STATISTICS IN MEDICINE
Statist. Med. **18**, 855–888 (1999)

TUTORIAL IN BIOSTATISTICS
AN INTRODUCTION TO HIERARCHICAL LINEAR MODELLING

LISA M. SULLIVAN[1]*, KIMBERLY A. DUKES[2] AND ELENA LOSINA[1]

[1] *Boston University School of Public Health, Department of Epidemiology and Biostatistics, 715 Albany Street, Boston, MA 02115, U.S.A.*
[2] *DM-STAT, Inc., 459 Broadway, Suite 204, Everett, MA 02149, U.S.A.*

SUMMARY

Hierarchical linear models are useful for understanding relationships in hierarchical data structures, such as patients within hospitals or physicians within hospitals. In this tutorial we provide an introduction to the technique in general terms, and then specify model notation and assumptions in detail. We describe estimation techniques and hypothesis testing procedures for the three types of parameters involved in hierarchical linear models: fixed effects, covariance components, and random effects. We illustrate the application using an example from the Type II Diabetes Patient Outcomes Research Team (PORT) study and use two popular PC-based statistical computing packages, HLM/2L and SAS Proc Mixed, to perform two-level hierarchical analysis. We compare output from the two packages applied to our example data as well as to simulated data. We elaborate on model interpretation and provide guidelines for model checking.

1. INTRODUCTION

Medical research applications often involve hierarchical data structures such as patients within hospitals or physicians within hospitals; for example, assessing differences in mortality rates across hospitals relative to a specific condition or procedure. Data are collected on random samples of patients nested within each hospital. In this application, it might be appropriate to adjust for covariates at both the patient-level (such as patient age, patient gender and the severity of the index diagnosis) and at the hospital-level (such as hospital size and hospital teaching status).

Hierarchical linear models, sometimes called multi-level linear models, nested models, mixed linear models or covariance components models, handle these hierarchical data structures. These models have historically been used in educational research where hierarchies occur naturally; students nested within classrooms, classrooms nested within schools and schools nested within

* Correspondence to: Lisa M. Sullivan, Boston University School of Public Health, Department of Epidemiology and Biostatistics, 715 Albany Street, Boston, MA 02115, U.S.A.

CCC 0277–6715/99/070855–34$17.50

Received January 1998
Accepted August 1998

districts. Recent advances in statistical computing capabilities have made these models more available to researchers across a variety of disciplines.

In this tutorial we provide an introduction to the technique relative to two-level hierarchical data structures. We provide references for readers interested in three-level structures. In Section 2 we motivate the application with an example and we illustrate the application using two popular statistical computing packages, HLM/2L[1] and SAS Proc Mixed.[2] In Section 3 we present notation, specify models in detail and discuss assumptions. In Section 4 we describe estimation techniques and hypothesis testing procedures. In Section 5 we provide data handling and programming statements to develop and test hierarchical linear models using HLM/2L and SAS Proc Mixed. In Section 6 we present results of analyses based on data from our example and based on simulated data. In Section 7 we provide a brief summary.

2. MOTIVATION

To illustrate the technique we have selected an example from the Type II Diabetes Patient Outcomes Research Team (PORT) study.[3] The PORT study was a 5-year longitudinal observational study of medical effectiveness in Type II diabetes. Three sites, representing diverse systems of care, were involved in this study. Stratified random samples of primary care physicians were selected at each site. Stratification criteria included age (< 45 versus $\geqslant 45$ years), gender and medical speciality (Internal Medicine, Family Practice, Endocrinology). Samples of Type II diabetic patients were selected from each of 224 enrolled physician's practices. A total of 3660 patients were enrolled.

Upon enrollment, each physician completed a self-administered physician background questionnaire. Physicians provided data related to their management styles, interpersonal care, job and practice satisfaction, personal background and training characteristics in the self-administered questionnaire. Patients completed self-administered questionnaires at 6-month intervals over the course of the study. Patients provided extensive data at baseline related to general health status and quality of life, diabetes history, medical complications and comorbidities, health habits and personal characteristics, compliance and regimen adherence, diabetes-specific health status, satisfaction with medical care, assessments of their physician's interpersonal style, health care utilization data and beliefs about health care. As part of the PORT study, utilization (for example, hospitalizations, primary care visits), laboratory (for example, glycated haemoglobin, total cholesterol) and pharmacy (prescriptions and over-the-counter medications) data were also collected on each patient from automated management information systems at each site. In our example we restrict our attention to physician and patient data collected at baseline from one site.

The objective of this analysis is to assess whether there is a significant difference in mean patient outcome scores across physicians adjusted for appropriate covariates. The specific outcome variable in the analysis is described in detail in Section 6. At this point, we simply wish to describe the structure of the data. Potentially important covariates include both patient characteristics (for example, patient's age, patient's gender) and physician characteristics (for example, physician's years in medical practice, physician's medical speciality). To appropriately model patient-level and physician-level covariates simultaneously, we develop a two-level hierarchical model. We describe the model development strategy, estimation techniques and hypothesis testing procedures in the following sections.

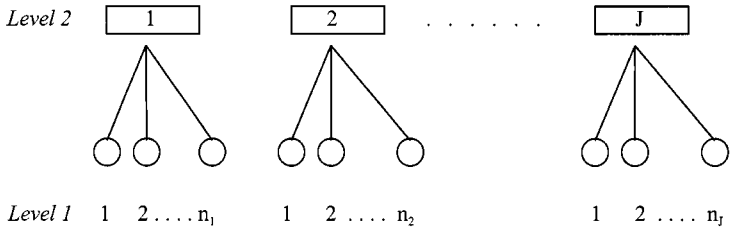

Figure 1. Data structure for the two-level hierarchical model

3. NOTATION

We consider two-level hierarchical data structures and follow the notation of Bryk and Raudenbush[4] who developed HLM/2L. For details of three-level models see Bryk and Raudenbush,[4] Gatsonis *et al.*[5] and Skene and Wakefield.[6]

Similar to our example (the PORT study, restricting attention to one site) consider an application involving a two-stage sampling strategy. In the simplest two-stage sampling strategy a simple random sample of primary sampling units (PSUs) is selected and then, within each PSU, a simple random sample of secondary sampling units (SSUs) is selected. The PSUs could reflect hospitals, physicians, clinics or some other entity and the SSUs could reflect physicians (within hospitals) patients (within physician practices or within clinics) or some other entity. In the PORT study, the PSU was the physician and the SSU was the patient at each site. In two-level hierarchical analyses, observations are classified as level 1 (within) or level 2 (between) units. In the PORT study, patients represent the level 1 units and physicians represent the level 2 units.

We use J to denote the number of level 2 units (in the PORT study, physicians) and within each of the $j = 1, \ldots, J$ level 2 units (physician practices) there are n_j level 1 units (in the PORT study, there are n_j patients in the jth physician practice). The data do not have to be balanced (that is, it is not necessary that $n_j = n_k$ for $j \neq k$). Figure 1 displays the data structure for the two-level hierarchical model.

3.1. Model Specifications

In two-level hierarchical models, separate level 1 models are developed for each of the J level 2 units. Consider the case of a continuous outcome or dependent variable, Y (for example, patient satisfaction), and a single, continuous level 1 predictor or covariate X (for example, patient's age). The *level* 1 *models* are of the form:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij} \tag{1}$$

where Y_{ij} is the dependent variable measured on the ith level 1 unit (for example, patient) nested within the jth level 2 unit (for example, physician), β_{0j} is the intercept for the jth level 2 unit (physician), X_{ij} is the level 1 predictor or covariate (for example, patient age), $\bar{X}..$ is the grand mean of X_{ij} (for example, the mean age of all patients in the sample), β_{1j} is the regression coefficient associated with level 1 predictor X for the jth level 2 unit (physician) and ε_{ij} is the random error associated with the ith level 1 unit nested within the jth level 2 unit.

The interpretation of estimated model parameters (particularly the intercept terms, $\hat{\beta}_{0j}$) depends upon the way in which the level 1 covariates or predictors are modelled. If the level 1 predictor, X_{ij},

is considered in its original metric, then the intercept, β_{0j}, is the expected value of the dependent variable when X_{ij} is equal to zero. If the level 1 predictor, X_{ij}, is centred about its overall (or grand) mean ($\bar{X}..$), then the intercept, β_{0j}, is interpreted as the mean of the jth level 2 unit adjusted for X (for example, the mean outcome score for physician j adjusted for patient age). If the level 1 predictor is centred at the respective level 2 unit mean, $\bar{X}._j$, called centring at the group mean, then the intercept, β_{0j}, is interpreted as the unadjusted mean of the dependent variable. Finally, if the level 1 predictor is centred at some other meaningful value, then the intercept, β_{0j}, is interpreted as the expected value of the dependent variable when the predictor is equal to that value.

In the case of a dichotomous dependent or outcome variable, Y, and a single, continuous level 1 predictor or covariate, X, the level 1 models are of the form:

$$p_{ij} = 1/\{1 + \exp(\beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij})^{-1}\}$$

where $p_{ij} = E(Y_{ij})$.

