if such a design is informative, in the sense that the inclusion probabilities depend on the response variable even after conditioning on the covariates, then standard maximum likelihood estimators are biased. In this paper, following the Pseudo Maximum Likelihood (PML) approach of Skinner (1989), we propose a probability-weighted estimation procedure for multilevel ordinal and binary models which eliminates the bias generated by the informativeness of the design. The reciprocals of the inclusion probabilities at each sampling stage are used to weight the log-likelihood function and the weighted estimators obtained in this way are tested by means of a simulation study for the simple case of a binary random intercept model with and without covariates. The variance estimators are obtained by a bootstrap procedure. The maximization of the weighted log-likelihood of the model is done by the MLMIXED procedure of the SAS, which is based on adaptive Gaussian quadrature. Also the bootstrap estimation of variances is implemented in the SAS environment.

Key Words: Informative design; Multilevel ordinal model; Multistage sampling; Pseudo Maximum Likelihood; Weighting.

1. Introduction

Multilevel models for ordinal responses, including binary responses as a special case, are frequently used in many areas of research for modeling hierarchically clustered populations. In fact, both in human and biological sciences, the status or the response of a subject may often be classified in two categories or in a set of ordered categories (ordinal or graded scale). At the same time, subjects are observed clustered in groups (*e.g.*, schools, firms, clinics, geographical areas). The hierarchical population structure is often also employed to design multistage sampling schemes, with unequal selection probabilities at some or all the stages of the sampling process. In the multilevel analysis of survey data, complex sampling schemes are often ignored even if they may cause the violation of the basic assumptions underlying multilevel models. In fact, in complex sampling designs both the subjects and the clusters at all levels could be selected with probabilities that, even conditionally on the covariates, do depend on the response variable; in other words, the sampling design might be informative.

For data that are clustered and obtained by multistage informative designs, proposals for fitting multilevel models have been formulated mainly for the case of continuous response variables. In particular, Pfeiffermann, Skinner, Holmes, Goldstein and Rasbash (1998) propose probability-weighting procedures of first and second level units that adjust for the effect of an informative design on the

estimation in two-level models with a continuous response variable. The method, known as Pseudo Maximum Likelihood (PML), consists in writing down a closed form expression for the census likelihood, estimating the log-likelihood function and then maximizing the estimated function numerically. The method needs the sampling weights for the sampled elements and clusters at all levels. The authors also develop appropriate ‘sandwich’ estimators for the variances of the estimators.

The work of Pfeiffermann *et al.* (1988) is mainly concerned with the implementation of the PML principle in the IGLS (Iterative Generalised Least Squares) algorithm (Goldstein 1986), which is suitable for linear multilevel models. The probability-weighted IGLS algorithm is available in the widespread package MLwiN (Rashbash, Browne, Goldstein, Yang, Plewis, Healy, Woodhouse and Draper 1999). However, the extension to nonlinear models is not trivial. For the nonlinear case the developers of MLwiN implemented a weighting procedure that parallels the one used for linear models with some ad hoc solution for the level 1 variation: for example, for binary responses the subject-level weights are included in the binomial denominator. The proposed method is straightforward to implement, but its properties have not been investigated yet. Moreover Renard and Molenberghs (2002) report the case of an application where the aforementioned algorithm for weighting in multilevel binary models did not converge or yielded implausible results.

1. Leonardo Grilli, Dipartimento di Statistica, Università di Firenze. E-mail: grilli@ds.unifi.it; Monica Pratesi, Dipartimento di Statistica e Matematica applicata all'Economia, Università di Pisa. E-mail: m.pratesi@ec.unipi.it.

The simulation study which we will use to judge the performance of the PML estimators will closely follow the lines of Pfeiffermann *et al.* (1988), since they use a similar approach for the linear model, so that some interesting comparisons are possible. However, when making the comparisons it should always be kept in mind that, while in the two-linear model the two variance components can be estimated separately, in the two-level binary model only a ratio of the two variance components is estimable, as discussed further on.

A recent paper which deals with the estimation of variance components is Korn and Graubard (2003), whose work is motivated by the substantial bias showed in small samples by several weighted estimators of variance components proposed to adjust for informative designs (Graubard and Korn 1996). Though the topic is same, the work of Korn and Graubard is different from ours in many respects: a) As Pfeiffermann *et al.* (1998), they consider only the linear multilevel model. b) In the context of the linear multilevel model, they focus on unbiased estimation of the variance components in small samples: in fact they propose some estimators for the variance components and only sketch how to derive similar estimators for the linear model with covariates, but without testing their performance. Anyway, the extension to nonlinear multilevel models is not trivial. c) The main estimators proposed by Korn and Graubard (2003), which are in closed form, showed good performance even in small samples. However they rely on the pairwise joint inclusion probabilities. When such probabilities are not available, which is often the case in practice, the authors propose a variant whose bias is substantial when the number of sampled clusters is moderate (33 in their simulation plan). In contrast, the PML method adopted in our work do not require joint inclusion probabilities. d) The informative design used by Korn and Graubard (2003) for their simulation study is quite different from ours: in fact, in their design the undersampling of the units depends on whether the model's random errors are greater than a certain threshold in absolute value, while in our design the criterion depends on whether the random errors are high or low. Therefore a comparison of the results is difficult.

The wide use of nonlinear multilevel models in many fields of application urges for a general and reliable weighted estimation method, which should be both effective and simple to implement, preferably in the framework of a standard statistical software. The present paper represents a contribution in this direction.

It is worth to note that the PML method we exploit is quite general, so it can be applied to a wide range of models. In the paper the focus is on models for ordinal and binary responses, since they are very common and can be

represented as a linear model for the latent response endowed with a set of thresholds (see section 2), facilitating the comparison with the existing results for the linear model. However the description of the PML approach is absolutely general and the estimation technique based on the NLMIXED procedure of SAS (reported in Appendix A) is easy to generalize.

The structure of the paper is as follows. Basic definitions for the multilevel ordinal model are out in section 2, while in section 3 the general PML approach is described, along with some details for fitting the model using SAS NLMIXED. In section 4 the properties of the various estimators for the random intercept binary model are evaluated by a simulation study. Section 5 concludes with some final remarks.

