**(Entry for *Encyclopaedia of Biostatistics,* Armitage, P., & Colton,T. (Eds.), 1998, Wiley.)**

# Multilevel Models

**by**

Harvey Goldstein

University of London

h.goldstein@ioe.ac.uk

## Introduction

Biostatistical data often have a hierarchical structure. Typically these structures are naturally occurring ones: animal populations are characterised by individuals nested within parents, themselves often nested within groups or herds which may also be nested within spatial entities. In other cases the structure may result from research designs, as in multi-centre clinical trials where patients are nested within clinics. In yet other cases, the data may not obviously seem to be nested, yet viewing it as such may yield new insights or more efficient analysis techniques. Examples are repeated measures designs where measurements are ‘nested’ within individual subjects and multivariate response data where measurements are ’nested’ within individuals.

In addition to nesting relationships among data units we may also have cross-classifications. For example, an individual cow may be nested within a herd of cattle, but also be the offspring of parent stock where any parent may contribute to several herds: individual cows are thus cross classified by parents as well as nested within their herds. A further complexity is also often present whereby individual units at one level of a data hierarchy may be nested within more than one higher level unit. An example is spatial data, where each individual person can be classified by the geographical locality where they live, but will also be influenced in terms, say, of their health or behaviour, by surrounding localities. In this case we regard them as belonging to a primary unit plus a number of secondary units.

In the following sections I shall develop a set of models for describing such data, increasing in complexity as they move from simple hierarchies with continuously distributed responses, to cross classifications and multivariate data and to discrete responses. Various extensions and special cases will also be considered. The emphasis is on model specification rather than estimation, although there is a brief section on the latter.

## The basic multilevel model

For simplicity consider a simple data structure where an outcome is measured on patients in a number of centres, together with one or more treatments or covariates. We wish to model a relationship between the outcome and the explanatory variables, taking into account the possibility that this relationship may vary across centres. We shall refer to the centres as higher level units and patients as lower level units. In the present case we just have two levels with centres as level 2 units and patients as level 1 units. A simple such model can be written as follows

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1 x_{ij} + u_{0j} + e_{ij} \\ \operatorname{var}(e_{ij}) &= \sigma^2_{e0} \\ \operatorname{var}(u_j) &= \sigma^2_{u0} \end{aligned} \qquad (1)$$

where $y_{ij}$ is the response and $x_{ij}$ the value of a single explanatory variable for the *i*-th patient in the *j*-th centre. The slope coefficient $\beta_1$ is for the present assumed to be the same at all the centres while the random variable $u_{oj}$ represents the departure of the *j*-th clinic's intercept from the overall population intercept term $\beta_0$. The first two terms on the right hand side of (1) constitute the fixed part of the model and the last two terms describe the random variation. We shall develop the model initially assuming that the random variables have a (multivariate) Normal distribution, and discuss the non-Normal case later. This model could be viewed as a standard analysis of covariance if we treated each $u_{oj}$ as a fixed parameter to be estimated. Such a model however will often be inappropriate, for the following reasons.

First, we may have a very large number of centres, leading to a very large number of separate parameters to estimate. Secondly, some of the clinics may have very few patients, so that their individual departures will be poorly estimated. Most importantly, we may be interested in treating the centres as a sample from a *population* of centres and wish to make general inferences about the likely behaviour of other centres in this population rather than, or in addition to, providing separate estimates for each centre in the sample. For all these reasons it will usually be more appropriate to regard $u_{0j}$ as random and to write

$$u_{0j} \sim N(0,\sigma_{u0}^2) \qquad e_{0ij} \sim N(0,\sigma_{e0}^2)$$

We can also elaborate (1) by allowing the coefficient $\beta_1$ to vary across centres and rewrite the model in the more compact form

$$\begin{aligned}
&y_{ij} = \beta_{0ij}x_0 + \beta_{1j}x_{1ij} \\
&\beta_{0ij} = \beta_0 + u_{0j} + e_{ij} \\
&\beta_{1j} = \beta_1 + u_{1j} \qquad (2) \\
&U = \{u_{0j}u_{1j}\} \quad E(U) = 0, \quad \operatorname{cov}(U) = \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix}, \quad \operatorname{var}(e_{ij}) = \sigma_e^2
\end{aligned}$$

This model is often referred to as a ‘random coefficient model’ by virtue of the fact that the coefficients $\beta_{0ij}$ and $\beta_{1j}$ in the first equation of (2) are random quantities. It is possible, however, to have random coefficient models which are only single level (see below), so that we shall drop this term in order to emphasise the hierarchical data structure.