Details of hierarchical models concerning dichotomous dependent variables can be found in Wong and Mason[7] and details of hierarchical models concerning count or rate data can be found in Christiansen and Morris.[8]

Here we focus on applications concerning continuous dependent variables, Y, which are assumed to be approximately normally distributed at each value of a single, continuous covariate, X, as described in (1). In these applications we have J models of the form shown in (1), each model potentially having different intercept and slope coefficients (β_{0j}, β_{1j}). In the level 2 models, we consider these regression coefficients (β_{0j} and β_{1j}) as dependent variables and relate each to appropriate level 2 covariates. In the case of a single, continuous level 2 predictor or covariate, W (for example, physician's years in medical practice), the *level 2 models* are of the form:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} W_j + \upsilon_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} W_j + \upsilon_{1j} \end{aligned} \tag{2}$$

where β_{0j} and β_{1j} are the intercept and slope for the jth level 2 unit (physician), γ_{00} and γ_{10} are the overall mean intercept and slope adjusted for W, respectively, W_j is the level 2 predictor or covariate (for example, physician's years in medical practice), γ_{01} and γ_{11} are the regression coefficients associated with the level 2 predictor W relative to the level 2 intercepts and slopes, respectively and υ_{0j} and υ_{1j} are the random effects of the jth level 2 unit (physician) on the intercept and slope, respectively, adjusted for W.

The level 2 predictor W can be modelled in its original metric (as in (2)) or centred about its grand mean (similar to the way in which the level 1 predictor is modelled in (1)).

Substituting (2) into (1) yields the *combined model*:

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10}(X_{ij} - \bar{X}..) + \gamma_{11} W_j(X_{ij} - \bar{X}..) + \upsilon_{0j} + \upsilon_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij}. \tag{3}$$

Notice that the combined model involves level 1 and level 2 covariates (X_{ij} and W_j, respectively), a cross-level term ($W_j(X_{ij} - \bar{X}..)$) and a complicated error term, $\upsilon_{0j} + \upsilon_{1j}((X_{ij} - \bar{X}..) + \varepsilon_{ij}$. Model (3) is not of a form in which ordinary least squares (OLS) can be used to estimate parameters, since OLS assumes errors are independent with mean zero and common variance σ^2. In model (3), the errors are not independent across level 1 units (patients), instead there is dependency among level 1 units nested within each level 2 unit (physician practice) in the terms

υ_{0j} and υ_{1j}. In addition, the variances of the errors may no longer be homogeneous if υ_{0j} and υ_{1j} take on different values within each level 2 unit (for example, physician practice). We describe estimation and hypothesis testing techniques which handle these dependencies in Section 4.

In Table I we present six distinct models, essentially special cases of the models (1)–(3). These models are presented in Bryk and Raudenbush[4] and we employ their notation and labelling conventions as they are also employed in HLM/2L. For each model we provide the level 1 model, the level 2 model, the combined model, and indicate which effects are fixed and which effects are random. The models are ordered from least to most complex. Models 1–4 are called random intercepts models, while models 5 and 6 are called random intercepts and slopes models. In Table II we present the same set of models using the notation employed in SAS Proc Mixed. SAS Proc Mixed uses a single model (analogous to the HLM/2L combined model).

3.2. Assumptions

For models concerning continuous dependent variables (Y_{ij} of (1)), we assume that the errors in the level 1 (for example patient-level) models are normal random variables with mean zero and common variance σ^2:

$$E(\varepsilon_{ij}) = 0 \quad \text{var}(\varepsilon_{ij}) = \sigma^2. \tag{4}$$

In the level 2 (for example, physician-level) models we assume that the parameters β_{0j} and β_{1j} are distributed as multivariate normal with means γ_{00} and γ_{10}, respectively, and variances τ_{00} and τ_{11}, respectively. The covariance of β_{0j} *and* β_{1j} is denoted τ_{01}. For simplicity, we consider the situation in which the errors are homogeneous at both levels 1 and 2, although more complicated error structures are allowed (see reference 9 pp. 583–586 for a variety of covariance structures with examples). Finally, level 1 and level 2 errors are uncorrelated. These assumptions are summarized below:

$$\begin{gathered} E(\upsilon_{0j}) = 0 \quad E(\upsilon_{1j}) = 0 \\ E(\beta_{0j}) = \gamma_{00} \quad E(\beta_{1j}) = \gamma_{10} \\ \text{var}(\beta_{0j}) = \text{var}(\upsilon_{0j}) = \tau_{00} \quad \text{var}(\beta_{1j}) = \text{var}(\upsilon_{1j}) = \tau_{11} \\ \text{cov}(\beta_{0j}, \beta_{1j}) = \text{cov}(\upsilon_{0j}, \upsilon_{1j}) = \tau_{01} \\ \text{cov}(\upsilon_{0j}, \varepsilon_{ij}) = \text{cov}(\upsilon_{1j}, \varepsilon_{ij}) = 0. \end{gathered} \tag{5}$$

4. ESTIMATION AND HYPOTHESIS TESTING IN HLM/2L AND IN SAS PROC MIXED

We now describe the estimation techniques and hypothesis testing procedures for two-level hierarchical models (3) used in HLM/2L and SAS Proc Mixed. We illustrate the estimation and testing procedures using matrix notation. Because HLM/2L and SAS Proc Mixed use somewhat different parameterizations (see Tables I and II), we first present notation which we use throughout this section.

Consider the level 1, level 2 and combined models (using HLM/2L notation) shown in (1), (2) and (3), respectively. Using matrix notation these models are represented as follows:

$$\text{HLM/2L level 1 model (1):} \qquad Y_j = X_j\beta_j + \varepsilon_j, \qquad j = 1, 2, \ldots, J \tag{6}$$

Table I. Hierarchical linear model specifications in HLM/2L

Model	Level 1 model	Level 2 model	Combined model	Fixed effects	Random effects
Random intercepts models					
1. One-way analysis of variance (ANOVA) with random effects	$Y_{ij} = \beta_{0j} + \varepsilon_{ij}$	$\beta_{0j} + \gamma_{00} + \upsilon_{0j}$	$Y_{ij} = \gamma_{00} + \upsilon_{0j} + \varepsilon_{ij}$	γ_{00}	υ_{0j}
2. Means as outcomes regression model	$Y_{ij} = \beta_{0j} + \varepsilon_{ij}$	$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \upsilon_{0j}$	$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \upsilon_{0j} + \varepsilon_{ij}$	γ_{00}, γ_{01}	υ_{0j}
3. One-way analysis of covariance (ANCOVA) with random effects	$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij}$	$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \upsilon_{0j}$ $\beta_{1j} = \gamma_{10}$	$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10}(X_{ij} - \bar{X}..) + \upsilon_{0j} + \varepsilon_{ij}$	$\gamma_{00}, \gamma_{01}, \gamma_{10}$	υ_{0j}
4. Non-randomly varying slopes model	$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij}$	$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \upsilon_{0j}$ $\beta_{1j} = \gamma_{10} + \gamma_{11} W_{ij}$	$Y_j = \gamma_{00} + \gamma_{01} W_j + \gamma_{10}(X_{ij} - \bar{X}..) + \gamma_{11} W_j(X_{ij} - \bar{X}..) + \upsilon_{0j} + \varepsilon_{ij}$	γ_{00}, γ_{01} γ_{10}, γ_{11}	υ_{0j}
Random intercepts and slopes models					
5. Random coefficients regression model	$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij}$	$\beta_{0j} = \gamma_{00} + \upsilon_{0j}$ $\beta_{1j} = \gamma_{10} + \upsilon_{1j}$	$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}..) + \upsilon_{0j} + \upsilon_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij}$	γ_{00}, γ_{10}	$\upsilon_{0j}, \upsilon_{1j}$
6. Intercepts and slopes as outcomes model	$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij}$	$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \upsilon_{0j}$ $\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + \upsilon_{1j}$	$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10}(X_{ij} - \bar{X}..) + \gamma_{11} W_j(X_{ij} - \bar{X}..) + \upsilon_{0j} + \upsilon_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij}$	γ_{00}, γ_{01} γ_{10}, γ_{11}	$\upsilon_{0j}, \upsilon_{1j}$