2. The Multilevel Ordinal Model

In order to ease the comparison with the results concerning the linear model (Pfeiffermann *et al.* 1998; Korn and Graubard 2003), it is useful to write the ordinal model in terms of a latent linear model endowed with a set of thresholds. Suppose that an observed ordinal response variable Y , with $k = 1, 2, \dots, K$ levels, is generated, through a set of thresholds, by a latent continuous variable \tilde{Y} following a variance component model (Hedeker and Gibbons 1994):

$$\tilde{Y}_{ij} = \beta' \mathbf{x}_{ij} + \omega u_j + \varepsilon_{ij}, \quad (1)$$

with $i = 1, 2, \dots, N_j$ elementary units (subjects) for the j^{th} cluster ($j = 1, 2, \dots, M$). In (1) \mathbf{x}_{ij} is a covariate vector and β is the corresponding vector of slopes; the random variables ε_{ij} and u_j are the disturbances, respectively at the first (subject) and second (cluster) level; and ω^2 is the second level variance component.

For the disturbances of model (1) we make the standard assumptions, *i.e.*, a) the ε_{ij} 's are iid with zero mean and unknown variance σ^2 ; b) the u_j 's are Gaussian iid with zero mean and unit variance; c) the ε_{ij} 's and u_j 's are mutually independent.

Note that model (1) leads to the simplest case of a multilevel ordinal model, with just two levels and a single random effect on the intercept; the extension to three or more levels and to multiple random effects is straight forward in principle (Gibbons and Hedeker 1997), but the complications in the formulae suggest to consider only the simplest case, which is sufficient for the discussion of the main conceptual issues.

The observed ordinal variable Y is linked to the latent one \tilde{Y} through the following relationship:

$$\{Y_{ij} = k\} \Leftrightarrow \{\gamma_{k-1} < \tilde{Y}_{ij} \leq \gamma_k\},$$

where the thresholds satisfy $-\infty = \gamma_0 \leq \gamma_1 \leq \dots \leq \gamma_{K-1} \leq \gamma_K = +\infty$. Therefore, conditional on u_j , the model probability for subject i of cluster j is

$$\begin{aligned} P(Y_{ij} = k | u_j) &= P(\gamma_{k-1} < \tilde{Y}_{ij} \leq \gamma_k | u_j) \\ &= P(\tilde{Y}_{ij} \leq \gamma_k | u_j) - P(\tilde{Y}_{ij} \leq \gamma_{k-1} | u_j), \end{aligned} \quad (2)$$

with

$$\begin{aligned} P(\tilde{Y}_{ij} \leq \gamma_k | u_j) &= P(\varepsilon_{ij} \leq \gamma_k - [\beta' \mathbf{x}_{ij} + \omega u_j] | u_j) \\ &= F\left(\frac{\gamma_k}{\sigma} - \left[\frac{1}{\sigma} \beta' \mathbf{x}_{ij} + \frac{\omega}{\sigma} u_j\right]\right) \\ &= F(\gamma_{\sigma, k} - [\beta'_{\sigma} \mathbf{x}_{ij} + \omega_{\sigma} u_j]), \end{aligned} \quad (3)$$

where $F(\cdot)$ is the distribution function of the standardized first level error term ε_{ij}/σ . All the model parameters are defined in terms of the unknown σ , the standard deviation of the first level error term, so only the ratios of the model parameters to the standard deviation of the first level error term are identifiable; we use the notation ψ_{σ} to indicate that the latent model parameter ψ is in σ units, *i.e.*, $\psi_{\sigma} \equiv \psi/\sigma$. Note that $F(\cdot)$ is also the inverse of the link function of the ordinal model: for example, the standard Gaussian distribution function yields the ordinal probit model.

As for identification, if β_{σ} includes the intercept, the estimable thresholds are $K-2$; so it is customary to set $\gamma_{\sigma, 1} = 0$. Alternatively, if the intercept is fixed to zero all the $K-1$ thresholds are estimable.

Now let θ denote the vector of all estimable parameters, which include β_{σ} , ω_{σ} and $K-2$ thresholds $\{\gamma_{\sigma, k} : k = 2, \dots, K-1\}$ ($\gamma_{\sigma, 1}$ is fixed to zero to insure identifiability). The conditional likelihood for subject i of cluster j is

$$L_{ij}(\theta | u_j) = \prod_{k=1}^K [P(Y_{ij} = k | u_j)]^{d_{ijk}}, \quad (4)$$

where $P(Y_{ij} = k | u_j)$ is defined by (2) and (3), while d_{ijk} is the indicator function for the event $\{Y_{ij} = k\}$. Then the marginal likelihood for cluster j is

$$L_j(\theta) = \int_{-\infty}^{+\infty} \prod_{i=1}^{N_j} L_{ij}(\theta | u) \varphi(u) du,$$

where φ is the standard Gaussian density function. Finally, the overall marginal likelihood is

$$L(\theta) = \prod_{j=1}^M L_j(\theta). \quad (5)$$

3. Probability-Weighted Estimation

3.1 Pseudo Maximum Likelihood (PML) Estimators

Suppose that the whole population of M clusters (level 2 units) with N_j elementary units (subjects or level 1 units) per cluster is not observed; instead the following two-stage sampling scheme is used:

- first stage: m clusters are selected with inclusion probabilities π_j ($j = 1, \dots, M$);
- second stage: n_j elementary units are selected within the j^{th} selected cluster with probabilities $\pi_{i|j}$ ($i = 1, \dots, N_j$).

The conditional sample inclusion probabilities are then $\pi_{ij} = \pi_{i|j} \pi_j$.

When the sampling mechanism is informative, *i.e.*, the π_j and/or the $\pi_{i|j}$ depend on the model disturbances and hence on the response variable, the maximum likelihood estimator of the parameters of the multilevel ordinal model defined in section 2 may be seriously biased.