As more explanatory variables are introduced into the model, so we can choose to make them random at the centre level thereby introducing covariances as well as variances at level 2, and this will lead to models with complex covariance structures. One of the aims of multilevel modelling is to explore such potential structures and also to attempt to explain them in terms of further variables. Having fitted such a model we can obtain posterior estimates for the individual 'residuals' $(u_{0j}, u_{1j}, e_{0ij})$ at either level by estimating their expected values (or other functions of their distributions), given the data and model estimates. Thus, for example, we can estimate $E(u_{0j}|Y, \beta, \theta)$ where

$$\beta^T = \{\beta_1, \beta_2\} \quad \theta = \{\sigma^2_{u0}, \sigma_{u01}\sigma^2_{u1}, \sigma^2_{e0}\} \tag{3}$$

The multilevel model is here described in non-Bayesian terms. For a full Bayesian specification of this model we would need to add prior distribution assumptions for the parameters in (3). The interested reader is referred, for example, to Gilks et al [3] for details with examples.

In the next section we shall look at a general formulation and then some important special cases. A fully detailed treatment of the topics is not possible here and the reader is referred to Longford [9] and Goldstein [4] for details of methodology with examples and a discussion of computer software. A World Wide Web site has been set up which contains information about current developments, references, etc. at *http://www.ioe.ac.uk/multilevel/* , and it is intended that this will be further developed.

## Cross Classifications

Many data structures are not purely hierarchical, but mixtures of hierarchies and cross classifications. For example, in a school health survey children may be assessed by raters each school having just one rater. Thus we have a structure where children are grouped within cells defined by the cross classification of raters by schools and we will wish to model the level 2 variation as a function of both the between-rater and between-school variation. If the design were changed, so that a separate team of raters visited each school and each child was measured by a single rater, then the cross classification would be that of raters by children nested within schools. If, again, there was a single team of raters who visited every school the cross classification would be of raters by children across the whole sample. In this case we have no separable hierarchy and we would wish to model the total response variation as a function of the between-child, between-school and between-rater variation.

Rasbash and Goldstein [12] discuss various examples of this kind and set out the appropriate models together with procedures for efficient estimation.

Corresponding to the first and second examples given above we can write the following models, using a more general notation for the fixed part of the model, where $i$ indexes children, $j_1$ indexes schools, and $j_2$ indexes raters. We write

$$y_{i(j_1j_2)} = X_{i(j_1j_2)}\beta + u_{j_1} + u_{j_2} + e_{i(j_1j_2)} \qquad (4)$$

for the first model with children nested within the level 2 cross classification and with the following level 2 covariance structure

$$\begin{aligned}
&\operatorname{cov}(y_{i(j_1j_2)}y_{i'(j_1j'_2)}) = \sigma^2_{u_1} \\
&\operatorname{cov}(y_{i(j_1j_2)}y_{i'(j'_1j_2)}) = \sigma^2_{u_2} \\
&\operatorname{var}(y_{i(j_1j_2)}) = \operatorname{cov}(y_{i(j_1j_2)}y_{i'(j_1j_2)}) = \sigma^2_{u_1} + \sigma^2_{u_2}
\end{aligned} \qquad (5)$$

The second model is written as

$$y_{(i_1i_2)j} = X_{(i_1i_2)j}\beta + u_j + e_{i_1j} + e_{i_2j}$$

(6)

In both (4) and (6) we have assumed an ‘additive’ model for the variance contributions and the adequacy of this can be tested against a model which includes an interaction term, for example

$$y_{(i_1i_2)j} = X_{(i_1i_2)j}\beta + u_j + e_{i_1j} + e_{i_2j} + e_{(i_1i_2)j} \qquad (7)$$

In addition, we can have further random coefficients and levels of nesting or crossing.