Table II. Hierarchical linear model specifications in HLM/2L and SAS Proc MIXED*

Model	Combined model (HLM/2L from Table I)	SAS Proc MIXED		
		Model specification	Fixed effects	Random effects
Random intercepts models				
1. One-way analysis of variance (ANOVA) with random effects	$Y_{ij} = \gamma_{00} + \upsilon_{0j} + \varepsilon_{ij}$	$Y_{1j} = \mu + \alpha_j + \varepsilon_{ij}$	μ	α_j
2. Means as outcomes regression model	$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \upsilon_{0j} + \varepsilon_{ij}$	$Y_{ij} = \mu + \alpha_j + \beta W_j + \varepsilon_{ij}$	μ, β	α_j
3. One-way analysis of covariance (ANCOVA) with random effects	$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10}(X_{ij} - X..) + \upsilon_{0j} + \varepsilon_{ij}$	$Y_{ij} = \mu + \alpha_j + \beta W_j + \delta(X_{ij} - \bar{X}..) + \varepsilon_{ij}$	μ, β, δ	α_j
4. Non-randomly varying slopes model	$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10}(X_{ij} - \bar{X}..) + \gamma_{11} W_j(X_{ij} - \bar{X}..) + \upsilon_{0j} + \varepsilon_{ij}$	$Y_{ij} = \mu + \alpha_j + \beta W_j + \delta(X_{ij} - \bar{X}..) + \beta\delta W_j(X_{ij} - \bar{X}..) + \varepsilon_{ij}$	μ, β, δ	α_j
Random intercepts and slopes models				
5. Random coefficients regression model	$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}..) + \upsilon_{0j} + \upsilon_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij}$	$Y_{ij} = \mu + \alpha_j + (\delta + \delta_j)(X_{ij} - \bar{X}..) + \varepsilon_{ij}$	μ, δ	α_j, δ_j
6. Intercepts and slopes as outcomes model	$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10}(X_{ij} - \bar{X}..) + \gamma_{11} W_j(X_{ij} - \bar{X}..) + \upsilon_{0j} + \upsilon_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij}$	$Y_{ij} = \mu + \alpha_j + \beta W_j + (\delta + \delta_j)(X_{ij} - \bar{X}..) + \beta\delta W_j(X_{ij} - \bar{X}..) + \varepsilon_{ij}$	μ, β, δ	α_j, δ_j

*SAS Proc Mixed source code for each model is provided in the Appendix

where

$$Y_j = \begin{bmatrix} Y_{ij} \\ Y_{2j} \\ . \\ . \\ Y_{n_j j} \end{bmatrix}, \qquad X_j = \begin{bmatrix} 1 & X_{ij} \\ 1 & X_{2j} \\ . & \\ . & \\ 1 & X_{n_j j} \end{bmatrix}, \quad \beta_j = \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix}, \quad \varepsilon_j = \begin{bmatrix} \varepsilon_{ij} \\ \varepsilon_{2j} \\ . \\ . \\ \varepsilon_{n_j j} \end{bmatrix}.$$

$$\text{HLM/2L level 2 model (2): } \beta_j = W_j\gamma + \upsilon_j, \qquad j = 1, 2, \ldots, J \tag{7}$$

where

$$\beta_j = \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix}, \quad W_j = \begin{bmatrix} 1 & W_j & 0 & 0 \\ 0 & 0 & 1 & W_j \end{bmatrix}, \quad \gamma = \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{bmatrix}, \quad \upsilon_j = \begin{bmatrix} \upsilon_{0j} \\ \upsilon_{1j} \end{bmatrix}.$$

$$\text{HLM/2L combined model (3): } \qquad Y_j = X_j(W_j\gamma + \upsilon_j) + \varepsilon_j$$

which is equivalent to

$$Y_j = X_j W_j \gamma + X_j \upsilon_j + \varepsilon_j \quad j = 1, 2, \ldots, J. \tag{8}$$

Note that the level 1 and level 2 covariates, X and W, respectively, can be considered in their original metric (as in (6), (7) and (8)) or centred (as in (1)).

The equivalent *SAS Model* is

$$Y_j = A_j\gamma + X_j\upsilon_j + \varepsilon_j, \quad j = 1, 2, \ldots, J \tag{9}$$

where $A_j = X_j W_j$, as follows:

$$A_j = \begin{bmatrix} 1 & W_j & X_{1j} & W_j X_{1j} \\ 1 & W_j & X_{2j} & W_j X_{2j} \\ . & & & \\ . & & & \\ 1 & W_j & X_{n_j j} & W_j X_{n_j j} \end{bmatrix}.$$

A_j and X_j are known design matrices, γ is a vector of fixed effects, υ_j is a vector of random effects, and ε_j is a vector of random errors. The assumptions of hierarchical linear models (outlined in (4) and (5)) are as follows:

$$\begin{aligned} &\varepsilon_j \sim \mathrm{N}(0, R), \quad R_j = \sigma^2 I_{nj} \\ &\upsilon_j \sim \mathrm{N}(0, G), \quad G = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix}. \end{aligned} \tag{10}$$

4.1. Estimation

Several estimation techniques are used in hierarchical linear modelling since the model comprises different types of parameters. Specifically, the level 1 coefficients, β_j, can be fixed (that is, equal to

a constant: β_{1j} of model 3 in Table I), non-randomly varying (that is, vary across level 2 units, for example, physicians, but solely as a function of a level 2 predictor W_j: β_{1j} of model 4 in Table I) or random (that is, vary across level 2 units, for example, physicians: β_{0j} *of models* 1–6, β_{1j} of models 5 and 6 in Table I). The level 2 coefficients, γ, are considered fixed effects and the level 1 and 2 variances and covariances (σ^2, τ_{00}, τ_{01} and τ_{11}) are called the covariance components. The estimation techniques for each type of paramter are outlined below. More theoretical details are available in Littell *et al.*[2] and in Searle *et al.*[10]

4.1.1. Estimating Fixed Effects (γ)

Weighted least squares (WLS) or generalized least squares (GLS) is used to estimate γ as shown below:

$$\hat{\gamma} = (A^{\mathrm{T}} \hat{V}^{-1} A)^{-1} A^{\mathrm{T}} \hat{V}^{-1} Y \tag{11}$$

where

$$V = \mathrm{var}(Y) = XGX^{\mathrm{T}} + R \tag{12}$$

A is the $N \times 4$ design matrix with $N = \sum_{j=1}^{J} n_j$ (see (9)), and $\hat{V}$ is V with G and R replaced by their maximum likelihood estimates. The elements of G and R (that is, τ_{00}, τ_{01}, τ_{11} and σ^2) are called the variance-covariance components and are estimated by maximum likelihood (ML) or restricted maximum likelihood (REML) as described in Section 4.1.2.

The variance of the estimator $\hat{\gamma}$ (11) is estimated by

$$\hat{\mathrm{var}}(\hat{\gamma}) = (A^{\mathrm{T}} \hat{V}^{-1} A)^{-1}. \tag{13}$$

Liang and Zeger[11] recommended the following as an alternative to (13), particularly in the case when the variances of Y are not homogeneous across level 2 units (for example, physicians):

$$\hat{\mathrm{var}}(\hat{\gamma}) = (A^{\mathrm{T}} \hat{V}^{-1} A)^{-1} A^{\mathrm{T}} \hat{V}^{-1} (Y - A\hat{\gamma})(Y - A\hat{\gamma})^{\mathrm{T}} \hat{V}^{-1} A (A^{\mathrm{T}} \hat{V}^{-1} A)^{-1}. \tag{14}$$

4.1.2. Estimating Covariance Components (R *and* G)

If the design is perfectly balanced (that is, n_j all equal and the distribution of level 1 predictors within each level 2 unit, (for example, within each physician practice, is the same) there are closed-form formulae for estimating the variance-covariance parameters.[10] When the design is unbalanced, iterative numerical procedures are used to obtain the estimates. Usually these procedures are based on maximum likelihood estimation techniques.

Maximum likelihood (ML) estimates of G and R are found by maximizing the following log-likelihood function (see also Littell *et al.*[2] and Searle *et al.*[10]):

$$l_{\mathrm{ML}}(G, R) = -\frac{1}{2} \log |V| - \frac{N}{2} \log r^{\mathrm{T}} V^{-1} r - \frac{N}{2} \left[1 + \log \frac{2\pi}{N} \right] \tag{15}$$

where

$$r = Y - A(A^{\mathrm{T}} V^{-1} A)^{\mathrm{T}} A^{\mathrm{T}} V^{-1} Y.$$

If the number of level 2 units, J is large then the estimates generated through maximum likelihood are approximately equal to estimates generated through restricted maximum

likelihood (REML). REML estimates of the covariance components are based on residuals which are computed after estimating the fixed effects (11) by WLS or by GLS and are estimates based on maximizing a marginal likelihood. REML estimates take into account the degrees of freedom used in estimating the fixed effects when estimating the covariance components. REML estimates of G and R are found by maximizing the following log-likelihood function (see also Littell *et al.*[2] and Searle *et al.*[10]):

$$I_{\mathrm{REML}}(G,R) = -\frac{1}{2}\log|V| - \frac{1}{2}\log|A^{\mathrm{T}}V^{-1}A| - \frac{(N-p)}{2}\log r^{\mathrm{T}}V^{-1}r - \frac{(N-p)}{2}\left[1+\log\frac{2\pi}{(N-p)}\right] \tag{16}$$

where $r = Y - A(A^{\mathrm{T}}V^{-1}A)^{\mathrm{T}}A^{\mathrm{T}}V^{-1}Y$ and $p = \mathrm{rank}(A)$.

HLM/2L generates REML estimates by default and uses the EM algorithm to maximize (16). SAS Proc Mixed also produces REML estimates by default and uses a ridge-stabilized Newton–Raphson algorithm to maximize the likelihood.[2] Maximum likelihood estimates can be requested in both HLM/2L and SAS Proc Mixed.