A standard solution to this problem is provided by the Pseudo Maximum Likelihood (PML) approach (Skinner 1989). However in the context of multilevel models the implementation of the PML approach is complicated by the fact that the population log-likelihood is not a simple sum of elementary unit contributions, but rather a function of sums across level 2 and level 1 units. This can be seen by writing the logarithm of the likelihood (5) as follows:

$$\log L(\theta) = \sum_{j=1}^M \log \int_{-\infty}^{+\infty} \left[\exp \left\{ \sum_{i=1}^{N_j} \log L_{ij}(\theta | u) \right\} \right] \varphi(u) du. \quad (6)$$

A design consistent estimate of the population log-likelihood (6) can be obtained applying the Horvitz-Thompson principle, *i.e.*, replacing each sum over the level 2 population units j by a sample sum weighted by $w_j \equiv 1/\pi_j$ and each sum over the level 1 units i by a sample sum weighted by $w_{i|j} \equiv 1/\pi_{i|j}$:

$$\log \hat{L}(\theta) = \sum_j w_j \log \int_{-\infty}^{+\infty} \left[\exp \left\{ \sum_i w_{i|j} \log L_{ij}(\theta | u) \right\} \right] \varphi(u) du, \quad (7)$$

where \sum^s denotes a sum over sample units.

Note that inserting the weights in the log-likelihood implies the use of a design consistent estimator of the population score function. In fact, the population score function $U(\theta) \equiv \partial/\partial\theta \log L(\theta)$ can be written as

$$\sum_{j=1}^M \frac{\int_{-\infty}^{+\infty} \left[\exp \left\{ \sum_{i=1}^{N_j} \log L_{ij} \right\} \right] \cdot \left\{ \sum_{i=1}^{N_j} \frac{\partial}{\partial \theta} \log L_{ij} \right\} \varphi(u) du}{\int_{-\infty}^{+\infty} \exp \left\{ \sum_{i=1}^{N_j} \log L_{ij} \right\} \varphi(u) du}, \quad (8)$$

where $L_{ij} = L_{ij}(\boldsymbol{\theta} | u)$, whose corresponding Horvitz-Thompson estimator $\hat{U}(\boldsymbol{\theta})$ is

$$\sum_j^s w_j \frac{\int_{-\infty}^{+\infty} \left[\exp \left\{ \sum_i^s w_{ij} \log L_{ij} \right\} \right] \cdot \left\{ \sum_i^s w_{ij} \frac{\partial}{\partial \theta} \log L_{ij} \right\} \varphi(u) du}{\int_{-\infty}^{+\infty} \exp \left\{ \sum_i^s w_{ij} \log L_{ij} \right\} \varphi(u) du}, \quad (9)$$

which equals the score obtained by differentiating the probability-weighted loglikelihood (7).

Under mild conditions, the solution $\hat{\boldsymbol{\theta}}_{\text{PML}}$ to the estimating equations $\hat{U}(\boldsymbol{\theta}) = \mathbf{0}$ is design consistent for the finite population maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ which, in turn, is model-consistent for the super-population parameter $\boldsymbol{\theta}$: therefore $\hat{\boldsymbol{\theta}}_{\text{PML}}$ is a consistent estimator of $\boldsymbol{\theta}$ with respect to the mixed design-model distribution (Pfeffermann 1993).

Note that general probability-weighted estimators for nonlinear multilevel models can also be devised by weighting suitable estimating functions, as in the work of Singh, Folsom and Vaish (2002) in the context of small area estimation.

The implementation of the PML approach requires the knowledge of the inclusion probabilities at both levels. Using only second level weights or only first level weights may be insufficient or may even worsen the situation, as shown by our simulations.

3.2 Scaling the Weights

A controversial issue discussed in Pfeffermann *et al.* (1998) and Korn and Graubard (2003) is the scaling of the weights to obtain estimators with little bias even in small samples. Obviously, scaling is not relevant for the level 2 weights, since from (7) and (9) it is clear that multiplying the w_j 's by a constant does not change the PML estimates (it simply inflates the information matrix by that constant). On the contrary, scaling the level 1 weights may have important effects on the small sample behavior of the PML estimator. In the simulation study discussed in section 4 we present the results for the following type of scaling (named 'scaling method 2' in Pfeffermann *et al.* 1998):

$$w_{ij}^{\text{scaled}} = \frac{w_{ij}}{\bar{w}_j}, \quad (10)$$

where $\bar{w}_j = (\sum_i^s w_{ij})/n_j$, so that, for the j^{th} cluster, the sum of the scaled weights equals the cluster sample size n_j . In the present paper we do not wish to discuss the relative merits of the various scaling methods, so we limit our simulations to scaled weights (10), which have an intuitive meaning and showed good performance in the study of Pfeffermann *et al.* (1998), although they may yield a substantial bias with certain designs, as discussed in Korn and Graubard (2003). The topic will be broached again in section 4.

3.3 Estimation Technique

The maximization of the weighted log-likelihood (7) involves the computation of several integrals which do not have a closed-form solution, so a numerical approximation technique is required. When the dimensionality of the integrals is low, a simple and very accurate technique is Gaussian quadrature, which is based on a summation over an appropriate set of points. The NLMIXED procedure of SAS (SAS Institute 1999) is a general procedure for fitting nonlinear random effects models using adaptive Gaussian quadrature. Various optimization techniques are available to carry out the maximization; the default, used in the simulations of section 4, is a dual quasi-Newton algorithm, where dual means that the upgrading concerns the Cholesky factor of an approximate Hessian (SAS Institute 1999).

Though the NLMIXED procedure does not include an option for PML estimation, it is still possible to insert the weights in the likelihood, using different tricks for level 1 and level 2 weights, as explained in Appendix A.

3.4 Variance Estimation

In standard maximum likelihood the estimation of the covariance matrix of the estimators is obtained by inverting the information matrix. However this conventional estimator is not appropriate when using the PML method since it does not take into account the variability stemming from the sampling design. To get a more reliable covariance matrix Skinner (1989) proposed the use of a robust 'sandwich' estimator, which is employed also by Pfeffermann *et al.* (1998).