## Multiple unit membership

We have assumed so far that each lower level unit, such as a school student or patient, belongs to just one higher level unit of a particular kind. In many cases, however, such units may belong to more than one higher level unit. For example, in a child growth study children may change schools from one occasion to the next, and a particular case is that of spatial data where an individual is influenced by the geographical unit where they live and also (with differing weights) by neighbouring areas.  We can write a simple 2-level model of this kind as follows where, for simplicity, we suppose the maximum number of level 2 units to which a level 1 unit may belong is two.

$$
\begin{aligned}
&y_{i(j_1 j_2)} = X_{i(j_1 j_2)}\beta + w_{1ij_1}u_{j_1} + w_{2ij_2}u_{j_2} + e_{i(j_1 j_2)} \\
&w_{1ij_1} + w_{2ij_2} = 1
\end{aligned}
$$

$$
\begin{aligned}
&\operatorname{var}(y_{i(j_1 j_2)}) = (w_{1ij_1}^2 + w_{2ij_2}^2)\sigma_u^2 + \sigma_e^2 \\
&\operatorname{cov}(y_{i(j_1 j_2)} y_{i'(j_1 j_2)}) = (w_{1ij_1} w_{1i'j_1} + w_{2ij_2} w_{2i'j_2})\sigma_u^2 \qquad (8) \\
&\operatorname{cov}(y_{i(j_1 j_2)} y_{i'(j_1' j_2)}) = w_{2ij_2} w_{2i'j_2} \sigma_u^2
\end{aligned}
$$

As before, we can further elaborate this model by allowing random coefficients, further hierarchical levels and further crossing factors. For example, in the example of children changing schools we may cross classify the schools by the neighbourhoods where the children live with the possibility of multiple neighbourhood membership in the above sense and across  time.

## Repeated measures data and multivariate data

An interesting special case of a 2-level structure is that of repeated measures models such as the following

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + e_{ij} \qquad (9)$$

where the response, say, is the weight of an animal related to a linear function of age (*x*) with the intercept and slope varying across animals. See 'Random coefficient repeated measures models' for further information.

Another important special case is that of multivariate data where the response is a vector. Consider first a 'single level' multivariate linear model, with two responses, height and weight, measured on a sample of males and females. For the *j*'th variable (*j* = 0 for height, *j* = 1 for weight) measured on the *i*'th subject we have the model equation

$$\begin{gathered} y_{ij} = \beta_{01} z_{1ij} + \beta_{02} z_{2ij} + \beta_{11} z_{1ij} x_j + \beta_{12} z_{2ij} x_j + u_{1j} + u_{2j} \\ z_{1ij} = \begin{Bmatrix} 1 \text{ if height} \\ 0 \text{ if weight} \end{Bmatrix}, \quad z_{2ij} = 1 - z_{1ij}, \quad x_j = \begin{Bmatrix} 1 \text{ if female} \\ 0 \text{ if male} \end{Bmatrix} \qquad (10) \\ \operatorname{var}(u_{1j}) = \sigma_{u1}^2, \quad \operatorname{var}(u_{2j}) = \sigma_{u2}^2, \quad \operatorname{cov}(u_{1j} u_{2j}) = \sigma_{u12} \end{gathered}$$

A part of the data matrix for this structure might be

|  |  | Intercepts (z) |  | Gender (x) |
|---|---|---|---|---|
| Individual | Response | Height | Weight |  |
| 1 (female) | $y_{11}$ | 1 | 0 | 1 |
| 1 | $y_{12}$ | 0 | 1 | 1 |
| 2 (male) | $y_{21}$ | 1 | 0 | 0 |
| 2 | $y_{22}$ | 0 | 1 | 0 |
| 3 (female) | $y_{31}$ | 1 | 0 | 1 |

so that at level 2 we have the variances and covariance of height and weight while there is no variation at level 1 and the fixed part of the model is defined using the relevant dummy variables associated with each response. Notice that in the data matrix, the third individual has no weight measurement. By specifying the multivariate model as in (10) we can implicitly fit data where some responses are missing: we simply omit the relevant level one unit corresponding to the missing observation. The model can be generalised readily in the ways already discussed, by allowing random coefficients, cross classifications etc. and further levels of nesting. An example of a multivariate model analysis will be given later.