4.1.3. Estimating Random Effects (υ)

Random effects are estimated using shrinkage estimators. SAS Proc Mixed generates estimates of random effects according to the following:

$$\hat{\upsilon} = \hat{G}X^{\mathrm{T}}\hat{V}^{-1}(Y - A\hat{\gamma}). \tag{17}$$

It is generally of interest to estimate individual (for example, physician-level) random coefficients (for example, $\hat{\beta}^{*}_{0j}$ and $\hat{\beta}^{*}_{1j}$). These can be obtained by substitution. For example, considering the model described in (2), $\hat{\beta}^{*}_{0j} = \hat{\gamma}_{00} + \hat{\gamma}_{01}W_j + \hat{\upsilon}_{0j}$, where $\hat{\upsilon}_{0j}$ is from (17). A shrinkage estimate (also referred to as an empirical Bayes (EB) estimate in HLM/2L or a best linear unbiased prediction (BLUP) in SAS Proc Mixed[2]) for the jth random coefficient (for example, β_{0j} of model 1 in Table I) is essentially an optimally weighted, linear combination of the estimated overall mean (for example, $\hat{\gamma}_{00}$) and the jth level 2 mean ($\bar{Y}_j$). The degree of shrinkage depends on the magnitude of the variation in level 2 means (which is related to the number of level 1 units within the specific level 2 unit used to generate the estimate). Thus, when n_j, the number of level 1 units within the jth level 2 unit, is small, the estimate of the jth random coefficient is close to the overall mean $\hat{\gamma}_{00}$, but, as an n_j increases, the estimate of the jth random coefficient moves closer to the level 2 mean $\bar{Y}_j$. BLUPs are seen as estimates subject to regression toward the overall mean ($\hat{\gamma}$) based on the covariance components of model effects (for example, $\hat{G}$, $\hat{V}$). For more details, see also Littell *et al.*[2] and Zeger *et al.*[12] Empirical Bayes estimates are computationally efficient and results are asymptotically equivalent to Bayes solutions.[13]

4.2. Hypothesis Testing

Several hypothesis tests are generally of interest in hierarchical models. The tests are implemented in slightly different ways in HLM/2L and in SAS Proc Mixed. In Section 6 we illustrate the specific tests produced by HLM/2L and SAS Proc Mixed using data from the PORT study and simulated data. Here we outline general testing procedures for fixed effects, covariance components and random effects. In general, t-statistics are produced for the fixed and random effects

while Wald Z and chi-square statistics are produced for the covariance components. These tests, particularly the tests for covariance components, are valid only asymptotically and therefore must be interpreted with caution, especially when the number of level 2 units (for example, physicians), J, is small.

4.2.1. *Hypothesis Tests for Fixed Effects* (γ)

The hypothesis of interest is of the form $H_0: \gamma_k = 0$. The test statistic is computed by taking the ratio of the ML (or REML) estimate to its estimated standard error as follows:

$$t = \frac{\hat{\gamma}_k}{\sqrt{\{\hat{\text{var}}(\hat{\gamma}_k)\}}}. \tag{18}$$

The above follows a t-distribution for balanced data and for some unbalanced data situations. In most cases, (18) only approximations a t-distribution and its degrees of freedom are estimated.[9]

4.2.2. *Hypothesis Tests for Covariance Components* (R *and* G)

The hypothesis of interest is of the form $H_0: \tau_{kl} = 0$, where τ_{kl} is an element of G. The test statistic is computed by taking the ratio of the ML (or REML) estimate to its asymptotic standard error as follows:

$$Z = \frac{\hat{\tau}_{kl}}{\sqrt{\{\hat{\text{var}}(\hat{\tau}_{kl})\}}}. \tag{19}$$

The asymptotic standard errors are computed from the second derivative of the likelihood with respect to covariance components.[2] SAS produces the Wald statistic (19) which is valid for large samples, while HLM/2L produces a chi-square statistic:

$$\chi^2_{J-1} = \sum_{j=1}^{J} \frac{(\hat{\beta}_{0j} - \hat{\gamma}_{00})^2}{\hat{V}_j} \tag{20}$$

where

$$\hat{V}_j = \frac{\hat{\sigma}^2}{n_j}.$$

4.2.3. *Hypothesis Tests for Random Effects* (υ)

The hypothesis of interest is of the form $H_0: \upsilon_k = 0$. The test statistic is computed by taking the ratio of the estimated random effect $\hat{\upsilon}_k$ to its estimated standard errors as follows:

$$t = \frac{\hat{\upsilon}_k}{\sqrt{\{\widehat{\text{var}}\,(\hat{\upsilon}_k)\}}}. \tag{21}$$

The above follows a t-distribution for balanced data and for some unbalanced data situations. In most cases, (21) only approximates at t-distribution and its degrees of freedom are estimated.[9] The estimated standard error $\sqrt{\{\widehat{\text{var}}(\hat{\upsilon}_k)\}}$, is larger under REML than ML especially if the number of level 2 units (for example, physicians), J, is small. Nevertheless, the test may be liberal (that is, reject more often than it should producing a higher type I error rate than the nominal rate) if J is small even using the REML estimates.

5. STATISTICAL SOFTWARE

There are several statistical computing packages currently available which handle hierarchical data structures. Kreft *et al.*[14] reviewed five popular packages – BMDP-5V; GENMOD; HLM; ML3, and VARCL – with respect to design, implementation, data set-up and handling, output and user frindliness. They also provide several examples in which they apply each package and compare results.

We focus on two popular packages, HLM/2L and SAS Proc Mixed, which were not as widely used in 1994. We review each package with respect to data handling, model specification and output. In Section 6 we apply each package to data from the PORT study and then to simulated data.

5.1. HLM/2L

HLM/2L is a PC-based program written by Bryk *et al.*[1] and is designed for two-level hierarchical linear modelling. The program is distributed with a companion program HLM/3L for three-level hierarchical linear modelling. An updated version of HLM/2L (and HLM/3L) for Windows, available since May 1996, significantly improved earlier versions.

5.1.1. Inputting Data

HLM/2L can be operated either interactively or in batch mode. In both modes, the input data must be summarized into a sufficient statistics matrices (SSM) file. The SSM file can be constructed by inputting ASCII data or by importing data in other database formats (for example SAS, SYSTAT). To illustrate the proper file formats, we proceed via ASCII files.

HLM/2L requires the user to prepare two separate data files, one containing level 1 data and a second containing level 2 data. Suppose that level 1 data consist of patient-level data and level 2 data consists of physician-level data. The level 1 file contains one record for each patient and the level 2 file contains one record for each physician. The level 1 (patient) file must contain, as its first column, the level 2 identifier (for example, the physician identifier). The following illustrates the format of the level 1 and level 2 files. In the level 1 (for example, patient) sample file, there are two level 1 variables not including the level 2 identifier (for example, the patient outcome or dependent variable, Y (for example satisfaction), and a patient covariate, X, such as age). In the level 2 sample file, there is one level 2 variable not including the level 2 identifier(for example physician's number of years in medical practice, W). The level 2 identifier appears in both files and is indicated in bold face below.

Level 1 file (total number of records $= \sum n_j$):

01	12	34
01	33	45
01	21	55
01	31	56
02	34	33
02	33	43
⋮	⋮	⋮
J	21	43

Level 2 file (total number of records = J).

01	10
02	12
.	.
.	.
J	6

The user is prompted to specify the formats of the level 1 and level 2 data files. The only requirement is that the level 2 identifier (appearing in the first columns of both the level 2 and level 2 data files) must be input as a character variable. The format statements for the level 1 and level 2 sample files shown above are (a2, 2x, f2.0, 2x, f2.0) and (a2, 2x, f1.0), respectively (where 'a2' indicates a 2-digit character variable, '2x' denotes 2 blank spaces between columns and 'f2.0' denotes a numeric variable occupying at most two columns with no digits following the decimal place). The user is also prompted for variable names and file locations.

Once the format of the data, variable names and the location of the ASCII data files are input into HLM/2L, an SSM file is constructed. The SSM file contains J matrices (one for each level 2 unit). Each matrix contains the number of level 1 units (n_j), the means of the level 1 variables, the sums of cross products of level 1 variables as well as the level 2 observations. Once the SSM file is constructed, subsequent run starts with the SSM file and proceed quickly. HLM/2L can handle up to 300 level 2 units ($J \leqslant 300$), 25 level 2 (for example, physician) variables, and 25 level 1 (for example, patient) variables.

5.1.2. Program Specifications

We describe the use of HLM/2L in interactive mode, as the logic of the modelling process is apparent to the user in this mode. (In batch mode, commands are stacked in a file and input. Batch mode can be more efficient for the more advanced user.) Once the program is initiated, the user is prompted to specify the name of the SSM file. The user is then prompted to choose the dependent or outcome variable, from the list of level 1 variables displayed on the screen. Next, the user is prompted to choose level 1 predictors or covariates from the list of level 1 variables displayed on the screen. Once the level 1 predictors are selected, the user is prompted to specify variables to be centred. HLM/2L offers three options: (i) centring about the grand mean ($\bar{X}..$); (ii) centring about the group (level 2) means ($\bar{X}._j$); or (iii) no centring. The user is then prompted to select level 2 predictors or covariates and offered two options for centring: (i) centring at the grand mean of the level 2 variable ($\bar{W}$); or (ii) no centring. The user is also offered options related to specific tests of hypothesis. These include, for example a test for homogeneity of level 1 variances and multivariate tests of covariance components. There are a number of other options available in the HLM/2L execution (see Bryk *et al.*[1]).