As noted in section 3.3, the NLMIXED procedure of SAS allows to fit the model with the PML approach, but the estimated covariance matrix, which is obtained by inverting the information matrix, is likely to be misleading in order to appreciate the actual variability of PML estimators. In the SAS framework the derivation of 'sandwich' estimators is not trivial. However, a simple and effective solution, requiring a bit of programming, is to empirically estimate the variance through the bootstrap technique for finite populations (Särndal, Swensson and Wretman 1992), which consists of the following steps: a) using the sample data, an artificial finite population is constructed, assumed to mimic

the real population; b) a series of independent bootstrap samples is drawn from the artificial finite population and for each bootstrap sample an estimate of the target parameter is calculated; c) the bootstrap variance estimate is obtained as the variance of the observed distribution of the bootstrap estimates.

The artificial finite population can be generated in the following way: i) for the j^{th} sampled cluster, each of the n_j sampled elementary units is replicated w_{ij} times, rounding the weight to the nearest integer, obtaining an artificial cluster of about N_j elementary units; ii) each of the m artificial clusters is replicated w_j times, rounding the weight to the nearest integer, obtaining an artificial population of about M clusters. Then the samples are selected from the artificial population in the following way: i) m clusters are resampled with probability proportional to π_j ; ii) for the j^{th} resampled cluster, n_j elementary units are resampled with probability proportional to π_{ij} .

When the sampling fraction m/M is low, most of the variance is due to the sampling of the clusters, so the bootstrap procedure described above could be simplified by omitting the steps concerning the elementary units, *i.e.*, step i) in the construction of the artificial population and step ii) in the resampling process.

A simpler resampling technique for variance estimation, considered by Korn and Graubard (2003), is the jackknife. In the case of clustered designs the technique entails the calculation of the variance from the set of point estimates obtained by deleting one cluster at a time, though the performance of the jackknife with correlated data is not always satisfactory (Shao and Tu 1995). In our simulation study the jackknife variance estimator seems unreliable, so it is not used. Further research is needed to fully evaluate the potentialities of the jackknife by testing some suitable modifications of the technique.

4. Simulation Study

4.1 Design of Experiment

The experiment reflects the two-stage scheme assumed for the observed variables: first, the finite population values are generated from the adequate superpopulation model (stage I) and then an informative or non-informative sample is selected from the finite population (stage II), with one sample per population. The two-stage selection scheme was repeated 1,000 times for each combination of sample size and type of informativeness. In order to compare our results with the ones obtained for the multilevel linear model, the experiment has been designed following the example of Pfeffermann *et al.* (1998, section 7).

The simulation study focused on a simple instance of the model defined in section 2, namely the random intercept probit binary model, which has only two categories for the response variable (*i.e.*, $K=2$) and one cluster-level Gaussian random error. To parallel the study of Pfeffermann *et al.* (1988) the main simulation plan refers to the model without covariates, but some additional simulations are conducted to assess the performance of the estimators in the model with one cluster-level covariate and one subject-level covariate.

The values of the binary response variable Y_{ij} were generated using the following two-stage scheme which parallels the one of Pfeffermann *et al.* (1998):

- Stage I. Finite population values Y_{ij} ($j=1, \dots, M$; $i=1, \dots, N_j$) were obtained by first generating a value from the superpopulation latent model $\tilde{Y}_{ij} = \beta + u_j + \varepsilon_{ij}$, with $u_j \sim N(0, \omega^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$, and then putting $Y_{ij} = 0$ if $\tilde{Y}_{ij} \leq 0$ or $Y_{ij} = 1$ if $\tilde{Y}_{ij} > 0$ (recall that the binary model has only one threshold which is set to zero to guarantee identifiability). The latent model parameter values employed in the simulation are $\beta = 0$, $\omega^2 = 0.2$ and $\sigma^2 = 0.5$, so that the parameters estimable from the binary model are $\beta_\sigma \equiv \beta/\sigma = 0$ and $\omega_\sigma \equiv \omega/\sigma = 0.632$ (see expression (3)). The hierarchical structure of the population comprises $M = 300$ clusters, while the cluster sizes N_j were determined by $N_j = 75 \exp(\tilde{u}_j)$, with \tilde{u}_j generated from $N(0, \omega^2)$, truncated below by -1.5ω and above by 1.5ω . As a result, in our population N_j lies in the range [38,147] with mean around 80.
- Stage II. Once the finite population values were obtained, we adopted one of the following sampling schemes:
 - (a) *Informative at both levels*: first, m clusters were selected with probability proportional to a ‘measure of size’ X_j , *i.e.*, $\pi_j = mX_j / \sum_{j=1}^M X_j$; the measure X_j was determined in the same way as N_j but with \tilde{u}_j replace by u_j , the random effect at level 2. The elementary units in the j^{th} sampled cluster were then partitioned into two strata according to whether $\varepsilon_{ij} > 0$ or $\varepsilon_{ij} \leq 0$ and simple random samples of sizes $0.25 n_j$ and $0.75 n_j$ were selected from the respective strata. The sizes n_j were either fixed, $n_j = n_0$, or proportional to N_j .
 - (b) *Informative only at level 2*: the scheme is the same as the previous one, except that simple random sampling was employed for the selection of level 1 units within each sampled cluster.

- (c) Non-informative: the scheme is the same as the previous one, except that the size measure X_j was set equal to N_j .

The simulation study included samples with $m = 35$ clusters and varying numbers of elementary units: large samples with fixed size $n_j = n_0 = 38$ and proportional allocation $n_j = 0.4N_j$, and small samples with fixed size $n_j = n_0 = 9$ and proportional allocation $n_j = 0.1N_j$ (mean of about 9).

The simulation study was carried out entirely within the SAS System (SAS Institute 1999), writing specific code with the macro language. The models were fitted with the NLMIXED procedure (see Appendix A), using 10–point adaptive Gaussian quadrature with a dual quasi-Newton algorithm, which reached convergence in a few iterations. As explained in appendix A, to avoid gross rounding errors the level 2 weights were pre-multiplied by a factor $k = 10,000$ and the estimated covariance matrix was then multiplied by the same factor.