## Modelling variances

In addition to specifying the average response as modelled in the fixed part of the model, we have discussed modelling the covariance structure at level 2 (and higher levels) by introducing random coefficients. We may also introduce random coefficients which vary across level 1 units and this provides a flexible general procedure for variance modelling. Consider the following model

$$\begin{aligned} &y_{ij} = \beta_0 + \beta_1 x_{ij} + (u_j + e_{0ij} + e_{1ij} x_{ij}), \\ &\operatorname{var}(e_{0ij}) = \sigma^2_{e0}, \quad \operatorname{var}(e_{1ij}) = 0, \quad \operatorname{cov}(e_{0ij} e_{1ij}) = \sigma_{e01} \end{aligned} \tag{11}$$

so that the level 1 contribution to the overall variance is the linear function

$$\sigma^2_{e0} + 2\sigma_{e01} x_{ij} .$$

Note that we have constrained one of the ‘variances’ at level 1 to be zero in order to give a linear rather than a quadratic variance function. In fact the parameters $\sigma^2_{e0}, \sigma^2_{e1}, \sigma_{e01}$ are not to be interpreted as separate variances and covariances, but simply as parameters defining the variance structure. The variable ($x$) may be any kind of explanatory variable. For example, if it were a dummy variable for gender the model would allow a separate level 1 variance for males and females. In this way it is possible to model the variance, as well as the mean, as functions of explanatory variables. Examples are given in Goldstein [4, Chapter 3].

In some circumstances, linear models for a variance such as implied by (11) are inappropriate because they may predict an overall level 1 variance which is negative for part of the range. In this case we can consider alternative models where the level 1 variance has the form

$$\mathrm{var}(e_{ij}) = \exp(\beta_0^* - \beta_1^* x_{ij}) \qquad (12)$$

which is non-negative and where we require estimates of the $\beta_0^*, \beta_1^*$. Goldstein [4, Appendix 5.1] shows how maximum likelihood estimates for such models can be obtained.

## Nonlinear and generalised linear models

We can write a 2-level generalised linear model in the form

$$\pi_{ij} = f(X_{ij}\beta_j) \qquad (13)$$

where $\pi_{ij}$ is the expected value of the response for the *ij*-th level 1 unit and *f* is a nonlinear function of the 'linear predictor' $X_{ij}\beta_j$ where we can have random coefficients at level 2. We need to specify a distribution for the *observed* response $y_{ij}|\pi_{ij}$: where the response is a proportion this is typically taken to be binomial and where the response is a count taken to be Poisson. Equation (13) is a special case of a nonlinear model which is completed by specifying a suitable link function f(). Thus, for binary response data we might have a simple model

$$\mathrm{logit}(\pi_{ij}) = \beta_0 + \beta_1 x_{1ij} + u_{0j} \qquad (14)$$

$$y_{ij} \sim Bin(\pi_{ij}, 1)$$

with a corresponding model for counts using a log link function. The random part of (14) can be elaborated with further random coefficients, cross classifications etc.

These models can be extended to multinomial (ordered or unordered) responses [4, Chapter 7].

## Survival models

Survival time data will often have a multilevel structure: for example we may measure illness durations within centres or waiting times in hospitals with variation across centres and hospitals. We may also have repeated duration episodes *within individuals* for example repeated periods of disease and remission, where different kinds of episode also may exist. See 'Survival data models' for more information. We briefly mention here three common types of model and their multilevel specification. Further details are given by Goldstein [4, Chapter 9].

The first type is the extension of the semiparametric Cox model, often referred to as a frailty model. When defining risk sets for this model we can choose to order our failure times across the whole data set or *within* level 2 units, say hospitals. In the former case the marginal relationship between the hazard and the covariates is not generally proportional, and in the latter case it is proportional within level 2 units.

At each failure time $l$ we define a response variate for each member of the risk set

$$y_{ijk(l)} = \begin{cases} 1 \ if \ i \ is \ the \ observed \ failure \\ 0 \ if \ not \end{cases}$$

where $i$ indexes the members of the risk set, and $j,k$ level 1 and level 2 units. The response is treated as a Poisson variate with mean function for a simple variance components model given by

$$\pi_{jk(l)} = \exp(\alpha_l + X_{jk}\beta + u_k) \tag{15}$$

where there is a 'blocking factor' $\alpha_l$ for each failure time.

The second type of model is a 'log duration' or 'accelerated failure time' model which can be written as

$$l_{ij} = \ln(t_{ij}) = X_{ij}\beta_j + e_{ij} \tag{16}$$

for the failure times $t_{ij}$. This is in the standard form for a 2-level random coefficient model. A complication is that we may have (level 1) censored observations, and this implies that we need a careful specification of the level 1 distribution to incorporate censoring information in the estimation. Some common choices are the Normal, extreme value and log-gamma distributions.