5.1.3. Output

HLM/2L generates extensive output, including the specifications of the level 1 and level 2 models using the notation in Section 3 which is extremely useful for model checking. HLM/2L outputs the starting values for parameter estimates, estimates of fixed effects along with t-statistics and significance levels for hypothesis testing, estimates of the covariance components along with chi-square statistics and significance levels. The user can request that empirical Bayes estimates of

the individual random effects be output into an ASCII file, a SAS file or a SYSTAT file for further analysis.

5.2. SAS Proc Mixed

Proc Mixed is a SAS procedure designed to handle mixed effects linear models. Detailed documentation is available in *SAS/STAT® Software: Changes and Enhancements through Release* 6.11[9] and in Littell *et al.*[2]

5.2.1. Inputting Data

Data can be input into SAS from ASCII files or from SAS data sets. Data for SAS Proc Mixed must reside in a single file containing one record for each level 1 unit. Both level 1 variables (for example, the patient-level outcome, Y, for example, satisfaction, and patient covariate age, X) and level 2 variables (for example, physician's years in medical practice, W) are contained in the *same* file. The following illustrates the proper format of the data file, using the sample files shown in Section 5.1.1. The level 2 identifier appears in the first column and is indicated in bold face. The level 2 variable, W, appears in the rightmost column.

Data file (total number of records $= \sum n_j$):

01	12	34	10
01	33	45	10
01	21	55	10
01	31	56	10
02	34	33	12
02	33	43	12
⋮	⋮	⋮	⋮
J	21	43	6

5.2.2. Program Specifications

The SAS Proc Mixed syntax is similar to the SAS Proc GLM syntax. The user specifies classification variables in the 'class' statement and the outcome or dependent variable for the analysis along with the fixed effects in the 'model' statement. The user specifies random effects in a 'random' statement and may include statements to perform repeated measures or time series analysis. SAS allows the user to specify a stratification variable in the 'by' statement. Specific tests of hypothesis for fixed or random effects can be requested using 'contrast' and 'estimate' statements. The user can specify initial or starting estimates for model parameters in the 'parms' statement. The user can also request that any part of the output be saved in a SAS data set using the 'make' option. (We provide the SAS code used in our analyses in the tables of results and in the Appendix.)

5.2.3. Output

SAS prints the output and also allows the user to save all or part of the output in a file for presentation. The output includes a description of the classification variables (for example,

number of levels and labels), estimates of the covariance components and significance tests (based on large sample theory), several statistics to assess model fit (including Akaike's information criterion and Schwarz's Bayesian criterion), estimates of the fixed effects along with *t*-statistics, and estimates of the individual random effects along with *t*-statistics and significance levels for hypothesis testing.

6. RESULTS

We now apply both HLM/2L and SAS Proc Mixed to data from the PORT study described in Section 2 and to simulated data, which allows for greater control over sources of variation. In Section 6.1 we present the results of hierarchical linear modelling analyses applied to data from the PORT study. In Section 6.2 we present our simulation models, strategy, analytic approach and results of hierarchical linear modelling on simulated data.

6.1. Analysis of PORT Study Data

In the PORT study, data were collected from physicians and patients at three sites. For this example, we restrict our attention to data collected at one site. Table III displays summary statistics on the physician ($J = 70$) and patient samples ($n = 1492$) involved in this analysis. The numbers of patients enrolled per physician ranged from 5 to 45, with a mean of 21·3 patients per physician. About one-quarter (27·1 per cent) of the physicians were female, their mean age was 43·9 years, most (92·6 per cent) were White, most were married (83·8 per cent), and they had been practising medicine for a mean of 20·5 years. About half (49·9 per cent) of the Type II diabetic patients were female, their mean age was 62·7 years, most (89·3 per cent) were White, a majority were married (70·9 per cent), almost half (47 per cent) had completed at least some college education, and about half (45·5 per cent) had an annual income of $30,000 or more.

The outcome or dependent variable in this analysis is a 14 item measure of patient satisfaction developed for the American Board of Internal Medicine's (ABIM) Patient Satisfaction Questionnaire. Sample items from the questionnaire include the following. How is your doctor at: telling you everything; letting you tell your story; listening carefully; treating you like an adult. Each item was rated on a five point Likert scale with the following response options: excellent; very good; good; fair, and poor. Item scores were aggregated into a single, composite measure and scaled to the range of 0–100 with higher scores indicative of better satisfaction. In a principal components analysis of the 14 items, a single component was produced which explained 75 per cent of the variance in the individual items. The Cronbach's alpha internal consistency reliability coefficient for the 14 items was 0·97. The mean satisfaction score (taken over all patients) was 67·7 with a standard deviation of 23·5 (see Table III). The distribution of satisfaction scores in the pooled sample ($n = 1492$) is negatively skewed as shown in Figure 2.

The objective of this analysis is to assess whether there is a significant difference in mean patient outcome scores across physicians adjusted for appropriate covariates. In this analysis we will consider one patient-level covariate and one physical-level covariate. The patient-level covariate is patient's age, since it has been shown that older patients generally report better satisfaction with medical care[15,16] and the physician-level covariate is the physician's number of years since graduation from medical school (a proxy for physician's experience in clinical practice). To appropriately model these patient-level and physician-level covariates simultaneously, we developed a two-level hierarchical model.

Table III. Description of physician and patient samples

Physician characteristics	Mean (SD)/per cent ($J = 70$)
Gender: % female	27·1
Mean age in years (SD)	43·9 (6·4)
Race: % African American	1·5
% Asian	4·4
% White	92·6
Marital status: % married	83·8
Mean years in practice (SD)	20·5 (7·1)
Mean satisfaction $\left(\sum_{j=1}^{70} \bar{Y}_{.j}/70\right)$	68·5* $\left(\sum_{j=1}^{70} s_j/70 = 22{\cdot}8^{\dagger}\right)$

Patient characteristics	Mean (SD)/per cent ($n = 1492$)
Gender: % female	49·9
Mean age in years (SD)	62·7 (11·9)
Race: % African American	4·5
% Other	6·2
% White	89·3
Marital status: % married	70·9
Education: % $\leqslant$ High school graduate	53·0
% $\leqslant$ College graduate	32·8
% Post-College graduate	14·2
Annual household income:	
% < $15,000	18·1
%$15,000–29,999	36·3
%$30,000 +	45·5
Mean satisfaction ($\bar{Y}..$)	67·7 (23·5)

* The range in physician-level means ($\bar{Y}_{.j}$) is 53·4 to 87·1
† The range in physician-level standard deviations (s_j) is 13·4 to 32·4

There are 70 primary care physicians in the analysis ($J = 70$). Physician-level mean satisfaction scores, $\bar{Y}_{.j}$, range from 53·4 to 87·1 (see Table III), and the mean of the physician-level means is 68·5. Physician-level standard deviations in satisfaction scores, which are assumed to be homogeneous, range from 13·4 to 32·4 and have a mean of 22·7.

Before fitting a two-level hierarchical model, we performed a fully stratified analysis, in which we developed separate regression equations relating patient satisfaction to patient age in each physician's practice, considered separately. The ordinary least squares (OLS) estimates of the intercepts and slopes for each physician practice are displayed in Table IV. The numbers of patients in each physician practice, n_j are also shown. Since the level 1 covariate X = patient age, was centred about its grand mean ($\bar{X}..$), the estimated intercepts ($\hat{\beta}_{0j}$) are interpreted as the mean satisfaction scores for each physician adjusted for patient age. There is substantial variation in the adjusted mean satisfaction scores among physicians which range from 54·0 to 87·9. Figure 3 displays the relationship between the OLS estimates of the intercepts and slopes. There is a very

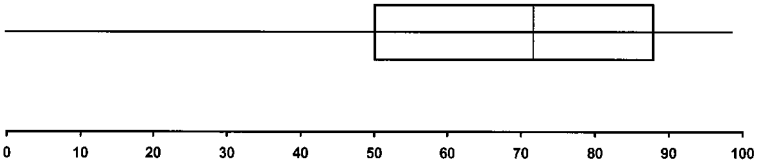

Figure 2. Distribution of patient satisfaction scores in pooled sample ($n = 1492$). Minimum = 0, $Q_1 = 50$, median = 71·4, $Q_3 = 87{\cdot}5$, maximum = 100. Mean satisfaction $\bar{Y}.. = 67{\cdot}7$, $s = 23{\cdot}5$

Table IV. Ordinary least squares (OLS) estimates of intercepts and slopes: results of fully stratified analysis