4.2 Results

The results of the simulations are shown in Tables 1 and 2. For each sampling design the behavior of the point estimators of the intercept β_σ and the second level standard deviation ω_σ is summarized by the mean and standard deviation of their Monte Carlo sampling distribution. The point estimators under study are the standard maximum likelihood unweighted estimator and the following three weighted version of it: a) *cluster-level weighted*: the weights are only at level 2 (i.e., varying w_j 's and constant w_{ij} 's); b) *unscaled fully weighted*: the weights are at both levels and the level 1 weights are unscaled; c) *scaled fully weighted*: the weights are at both levels and the level 1 weights are scaled according to (10), i.e., 'scaling method 2' of Pfeffermann *et al.* (1998).

Our results are shown and discussed according to the following three scenarios: 1) *Base scenario*: the sampling design is non-informative. In this situation all the basic assumptions underlying the random intercept binary model are fulfilled, so this case can be assumed as a benchmark for judging the subsequent results. 2) *Informative/Unweighted scenario*: the sampling design is informative, while the estimator is unweighted. In this situation the basic assumptions underlying the random intercept binary model are violated because of the informativeness of the design and no adjustment is used. 3) *Informative/Weighted scenario*: the sampling design is informative and the estimator is weighted. Also in this case the basic assumptions underlying the random intercept binary model are violated, but the weights are introduced as a tentative adjustment for the bias of the standard estimator.

4.2.1 Base Scenario

When the sampling design is non-informative the standard maximum likelihood unweighted estimator is asymptotically unbiased (Tables 1 and 2: rows 9–12, column 1). However for small samples ($n_j = 9$ and $n_j = 0.1N_j$) there is an appreciable negative bias in the estimation of ω_σ .

If the weights are introduced when there is no need to adjust for the effect of the design (Tables 1 and 2: rows 9–12, columns 2–4), we face a slight increase in the variability of the estimators, which is more pronounced when the unscaled fully weighted estimator is used in small samples. Note that, still in small samples, the unscaled fully weighted estimator of ω_σ is upward biased.

4.2.2 Informative/Unweighted Scenario

The informativeness of the sampling design procedures biased and unstable estimates. The bias is still evident for large samples (Tables 1 and 2: rows 1–8, column 1). The conclusions are the same for both types of informative designs, though the bias tends to have a different sign. Moreover the informativeness of the design inflates the variability of the standard estimator with respect to the base scenario: in particular, when the design is informative at both levels the standard error of the estimator of β_σ is doubled.

4.2.3 Informative/Weighted Scenario

Estimation of β_σ .

The results in Table 1 show that, when the design is informative, the weighted-based adjustment is effective in removing the bias in the estimation of β_σ .

Particularly, when the design is informative only at level 2 (Table 1: rows 5–8, columns 2–4) and the weights are introduced only at this level (cluster-level weighted estimator), the bias in the estimation is corrected with no important increase in the sampling variance. The result is valid also for fully weighted estimators (unscaled or scaled). The bias correction works for small samples too.

When the design is informative at both levels (Table 1: rows 1–4, columns 2–4) and the weights are introduced at both levels (fully weighted estimators), the bias in the estimation of β_σ is corrected. Moreover, the fully weighted estimators have smaller sampling variance than the unweighted counterpart, except for the unscaled version in small samples. The scaled version is preferable especially in small samples. The scaled version is preferable especially in small samples, since it allows to achieve an unbiased estimator with a substantial lower sampling variance. It should be noted that when the design is informative at both levels, the cluster-level weighted estimator is worse than the standard unweighted estimator.

Table 1
Simulation Means and Standard Deviations (in parenthesis) of Point Estimators of the Intercept
(true value 0, number of replicates 1,000)

Sampling design	Unweighted estimator	Cluster-level weighted	Weighted estimators	
			Unscaled fully weighted	Scaled fully weighted
Informative at both levels				
Fixed size $n_j = 38$	-0.120 (0.212)	-0.411 (0.202)	0.014 (0.193)	0.015 (0.188)
Prop. size $n_j = 0.4N_j$	-0.163 (0.212)	-0.453 (0.200)	0.018 (0.190)	0.021 (0.183)
Fixed size $n_j = 9$	-0.214 (0.204)	-0.512 (0.190)	-0.062 (0.258)	0.000 (0.185)
Prop. size $n_j = 0.1N_j$	-0.164 (0.220)	-0.450 (0.209)	-0.074 (0.294)	0.008 (0.203)
Informative only at cluster level (level 2)				
Fixed size $n_j = 38$	0.281 (0.169)	0.018 (0.168)	0.017 (0.170)	0.017 (0.169)
Prop. size $n_j = 0.4N_j$	0.274 (0.169)	0.014 (0.178)	0.014 (0.182)	0.014 (0.181)
Fixed size $n_j = 9$	0.274 (0.187)	0.010 (0.195)	0.010 (0.212)	0.009 (0.196)
Prop. size $n_j = 0.1N_j$	0.269 (0.179)	0.007 (0.179)	0.007 (0.203)	0.006 (0.182)
Non-informative				
Fixed size $n_j = 38$	0.000 (0.108)	0.000 (0.114)	0.001 (0.115)	0.001 (0.115)
Prop. size $n_j = 0.4N_j$	0.003 (0.113)	0.004 (0.120)	0.003 (0.123)	0.003 (0.122)
Fixed size $n_j = 9$	-0.007 (0.108)	-0.009 (0.115)	-0.010 (0.125)	-0.010 (0.117)
Prop. size $n_j = 0.1N_j$	-0.002 (0.110)	-0.002 (0.114)	-0.004 (0.132)	-0.003 (0.117)

Table 2
Simulation Means and Standard Deviations (in parenthesis) of Point Estimators of the Second Level Standard Deviation
(true value 0.632, number of replicates 1,000)