The third type of model, which leads to a particularly simple form, is the discrete time proportional hazards model. For a 2-level model we write

$$\log\{-\log(1-\pi_{jk(l)})\} = X_{jk}\beta_k + \alpha_{(l)} \tag{17}$$

where, as before, the $\alpha_{(l)}$ are constants to be estimated, one for each time interval. This leads to a model where the response is a binomial variate, being the number of deaths divided by the number in the risk set at the start of the interval. As with the first type, any censored observations in an interval are excluded from the risk set.

## Estimation

The basic model assumes multivariate Normality and standard (as well as restricted) maximum likelihood methods are available using Fisher scoring, iterative generalised least squares or the EM algorithm. Bayesian estimation is available using Gibbs sampling (Gilks et al, [3]) which is also available for generalised linear models with the appropriate distributional assumptions. An alternative in this case is to use quasi-likelihood estimation together with appropriate bias correction procedures (Goldstein and Rasbash, [5]), or the related GEE procedure (Liang et al, [8]). For inference, interval estimates are obtained directly from Gibbs sampling and via large sample deviance statistics or bootstrapping for likelihood estimation.

## An example

To illustrate the flexibility of multilevel models we fit a bivariate 2-level model where one response is Normal and the other is binary.

The data are part of the ‘Health and Lifestyle Survey’, a sample of 9003 individuals within households nested within 396 electoral wards in Britain and carried out in 1984/85. For present purposes data on smoking habits is analysed using information about gender and age. Further details are given by Duncan et al. [2]. The information about smoking behaviour consists of whether or not the respondent smoked and if they did how many per day. Sixty five percent did not smoke and the mean number smoked for those who did is 15.2 with a standard deviation of 9.3. The distribution of the number smoked is positively skewed which suggests a Normalising transformation. The use of this, however, does not substantially alter the results and the analysis is presented in terms of the actual number smoked.

One aim of the analysis is to ascertain how the probability of smoking and the number smoked each relate to the explanatory variables. The other is to estimate the between-area variation and in particular to see whether areas where the proportion of non smokers is high are also the areas where smokers tend to smoke greater numbers of cigarettes. We write the model as

$$
\begin{aligned}
y_{ijk} = \quad & \delta_i [\{1+\exp[-(X_1\beta_1)_{ijk} - u_{1k}]\}^{-1} + e_{1jk}] \\
& + (1-\delta_i)[(X_2\beta_2)_{ijk} + u_{2k} + e_{2jk}] \\
\delta_i = 1 & \ \mathit{if\ binary},\ 0\ \mathit{if\ continuous}
\end{aligned}
\qquad (18)
$$

where $u_{1k}, u_{2k}$ respectively refer to the ward level contributions to the discrete and continuous parts of the model. The $e_{1jk}, e_{2jk}$ are similarly defined for the variation among individuals. This model combines a model for smokers where the response is the number of cigarettes smoked and a model with a binary response which is whether or not the subject smoked. Thus, each smoker will have two responses, a ‘1’ for the binary response variable and the number smoked for the continuous response. Each non-smoker will have just one response, a ‘0’ indicating that they are a non-smoker.

This model has been fitted with the MLn software package (Rasbash and Woodhouse, [13]) using specially written macros. The bivariate structure is modelled as level 1, where there is no random variation, so that the full model is 3-level.

**Bivariate model for smoking/non-smoking and number smoked. Gender is coded 1 for male and 0 for female: age is measured about the mean of 45.9 years. The level 1 variance is constrained to 1.0 which corresponds to binomial variation.**

| **Parameter:** | **Response** | |
|---|---|---|
| ***Fixed*** | Binary (s.e.) | Continuous (s.e.) |
| Intercept | -0.54 | 15.7 |
| Gender | 0.14 (0.05) | 2.82 (0.32) |
| Age | -0.03 (0.03) | 1.22 (0.21) |
| $(\text{Age})^2$ | 0.0011 (0.0007) | -0.02 (0.005) |
| $(\text{Age})^3$ | -0.000012 (0.000005) | 0.00009 (0.00003) |
| ***Random*** | | |
| Level 2: | | |
| Intercept variance | 0.17 (0.03) | 1.45 (0.81) |
| covariance | 0.40 (0.11) | |
| Level 1: | | |
| Intercept | 1.0 | 79.2 (2.1) |