Physician identification number	Number of patients (n_j)	OLS intercept ($\hat{\beta}_{0j}$)	OLS slope ($\hat{\beta}_{1j}$)
001	21	69·6	0·07
003	25	54·0	0·21
005	28	71·4	0·32
006	45	67·9	0·05
007	16	65·1	0·10
008	11	66·6	− 0·37
009	13	65·7	− 0·42
010	34	77·1	− 0·28
011	14	73·2	− 0·22
012	17	60·1	1·13
013	27	87·2	− 0·11
015	7	70·9	− 0·55
016	25	81·8	0·06
017	5	50·8	1·52
018	30	71·4	0·01
019	15	72·8	− 0·37
021	36	60·9	1·07
022	14	62·6	− 0·43
023	25	61·8	1·88
024	16	58·7	− 1·16
026	16	70·5	0·72
028	14	81·2	0·18
029	8	87·9	0·96
030	30	64·5	− 0·24
031	28	69·1	− 0·15
032	20	58·1	− 0·24
033	19	72·6	− 0·25
034	27	58·1	− 0·34
035	11	66·0	1·00
.	.	.	.
.	.	.	.
.	.	.	.
074	29	63·3	0·05
075	31	54·0	0·67
078	22	77·3	− 0·32
080	9	66·1	− 0·63
081	23	56·8	0·84

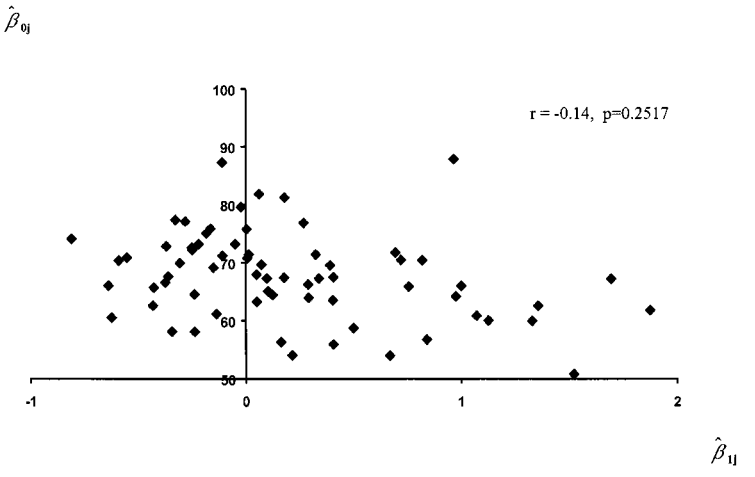

Figure 3. Relationship between OLS intercepts ($\hat{\beta}_{0j}$) and slopes ($\hat{\beta}_{1j}$)

slight negative orientation to the scatter ($r = -0{\cdot}14$, $p = 0{\cdot}2517$) indicating that higher adjusted mean satisfaction scores ($\hat{\beta}_{0j}$) are associated with smaller slope coefficients ($\hat{\beta}_{1j}$) (that is, the effect of age on satisfaction is slightly more pronounced in the presence of lower adjusted means).

The objective of this analysis is to assess whether there is a significant difference in mean patient outcome scores across physicians adjusted for appropriate covariates. A comparison of the adjusted mean satisfaction scores ($\hat{\beta}_{0j}$) from the fully stratified analysis is not optimal because in some physician practices these estimates are based on very few patients (for example, $n_{015} = 7$, $n_{017} = 5$, $n_{029} = 8$). We use two-level hierarchical modelling techniques to generate more precise estimates of individual physician's mean satisfaction scores, adjusted for patient's age and the physician's years in medical practice. We begin with the most general model.

6.1.1. Intercept and Slopes as Outcomes Model

The first two-level hierarchical model we consider is the intercepts and slopes as outcomes model (model 6 of Tables I and II). We consider the level 1 and level 2 covariates, patient's age and physician's years in medical practice, respectively, and centre each about their respective grand means ($\bar{X}.. = 62{\cdot}7$ and $\bar{W}. = 20{\cdot}5$). We model the intercepts (physician's adjusted mean satisfaction scores) and slopes (relating age to satisfaction for each physician) as random effects. The HLM/2L and SAS Proc Mixed output for the intercepts and slopes as outcomes model are shown in Tables V and VI, respectively. (The source code for the SAS Proc Mixed run is also provided in Table VI.) Before we compare estimates and hypothesis tests of fixed effects, covariance components and random effects from HLM/2L and SAS Proc Mixed, we outline portions of the outputs which are distinct.

In the last section of Table V HLM/2L provides a test of homogeneity of level 1 variances. The test is not significant ($p = 0{\cdot}370$) supporting our assumption of homogeneity of variances (that is the variances in patient's satisfaction scores across physicians are not significantly different).

SAS Proc Mixed produces several statistics which are useful for assessing model fit (middle section of Table VI). The most widely used are Akaike's information criterion (AIC) and

Table V. Intercept and slopes as outcomes model: HLM/2L

Final estimation of fixed effects:

Fixed effect		Coefficient	Standard error	*T*-ratio	*P*-value
For INTRCPT1	B0				
INTRCPT2	G00	67·9796	0·8626	78·812	0·000
MD_YEARS	G01	− 0·0476	0·1186	− 0·401	0·688
For PT_AGE slope	B1				
INTRCPT2	G10	0·1476	0·0592	2·492	0·013
MD_YEARS	G11	0·0053	0·0084	0·633	0·527

Final estimation of variance components:

Random effect	Standard deviation	Variance component	D.F.	Chi-square	*P*-value
INTRCPT1, U0	5·0063	25·0631	68	143·4043	0·000
PT_AGE slope, U1	0·2363	0·0558	68	99·3944	0·008
Level-1, R	22·7633	518·1695			

Statistics for current covariance components model: deviation = 13636·795; number of estimated parameters = 4.
Test of homogeneity of level-1 variance: chi-square statistic = 72·286; number of degrees of freedom = 69; *p*-value = 0·370

Schwarz's Bayesian criterion (BIC). AIC can be used to compare models with the same fixed effects but different covariance structures. Larger values of the AIC indicate better models. Schwarz's BIC is used for the same purpose and is interpreted in the same manner (that is, larger values indicate better models). However, the two criteria are computed in slightly different ways (for example, Schwarz's BIC involves a larger penalty in models with more covariance parameters) and may lead to different conclusions. We can also use a likelihood ratio test to compare models which are submodels of other models by taking -2 times the difference in log-likelihoods which follows a χ^2 distribution with p degrees of freedom, where p reflects the difference in the number of parameters estimated between the two models. SAS automatically generates -2 times the log-likelihood for each model. We use these criteria to compare different models. We now describe estimates and hypothesis tests for fixed effects, covariance components and random effects from HLM/2L and SAS Proc Mixed.

Fixed Effects. Both HLM/2L and SAS Proc Mixed produce similar estimates of the fixed effects and covariance components. For example, the overall adjusted mean satisfaction score is estimated as $\hat{\gamma}_{00} = 67{\cdot}9796$ by HLM/2L and $\hat{\gamma}_{00} = 67{\cdot}9852$ by SAS. Physician's years in medical practice has little effect on this overall adjusted mean; the regression coefficient associated with the level 2 covariate, physician's years in medical practice, is estimated as $\hat{\gamma}_{01} = -0{\cdot}0476$ by HLM/2L and $\hat{\gamma}_{01} = -0{\cdot}0474$ by SAS ($p = 0{\cdot}688$ and $p = 0{\cdot}6896$ respectively). There is a significant positive association between patient's age and satisfaction with care. The overall slope coefficient relating the level 1 covariate, patient's age, to satisfaction with care is estimated as $\hat{\gamma}_{10} = 0{\cdot}1476$ by both HLM/2L and SAS ($p = 0{\cdot}013$ and $p = 0{\cdot}0150$, respectively). Patient's age is

Table VI. Intercept and slopes as outcomes model: SAS Proc Mixed

Covariance parameter estiomates (REML)					
Cov Parm	Subject	Estimate	Standard error	Z	$\Pr > \lvert Z \rvert$
UN(1, 1)	phys	25·0099	8·4684	2·95	0·0031
UN(2, 1)	phys	− 0·9025	0·4390	− 2·06	0·0398
UN(2, 2)	phys	0·0550	0·0437	1·26	0·2087
Residual		518·3230	19·9024	26·04	0·0001

Model fitting information for Y	
Description	Value
Res Log Likelihood	− 6814·79
Akaike's Information Criterion	− 6818·79
Schwarz's Bayesian Criterion	− 6829·40
− 2 Res Log Likelihood	13629·58
Null Model LRT Chi-square	29·5605
Null Model LRT DF	3·0000
Null Model LRT P-value	0·0000

Solution for fixed effects					
Effect	Estimate	Standard error	D.F.	t	$\Pr > \lvert t \rvert$
INTERCEPT	67·9852	0·8621	68	78·86	0·0001
PT_AGE	0·1476	0·0591	68	2·50	0·0150
MD_YEARS	− 0·0474	0·1185	1352	− 0·40	0·6896
PT_AGE*MD_ YEARS	0·0053	0·0084	1352	0·64	0·5255

SAS Code:
```
proc mixed covtest;
class phys;
model y = pt_age md_years pt_age*md_years/s;
random int pt_age/type = un subject = phys s;
run;
```

positively associated with satisfaction with care; with each additional year of age associated with a 0·15 unit increase in satisfaction. Physician's years in medical practice has little effect on this relationship between patient's age and satisfaction; the regression coefficient associated with the level 2 covariate, physician's years in medical practice, is estimated as $\hat{\gamma}_{11} = 0{\cdot}0053$ by both HLM/2L and SAS ($p = 0{\cdot}527$ and $p = 0{\cdot}5255$, respectively). HLM/2L and SAS Proc Mixed indicate that the physician's years in medical practice is not significantly associated with either their patients' mean satisfaction scores nor the relationship between patient's age and satisfaction. In subsequent models, we will drop physician's years in medical practice.