Sampling design	Unweighted estimator	Cluster-level weighted	Weighted estimators	
			Unscaled fully weighted	Scaled fully weighted
Informative at both levels				
Fixed size $n_j = 38$	0.671 (0.106)	0.638 (0.112)	0.637 (0.137)	0.604 (0.128)
Prop. size $n_j = 0.4N_j$	0.673 (0.108)	0.636 (0.112)	0.645 (0.142)	0.592 (0.130)
Fixed size $n_j = 9$	0.644 (0.145)	0.584 (0.172)	0.920 (0.289)	0.536 (0.222)
Prop. size $n_j = 0.1N_j$	0.598 (0.164)	0.546 (0.183)	1.002 (0.317)	0.498 (0.242)
Informative only at cluster level (level 2)				
Fixed size $n_j = 38$	0.595 (0.100)	0.596 (0.110)	0.605 (0.111)	0.601 (0.111)
Prop. size $n_j = 0.4N_j$	0.582 (0.096)	0.582 (0.115)	0.603 (0.113)	0.596 (0.113)
Fixed size $n_j = 9$	0.547 (0.121)	0.548 (0.135)	0.671 (0.144)	0.563 (0.133)
Prop. size $n_j = 0.1N_j$	0.538 (0.122)	0.535 (0.142)	0.696 (0.158)	0.551 (0.139)
Non-informative				
Fixed size $n_j = 38$	0.611 (0.086)	0.612 (0.092)	0.621 (0.090)	0.617 (0.091)
Prop. size $n_j = 0.4N_j$	0.609 (0.084)	0.606 (0.088)	0.626 (0.088)	0.618 (0.088)
Fixed size $n_j = 9$	0.561 (0.105)	0.561 (0.112)	0.685 (0.119)	0.575 (0.111)
Prop. size $n_j = 0.1N_j$	0.551 (0.109)	0.546 (0.113)	0.703 (0.134)	0.559 (0.112)

Estimation of ω_{σ} .

The results in Table 2, concerning ω_{σ} , are more difficult to interpret (Table 2: rows 1–8, columns 2–4). First note that also in the base scenario the estimation of ω_{σ} is biased, especially in small samples. Therefore the weight-based adjustment should be judged as effective if it is able to reproduce the same bias which is observed in the base

scenario. On these grounds the behavior of the scaled fully weighted estimator is satisfactory in nearly all situations, with the exception of the small samples when the design is informative at both levels. In that case there is also a not negligible number of replications which yielded a zero estimate for ω_{σ} (4.5% for the design with fixed size and 2% for the design with proportional size). The unscaled fully

weighted estimator does not suffer from the problem of null estimates, but, apart from having a larger variance than the scaled version, tends to overestimate ω_σ , showing a relative bias of about 50% in small samples when the design is informative at both levels. Note also that the scaled fully weighted estimator outperforms the cluster-level weighted estimator even when the design is informative only at level 2.

4.2.4 Additional Simulations Using the model with Covariates

Some additional simulations were conducted to assess the performance of the scaled fully weighted estimator in the model with one cluster-level covariate and one subject-level covariate. The model is the same used in the main simulation plan, except for the inclusion of a covariate at each hierarchical level. For each covariate the values are generated from a standard Gaussian distribution, while the corresponding regression coefficient is fixed to 0.1.

As shown by Tables 3 and 4, the scaled fully weighted estimator is effective in removing the bias induced by the informative design. Relative to the unweighted estimator the sampling variance is higher, especially for the subject-level regression coefficient. Overall, the performance of the weighted estimator is satisfactory.

Table 3

Simulation Means and Standard Deviations (in parenthesis) of Point Estimators of the Regression Coefficient of the Subject-Level Covariate (true value 0.1, number of replicates 1,000)

Sampling design	Non informative	Informative at both levels	
	Unweighted estimator	Unweighted estimator	Scaled fully weighted estimator
Fixed size $n_j = 38$	0.101 (0.028)	0.117 (0.040)	0.098 (0.050)
Prop. size $n_j = 0.4N_j$	0.099 (0.026)	0.117 (0.043)	0.098 (0.052)
Fixed size $n_j = 9$	0.099 (0.055)	0.119 (0.083)	0.100 (0.104)
Prop. size $n_j = 0.1N_j$	0.098 (0.056)	0.116 (0.089)	0.098 (0.107)

Table 4

Simulation Means and Standard Deviations (in parenthesis) of Point Estimators of the Regression Coefficient of the Cluster-Level Covariate (true value 0.1, number of replicates 1,000)

Sampling design	Non informative	Informative at both levels	
	Unweighted estimator	Unweighted estimator	Scaled fully weighted estimator
Fixed size $n_j = 38$	0.096 (0.119)	0.117 (0.130)	0.102 (0.142)
Prop. size $n_j = 0.4N_j$	0.102 (0.110)	0.106 (0.133)	0.106 (0.142)
Fixed size $n_j = 9$	0.094 (0.117)	0.116 (0.141)	0.105 (0.150)
Prop. size $n_j = 0.1N_j$	0.094 (0.119)	0.115 (0.144)	0.095 (0.158)

4.2.5 General Remarks

Our simulations showed that the PML approach is, in most cases, a simple and effective strategy to deal with informative sampling designs. The only requirement is the knowledge of the inclusion probabilities at every stage of the sampling process (except when the informativeness does not concern all the levels).

As for the regression parameters, the scaled version of the fully weighted estimator showed good performance in our simulations, achieving a low bias with a modest increase in the sampling variance (in some cases the variance even diminished). Even when weighting is superfluous, the loss of efficiency due to the inclusion of scaled weights is very low.

While for the estimation of the regression parameters weighting seems to be always effective, for the variance component ω_σ attention should be paid to the sample size: in fact, weighting leads to satisfactory results only when the cluster size is high, *i.e.*, when it allows a good representation of the complex variance structure. However the sample size is crucial in the estimation of ω_σ also when all the basic assumptions of the multilevel ordinal model are satisfied.

The differences induced by the type of clusters in the sample, fixed or variable size, are minimal, with equal sized clusters leading to slightly better estimators; however, as already noted, the important differences are largely due to the average size of the clusters in the sample.

The results of our simulation study confirm the findings of Pfeffermann *et al.* (1998) on the random intercept linear model: probability-weighted estimators are good for the intercept, while some relevant bias remains in the estimation of the variance components when the sample is small. As was to be expected, when passing from a linear to a nonlinear model the performance of the estimators slightly worsen, but the direction and importance of the bias in the various cases are similar. Also the advantages of scaling are confirmed.

The rise in the sampling variance due to the inclusion of the weights often has a magnitude which is in line with the results of Pfeffermann *et al.* (1998), though in some cases we found a reduction in the sampling variance, notably for the intercept when the weights are scaled and the design is informative at both levels. An interesting difference with respect to Pfeffermann *et al.* (1998) is the role of scaling in reducing the sampling variance: in this respect, scaling seems to be more effective in the binary model than in the linear model.