At the electoral ward level there is a high correlation (0.81) between the proportion of smokers and the number smoked. Men are more likely to be smokers and to smoke more and there is an age effect for the number smoked, with a maximum among 50 year olds, and declining thereafter. The relationship is weaker for the probability of smoking. A model which allowed gender to have a random coefficient at level 2 was fitted but a large sample test for the extra variance and two covariance terms gave a chi squared value of 6.8 on 3 degrees of freedom (P=0.08). Attempting to fit the age coefficient as random at level 2 produces a zero estimated variance. We can also test the assumption of binomial variation for the smoking response by fitting extra binomial variation in the form of an estimated rather than a theoretical unit variance at level 1. This is estimated as 0.98 with a standard error of 0.015, providing little evidence of extra binomial variation.

## Further topics

Finally we mention briefly some further topics, most of which are currently the subject of methodological research.

The standard *meta analysis* model can be viewed as a special case of a general multilevel model. For the *j-th* study in such an analysis we can define the standardised effect $d_j$ where this is a dimensionless quantity. It may, for example, be a correlation coefficient, a standardised regression coefficient, group difference, or weighted group difference. We can write a simple model as follows

$$d_j = \delta + v_j + u_j, \qquad \mathrm{var}(u_j) = \sigma_j^2, \qquad \mathrm{var}(v_j) = \sigma_v^2 \tag{19}$$

where in the usual case $\sigma_j^2$ is assumed known and is treated as an offset in the random part of the model, but may also in some circumstances be estimated. The parameter $\delta$ is the population parameter of interest and $\sigma_v^2$ is the between study (level 2) variance of the standardised effect. We can add random coefficients and covariates representing study factors to (19) in an attempt to explain between-study differences which is a further aim of meta analysis studies.

As in single level models, diagnostics are important. We can estimate standardised residuals at any level of a data hierarchy and study these together with looking for influential units. A detailed discussion is given by Lewis and Langford [7].

Further important issues are those concerned with missing units and missing data generally, especially where the missingness is informative, and research is being conducted in this area. Another topic which is actively being researched is that of multilevel structural equation modelling (McDonald and Goldstein, [10], Muthen, [11]).

## Software

Some of the major software packages, for example SAS and GENSTAT, can handle many, although not all of the models described in this article. Several general purpose software packages have been written; HLM (Bryk and Raudenbush, [1]), VARCL (Longford, [9]) and MLn (Rasbash and Woodhouse, [13]). A review of these packages has been carried out by Kreft et al [6].

## References

1. Bryk, A. S. and S. W. Raudenbush (1992). *Hierarchical Linear Models*. Newbury Park, California, Sage.

2. Duncan, C., K. Jones, and Moon, G. (1996). Health-related behaviour in context: a multilevel modelling approach. *Social Science and Medicine*. 42, 817-30

3. Gilks, W., Richardson, S. And Spiegelhalter, D. (1996) *Markov Chain Monte Carlo in Practice*. London, Chapman and Hall.

4. Goldstein, H. (1995). *Multilevel Statistical Models*. London, Edward Arnold. New York, Halsted Press.

5. Goldstein, H. and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. J. Royal Statistical Society, A, 159, (to appear)

6. Kreft, G. G., J. DeLeeuw and R. VanderLeeden (1994). Review of five multilevel analysis programs. *The American Statistician*, **48**: 324-335.

7. Lewis, T. and Langford, I. (1996). Outliers in multilevel data. (Submitted for publication).

8. Liang, K-Y., Zeger, S.L. and Qaqish, B. (1992) Multivariate regression analyses for categorical data (with discussion). *Journal of the Royal Statistical Society*, B, 54, 3-40.

9. Longford, N. T. (1993). *Random Coefficient Models.* Oxford, Clarendon Press.

10. McDonald, R. P. and H. Goldstein (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of mathematical and statistical psychology,* **42**: 215-232.

11. Muthen, B. O. (1994). Multilevel covariance structure analysis. *Sociological methods & research,* **22**: 376-398.

12. Rasbash, J. and H. Goldstein (1994). Efficient analysis of mixed hierarchical and cross classified random structures using a multilevel model. *Journal of Educational and Behavioural statistics,* **19**: 337-50.

13. Rasbash, J. and Woodhouse, G. (1995). *MLn Command Reference*. London, Institute of Education.