Covariance Components. The estimates of the covariance components address the magnitude of the variation in the intercepts (adjusted mean satisfaction scores), slopes (relating patient's age to satisfaction) and the patient-level random error. The estimates of the covariance components are similar in HLM/2L and SAS Proc Mixed. HLM/2L estimates the variation in adjusted mean satisfaction scores (intercepts) as $\hat{\tau}_{00} = 25{\cdot}063$, while SAS produces $\hat{\tau}_{00} = 25{\cdot}010$. Both packages indicate that this variation in adjusted mean satisfaction scores among physicians is significant ($p = 0{\cdot}000$ and $p = 0{\cdot}0031$, respectively). HLM/2L estimates the variation in slopes relating patient age to satisfaction among physicians as $\hat{\tau}_{11} = 0{\cdot}0558$ and SAS produces $\hat{\tau}_{11} = 0{\cdot}0550$. HLM/2L indicates that this variation is significant ($p = 0{\cdot}008$), while SAS Proc Mixed does not ($p = 0{\cdot}2087$). The apparent inconsistency in the significance levels of the tests may be due in part to the fact that the data are highly unbalanced (the number of patients per physician ranges from 5 to 45). In addition, SAS produces a Wald statistic (19) which is valid for large samples while HLM/2L produces a chi-square statistic (20).

Random Effects. The estimates of the individual random effects are not shown, as this is not our final model. Estimates of individual random effects will be provided once a final model is determined.

6.1.2. Random Coefficients Regression Model

Because the significance tests for the fixed effects indicate that the level 2 covariate, physician's years in medical practice, is not significantly ($p > 0{\cdot}05$) associated with either the adjusted mean satisfaction scores (intercepts) or the effects of patient's age on satisfaction (slopes), we remove physician's years in medical practice as a level 2 covariate and estimate the random coefficients regression model (model 5 of Tables I and II). The HLM/2L and SAS Proc Mixed output for the random coefficients regression model are shown in Tables VII and VIII, respectively. (The source code for the SAS Proc Mixed run is also provided in Table VIII.)

Fixed Effects. Again both HLM/2L and SAS Proc Mixed produce similar estimates of the fixed effects and covariance components. The overall adjusted mean satisfaction score is estimated as 67·9950 in HLM/2L and as 68·0013 in SAS. Both packages indicate that the adjusted mean satisfaction score is significantly different from zero ($p = 0{\cdot}000$ and $p = 0.0001$, respectively). The regression coefficient associated with the level 1 covariate, patient's age, is estimated as 0·1472 in both packages and is also significantly different from zero ($p = 0{\cdot}013$ and $p = 0{\cdot}0150$, respectively).

Covariance Components. Similar to the previous model, there is significant variation in adjusted mean satisfaction scores ($\hat{\tau}_{00} = 24{\cdot}5087$ in HLM/2L and $\hat{\tau}_{00} = 24{\cdot}4360$ in SAS, $p = 0{\cdot}000$ and $p = 0.0032$, respectively) and although the estimates of variation in the effects of age on satisfaction (slopes) are similar ($\hat{\tau}_{11} = 0{\cdot}0538$ in HLM/2L and $\hat{\tau}_{11} = 0{\cdot}0524$), HLM/2L indicates that this variation is significant ($p = 0{\cdot}009$), while SAS indicates that this variation is not significant ($p = 0{\cdot}2212$). Again, the apparent inconsistency in the tests may be due to the unbalanced nature of the data or the asymptotic nature of the Wald statistic.

Random Effects. The estimates of the individual random effects are not shown, as this is not our final model. Estimates of individual random effects will be provided once a final model is determined.

Table VII. Random coefficients regression model: HLM/2L

Final estimation of fixed effects:

Fixed effect		Coefficient	Standard error	T-ratio	P-value
For INTRCPT1	B0				
INTRCPT2	G00	67·9950	0·8562	79·418	0·000
For PT_AGE slope	B1				
INTRCPT2	G10	0·1472	0·0589	2·499	0·013

Final estimation of variance components:

Random effect	Standard deviation	Variance component	D.F.	Chi-square	P-value
INTRCPT1, U0	4·9506	24·5087	69	144·0142	0·000
PT_AGE slope, U1	0·2320	0·0538	69	99·5919	0·009
Level-1, R	22·7618	518·1005			

Statistics for current covariance components model: deviance = 13623·295; number of estimated parameters = 4

6.1.3. One-Way Analysis of Covariance Model With Random Effects

Based on the small estimates of the variation in the effect of age on satisfaction among physicians (slopes), we model the regression slopes as fixed (that is, model a constant effect of age on satisfaction, or a constant slope, across physicians) and fit the one-way analysis of covariance (ANCOVA) model with random effects (model 3 of Tables I and II without the level 2 covariate, physician's years in medical practice). The SAS Proc Mixed output (and source code) for the one-way ANCOVA model with random effects is shown in Table IX (the output from HLM/2L was comparable and is not shown). Using the model fitting information, AIC suggests that the random coefficients regression model (Table VIII) is the best of the three models considered (although there is not a substantial difference in values), while Schwarz's BIC suggests that the one-way ANCOVA model with random effects (Table IX) is the best of the three considered.

Fixed Effects. The estimate of the overall adjusted mean satisfaction score is 68·0081 which is significantly different from zero ($p = 0{\cdot}0001$). The effect of age on satisfaction is estimated as 0·1533 and is also significantly different from zero ($p = 0{\cdot}0027$). Older patients report significantly higher satisfaction, with each additional year of age associated with an increase of 0·15 units in satisfaction.

Covariance Components. There is significant variation in adjusted mean satisfaction scores ($\hat{\tau}_{00} = 24{\cdot}7819$, $p = 0{\cdot}0031$). We will explore these individual estimates below.

Random Effects. The estimates of the random effects are shown in the last section of Table IX. A sample of the estimates of the random effects, the physician identifier, estimates of standard

Table VIII. Random coefficients regression model: SAS Proc Mixed

Covariance parameter estimates (REML)					
Cov Parm	Subject	Estimate	Standard error	Z	Pr > |Z|
UN(1, 1)	phys	24·4360	8·2939	2·95	0·0032
UN(2, 1)	phys	− 0·9004	0·4288	− 2·10	0·0358
UN(2, 2)	phys	0·0524	0·0429	1·22	0·2212
Residual		518·3323	19·8978	26·05	0·0001

Model fitting information for Y	
Description	Value
Res Log Likelihood	− 6809·88
Akaike's Information Criterion	− 6813·88
Schwarz's Bayesian Criterion	− 6824·49
− 2 Res Log Likelihood	13619·75
Null Model LRT Chi-square	29·5137
Null Model LRT DF	3·0000
Null Model LRT P-value	0·0000

Solution for fixed effects					
Effect	Estimate	Standard error	D.F.	t	Pr > |t|
INTERCEPT	68·0013	0·8555	69	79·49	0·0001
PT_AGE	0·1472	0·0587	69	2·51	00145

```
SAS Code: proc mixed covtest;
          class phys;
          model y = pt_age/s;
          random int pt_age/type = un subject = phys s;
      run;
```

errors, degrees of freedom, t-statistics and two-sided significance levels are also provided. SAS Proc Mixed's estimates of the random effects correspond to the $\hat{\upsilon}_{0j}$'s of Section 4.1.3. These estimates of random effects vary from − 7·67 (physician 003) to 10·57 (physician 013, not shown).

Estimates of Random Coefficients. The estimates of the random effects can be used to generate shrinkage estimates of the random coefficients ($\hat{\beta}^{*}_{0j}$) or estimates of the individual physician adjusted mean satisfaction scores using $\hat{\beta}^{*}_{0j} = 68{\cdot}0081 + 0{\cdot}1533^{*}\,(\bar{X}._{j} - \bar{X}..) + \hat{\upsilon}_{0j}$, where $\bar{X}._{j}$ is the mean age of patients in physician j's practice and $\bar{X}..$ is the mean age of patients in the sample ($n = 1492$). For physicians 003 and 013, the estimates of their mean satisfaction scores adjusted for patient age are $\hat{\beta}^{*}_{0\,003} = 60{\cdot}4$ and $\hat{\beta}^{*}_{013} = 78{\cdot}6$. The OLS estimates of the adjusted mean satisfaction scores (intercepts) for these physicians (see Table IV) were 54·0 and 87·2, respectively. The discrepancies between the OLS estimates and the BLUPs can be attributed, in part, to shrinkage.