As already noted, the critical point in the random intercept binary model is the estimation of the cluster-level variance ω_σ , which represents a difficult task also when the

design is non-informative. Using the threshold formulation outlined in section 2, ω_σ is defined as ω/σ , so estimation of ω_σ involves the problems observed in the linear model associated with estimation of the two variance components. The simulations showed that the performance of the scaled weighted estimator of ω_σ is not entirely satisfactory in the case of small sample sizes. A possible way to improve the performance of the estimator is the adoption of a different scaling method. Korn and Graubard (2003) investigated the issue of scaling in the context of the linear model and warned that the scaling method here adopted ('scaling method 2' of Pfeiffermann *et al.* 1998) may be badly biased under some designs, even if the sample size of clusters and sample sizes within the clusters are large. To get an idea of the extent of the bias we performed a short simulation study under the unfavorable scenario outlined by Korn and Graubard (2003), namely a simple random sample of clusters whose population sizes are all equal, and a simple random sample of individuals within each sampled cluster that is of size $2m$ or $m/2$ for a fixed m , depending on whether the observed variability of the individuals within the clusters tends to be large or small, respectively. In this case the scaled weights at subject level are all equal to 1, so weighting becomes ineffective. As a consequence, in the linear variance component model the within variance will be biased high. To see how this behavior extends to the random intercept binary model we simulation 1,000 datasets with 80 clusters and cluster sizes of 36 or 9 depending on whether the binomial variance of the responses of the cluster is over or under the median, respectively. Under the same super population model as in the main simulations, the simulation means (and standard deviations) are -0.003 (0.098) for β_σ and 0.451 (0.144) for ω_σ . The cluster-level variance is heavily underestimated, though its value is not so far from the worst case of the main simulations (0.498 under the informative design with $n_j = 0.1N_j$). Therefore, it seems unlikely to encounter situations where the bias is much greater than already shown by our simulations. Obviously, if estimation of the variance components is of primary interest it is important to improve the method, but this requires further research.

4.2.6 Bootstrap Variance Estimation

The estimated covariance matrix of the parameter estimates obtained by inversion of the information matrix, yielded by default by the NLMIXED procedure, is not reliable when using the weighted estimators to adjust for an informative design. For example, the estimated standard error of the scaled fully weighted estimator under the design informative at both levels with $n_j = 0.4N_j$ is 0.109 for β_σ (compared with a Monte Carlo value of 0.183) and 0.089 for ω_σ (compared with a Monte Carlo value of 0.130). For

the other samling sizes similar downward biases arise, so an alternative variance estimator should be devised.

The bootstrap procedure described in section 3.4 has been applied to estimate the sampling standard deviations of the weighted point estimators of β_σ and ω_σ . We limited the analysis to the scaled fully weighted estimator and to designs that are informative at both levels. To save computational resources we implemented a bootstrap procedure which omits the steps concerning the elementary units, *i.e.*, only the clusters are resampled. This procedure is expected to produce sufficiently accurate results, given the low sampling fraction (35/300) of the clusters (see section 3.4). Each simulation comprises 1,000 replications. For every replication the values of the response variable are generated through the two-stage scheme described in section 4.1 and 200 bootstrap samples are selected. Table 5 reports, for each parameter, the Monte Carlo standard error of the sampling distribution of the scaled weighted estimator on 1,000 replications of the complex design (see Tables 1 and 2), the corresponding average bootstrap estimate and the relative bias.

Table 5
Simulation Standard Deviations of the Scaled Weighted Point Estimators of the Intercept and of the Second Level Standard Deviation and Corresponding Bootstrap Estimates (with 200 Bootstrap Samples Each) for Designs Informative at Both Levels (1,000 Replicates for Each Design)

Sampling design Inform. Both levels	β_σ			ω_σ		
	Simul. s.d.	Boot. Estim.	Relative error	Simul. s.d.	Boot. Estim.	Relative error
Fixed size $n_j = 38$	0.185	0.175	-5.4%	0.124	0.106	-14.5%
Prop. size $n_j = 0.4N_j$	0.183	0.173	-5.5%	0.140	0.129	-7.9%
Fixed size $n_j = 9$	0.200	0.167	-16.5%	0.234	0.599	156.0%
Prop. size $n_j = 0.1N_j$	0.195	0.173	-11.3%	0.247	0.538	117.8%

Due to the extremely long computational time, we limited our experiment to a specific bootstrap procedure based on only 200 bootstrap samples. Further work is needed to calibrate the number of bootstrap samples and to explore possible variants of the method. Nonetheless, the entries of Table 5 give some hints about the behavior of bootstrap estimators.

The performance is better for the estimation of the sampling standard deviation of the estimator of β_σ , rather than of ω_σ . Especially for ω_σ the sample size is the critical factor: for small cluster sizes ($n_j = 9$ and $n_j = 0.1N_j$) the bootstrap estimate is completely unreliable. On the contrary with large cluster sizes ($n_j = 38$ and $n_j = 0.4N_j$) the results are quite good, since for both β_σ and ω_σ the bootstrap produces a slight underestimation of the true variance. Note, however, that the bad performance of the variance

estimator for ω_{σ} is not as critical since Wald tests for variance parameters are not generally recommended in ordinary situations anyway.

5. Final Remarks

The wide use of multilevel ordinal and binary models in many fields of application has motivated our study on the effects of complex sampling designs on the fitting of such models. In the paper we showed, by means of simulations, the bias induced by a two-stage complex sampling design on the fitting of a simple random intercept binary model when the clusters and/or the subjects are selected with probabilities that depend on the model's random terms. The simulation study also showed that in such situations the bias can be reduced in an effective manner by the probability-weighted estimation procedure (PML) described in the paper, which is easily implemented in the SAS environment. In particular, the scaled version of the weighted estimator achieved, for both fixed and random parameters, a low bias with a modest increase in the sampling variance. Even when weighting is superfluous, the loss of efficiency due to the inclusion of scaled weights seems to be very low.

The application of the proposed methodology to real life examples requires an operational strategy which depends on the extent of the available information on the sampling design. Two extreme cases can be envisaged: a) for each stage of the sampling plan, the probabilities of inclusion and the adjustments for poststratification and nonresponse are exactly known; b) the information is limited to the final overall weights, which also include adjustments for post-stratification and nonresponse.

In case a) the weights can be calculated at each sampling stage as the reciprocals of the product of sample selection probabilities and response probabilities given the sample selection, with a further correction for a possible postratification. This is the idea behind the real life application presented in Pfefferman *et al.* (1998).

In case b) the lack of information is critical, since, even in the absence of nonresponse and poststratification, it is not possible to disentangle the cluster-level and the (conditional) subject-level weights, at least without strong assumptions. As a result, weighted estimation cannot be performed.

Between the two extreme cases just outlined there are many possible intermediate situations which require *ad hoc* solutions. For example, a common case arises when the researcher has access to the cluster-level inclusion probabilities (π_j) and to the final overall subject-level weights (w_{ij}), which also include adjustments for postratification and nonresponse. When the poststratification and non-response affect only the subject level, then the subject-level

(conditional) weights can be calculated as $w_{i|j}^* = w_{ij} \cdot \pi_j$. Another more complex situation is described by Korn and Graubard (2003).

A drawback of probability-weighted estimation is the need for special procedures to estimate the variability of the estimators. In our application we adopted a bootstrap technique, which is conceptually simple and easy to program, but requires some computational effort. Our limited simulation study suggests that its performance is good only for large sample cluster sizes; however more simulations would be needed to fully understand the behavior of the bootstrap estimator.

Another open question is the choice of the most effective scaling method for reducing the bias of the estimator of the variance components when the sample size is small.

The PML approach described in the paper is absolutely general and the estimation technique based on the NLMIXED procedure of SAS is easy to generalize to other nonlinear models. Therefore it would be of interest to assess the performance of the method in models other than the random intercept binary model here considered.

Appendix A

We report the SAS code used for implementing the probability-weighted (PML) estimators described in the paper. The essential part of the code is the NLMIXED procedure of SAS, which is a general procedure for fitting nonlinear random effects models using adaptive Gaussian quadrature. Though the NLMIXED procedure does not include an option for PML estimation, it is still possible to insert the weights in the likelihood, using different tricks for level 1 and level 2 weights. To insert level 1 weights it is necessary to exploit the option which allows to write down the expression for the conditional likelihood of the model: then one should simply translate in SAS programming statements the expression $w_{i|j} \log L_{ij}(\theta|u)$ (see section 3.1). On the other hand, level 2 weights can be inserted in the likelihood through the `replicate` statement. Unfortunately, this statement is limited to integer weights, so to avoid gross approximations it is advisable to proceed as follows: a) inflate all the level 2 weights by an arbitrary constant k (equal to 10,000 in our application); b) insert the integer part of the inflated weights in the likelihood through the `replicate` statement; c) multiply the estimated covariance matrix by k by means of the `cfactor` option. This trick relies on the fact that multiplying the level 2 weights by a constant has the only effect of inflating the information matrix by that constant, leaving the estimates unchanged. Anyway, when using the weighted estimation

method to adjust for an informative design the estimated covariance matrix of the parameter estimates is not reliable.

In the following the SAS code is reported, where the symbols/* and */ include the comments:

```
proc nlmixed data=dataname qpoints=10
cfactor=10,000;
/* cfactor is a constant multiplying the
estimated covariance matrix of the parameter
estimates */
parms b0=0 sd=0.5; /* initial values */
bounds sd >= 0;
eta=b0+randeff*sd;
if (yobs=1) then z=probnorm(eta);
else if (yobs=0) then z=1-probnorm(eta);
if (z >1e-8) then ll=log(z); else ll=-1e100;
/*to avoid numerical problems if z becomes
too small*/
ll=ll*w1_2;/*inclusion of level 1 weights */
model yobs~general(ll);
random randeff ~normal(0,1) subject=j;
/* j is the cluster identifier */
replicate w2; /* inclusion of level 2 weights
(only integers) */
ods output ParameterEstimates=pe
ConvergenceStatus=cs;
run;
```

Acknowledgements

We wish to thank two anonymous referees for their suggestions which contributed to substantially improve the paper.

References

- Gibbons, R.D., and Hedeker, D. (1997). Random effects probit and logistic regression models for three-level data. *Biometrics*, 53, 1527-1537.
- Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73, 43-56.
- Graubard, B.I., and Korn, E.L. (1996). Modelling the sampling design in the analysis of health surveys. *Statistical Methods in Medical Research*, 5, 263-281.
- Hedeker, D., and Gibbons, R.D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics*, 50, 933-944.
- Korn, E.L., and Graubard, B.I. (2003). Estimating variance components by using survey data. *Journal of the Royal Statistical Society B*, 65(1), 175-190.
- Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review*, 61(2), 317-337.
- Pfeffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H. and Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society B*, 60(1), 23-40.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G. and Draper, D. (1999). A users guide to MLwiN. London: Multilevel models project, Institute of Education, University of London.
- Renard, D., and Molenberghs, G. (2002). Multilevel modeling of complex survey data. In *Topics in Modelling of Clustered Data* (Ed. M. Aerts M). London: Chapman and Hall, 263-272.
- Särndal, C.-E., Swensson, B. and Wretman, J. (1992). Model assisted Survey Sampling. New York: Springer-Verlag.
- SAS Institute (1999). SAS/STAT User's Guide Version 8. Cary: SAS Institute Inc.
- Shao, J., and Tu, D. (1995). *The Jackknife and the Bootstrap*. New York: Springer-Verlag.
- Singh, A.C., Folsom, R.E. and Vaish, A.K. (2002). A hierarchical Bayes generalization of the Fay-Herriot method to unit level nonlinear mixed models for small area estimation. *Proceedings of the Section on Survey Research Methods, American Statistical Association*. New York, August 10-13. 3258-3263.
- Skinner, C.J. (1989). Domain means, regression and multivariate analysis. In *Analysis of Complex Surveys*, (Eds. C.J. Skinner, D. Holt and T.M.F. Smith). Chichester: Wiley, 59-87.