Table IX. One-way analysis of covariance (ANCOVA) model with random effects: SAS Proc Mixed

Covariance parameter estimates (REML)					
Cov Parm	Subject	Estimate	Standard error	Z	Pr > |Z|
UN(1, 1)	phys	24·7819	8·3868	2·95	0·0031
Residual		524·9288	19·6453	26·72	0·0001

Model fitting information for Y	
Description	Value
Res Log Likelihood	− 6813·39
Akaike's Information Criterion	− 6815·39
Schwarz's Bayesian Criterion	− 6820·69
− 2 Res Log Likelihood	13626·77
Null Model LRT Chi-Square	22·4948
Null Model LRT DF	1·0000
Null Model LRT P-value	0·0000

Solution for fixed effects					
Effect	Estimate	Standard error	D.F.	t	Pr > |t|
INTERCEPT	68·0081	0·8588	69	79·19	0·0001
PT_AGE	0·1533	0·0511	1421	3·00	0·0027

Solution for random effects						
Effect	Phys	Estimate	SE Pred	D.F.	t	Pr > |t|
INTERCEPT	001	0·914	3·554	1421	0·26	0·7970
INTERCEPT	003	− 7·666	3·404	1421	− 2·25	0·0245
INTERCEPT	005	2·462	3·307	1421	0·74	0·4567
INTERCEPT	006	− 0·195	2·877	1421	− 0·07	0·9460
INTERCEPT	007	− 1·209	3·776	1421	− 0·32	0·7488
INTERCEPT	008	− 0·814	4·050	1421	− 0·20	0·8408
.	.	.	.	.	.	.
INTERCEPT	081	− 5·869	3·476	1421	− 1·69	0·0915

```
SAS Code: proc mixed covtest;
          class phys;
          model y = pt_age/s;
          random int pt_age/type = un subject = phys s;
        run;
```


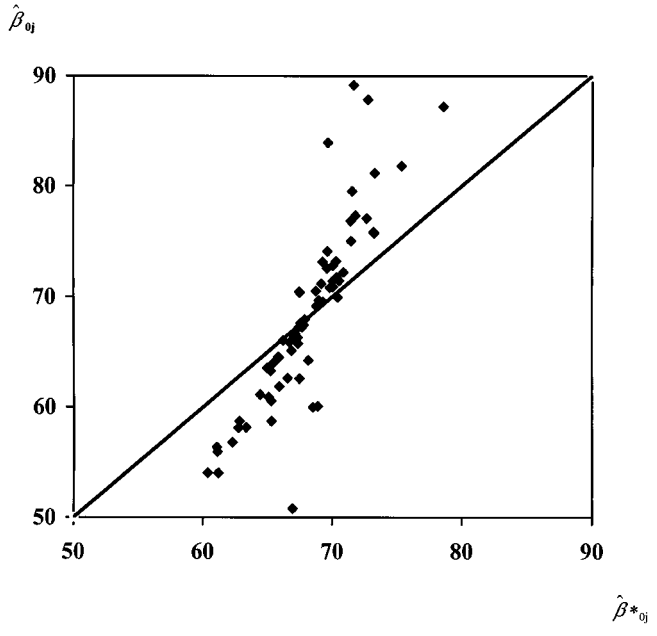

Figure 4. Relationship between OLS intercepts ($\hat{\beta}_{0j}$) and best linear unbiased predictions ($\hat{\beta}^*_{0j}$): shrinkage

Shrinkage. Ordinary least squares (OLS) can be used to estimate the regression equations within each level 2 unit (that is, within each physician practice in our example). If the level 1 covariate is centred about its grand mean, then the estimated intercepts are interpreted as the adjusted mean outcome scores for each level 2 unit (physician). The estimated slopes represent the change in outcome scores associated with a unit change in the level 1 covariate. OLS estimates have desirable statistical properties such as unbiasedness, uniqueness, minimum variance, and so on. However, the OLS estimates may be poor if, for example, the number of level 1 units (for example, patients) within a particular level 2 unit (for example, physician practice) is small. Estimates of individual level 2 effects (for example, the intercepts ($\hat{\beta}^*_{0j}$'s) derived from the $\hat{\upsilon}_{oj}$ of Table IX) produced by HLM/2L and SAS Proc Mixed take into account the precision (or lack thereof) of the estimates within each level 2 unit. Figure 4 displays the relationship between the OLS estimates of the intercepts (from the fully stratified analysis in Table IV) and the BLUPs from the one-way ANCOVA model with random effects in Table IX. Notice that the range of the OLS estimates is from 50·8 to 89·1, while the range of the BLUPs is from 60·4 to 78·6. The BLUPs are pulled toward the estimated overall mean $\hat{\gamma}_{00} = 68{\cdot}008$ (see Table IX). The degree of shrinkage depends on the precision of the OLS estimate, which is related to the number of level 1 units used to generate the estimate. For example, physician 017 has only 5 patients. The OLS estimate of the intercept for physician 017 is 67·0 while the BLUP is 50·8. In contrast, physician 006 has 45 patients, and the OLS estimate and BLUP are closer in value (that is, less subject to shrinkage), 67·8 and 67·9, respectively. (For more details see Chapter 6 of Littell *et al.*[2])

6.2. Analysis of Simulated Data

We simulated data and estimated hierarchical linear models in HLM/2L and SAS Proc Mixed to: (i) compare the parameter estimates and hypothesis tests from HLM/2L and SAS Proc Mixed to one another; (ii) to compare the parameter estimates from each package to the true values; and (iii) to assess the effect of small and unequal (that is, unbalanced designs) samples on parameter estimates and hypothesis tests.

In our simulations, we generated data for a two-level hierarchical structure. We considered a continuous outcome variable, Y, a single level 1 predictor X, and a single level 2 predictor W. Using the PORT study example described in Section 2 as a framework for the simulation, we considered an application with patients as the level 1 units and physicians as the level 2 units. We simulated data in which the level 1 (patient-level) outcome, Y, might reflect self-reported satisfaction scored from 0–100 with higher scores indicative of better satisfaction. We modelled patient age, recorded in years, as the continuous, level 1 covariate. Based on other studies, we considered a positive relationship between age and self-reported satisfaction. In our simulation, we considered patient ages in the range of 30 to 70 years. We modelled physician's years in medical practice as the continuous level 2 covariate and hypothesized that more experienced physicians would be associated with patients who report better satisfaction. In our simulation model, we considered physician's years in practice in the range of 0 to 20 years.

The structure of the level 1 and level 2 models for the simulation study are given below along with the parameters and distribution functions we used to simulate data.

Models and parameters
Level 1 model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij}.$$

Simulation parameters and distribution functions:

$$\varepsilon_{ij} \sim \mathrm{N}(0, 15^2)$$

$$X \sim \mathrm{U}(30, 70).$$

Level 2 models:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(W_j - \bar{W}.) + \upsilon_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(W_j - \bar{W}.) + \upsilon_{1j}.$$

Simulation parameters and distribution functions:

$$\gamma_{00} = 50{\cdot}0, \gamma_{01} = 1{\cdot}5, \gamma_{10} = -4{\cdot}0, \gamma_{11} = 0{\cdot}25$$

$$\upsilon_{0j}, \upsilon_{1j} \sim \mathrm{BVN}(0, 0, 64{\cdot}0, 9{\cdot}0, -5{\cdot}0)$$

$$W \sim \mathrm{U}(0, 20)$$

$$\text{(that is, } \beta_{0j} = 50.0 + 1{\cdot}5(W_j - \bar{W}.) + \upsilon_{0j} \qquad \beta_{1j} = -4{\cdot}0 + 0{\cdot}25(W_j - \bar{W}.) + \upsilon_{1j}).$$

The parameters we modelled reflect the following scenario. Adjusting for physician's years in medical practice, the overall physician-level mean satisfaction score is 50·0 (γ_{00}). For every additional year in medical practice, this adjusted mean satisfaction score increases by 1·5 units

(γ_{01}, physician's experience is positively associated with patient's satisfaction). Adjusting for physician's years in medical practice, the slope relating patient's age to satisfaction is $-4{\cdot}0$ (γ_{10}, each additional year of age is associated with a decrement of 4·0 units in the slope relating patient's age to satisfaction). For every additional year in medical practice, this slope coefficient increases by 0·25 units (γ_{11}).

Simulation Strategy. Data were simulated in SAS according to the models displayed above. Four sample size configurations were considered:

(i) balanced design: large level 2 sample and large level 1 samples, $J = 100$, $n_j = 50$;
(ii) balanced design: small level 2 sample and large level 1 samples, $J = 10$, $n_j = 50$;
(iii) balanced design: large level 2 sample and small level 1 samples, $J = 10$, $n_j = 10$;
(iv) unbalanced design: large level 2 sample and varying level 1 samples, $J = 100$, $n_j = 10(10)50$, where 10(10)50 indicates that n_j ranges from 10 to 50 in increments of 10.

Data were simulated as follows. For each of the J physicians considered, we drew random physician effects ($\upsilon_{0j}, \upsilon_{1j}$) from a bivariate normal distribution (BVN(0, 0, 64·0, 9·0, $-5{\cdot}0$)). We then drew a value for the physician-level covariate, W (for example, physician's years in medical practice), from a uniform distribution over the range of 0–20 and centred this value by subtracting $\bar{W}. = 10$. We then computed the intercepts and slopes for each physician (β_{0j}, β_{1j}). For each of the n_{ij} patients within each physician practice, we drew a random error (ε_{ij}) from a normal distribution with mean 0 and standard deviation 15. We then drew a value for the patient-level covariate, X (for example, patient's age in years), from a uniform distribution over the range of 30–70 and centred this value by subtracting $\bar{X}.. = 50$. We then computed the outcome or dependent variable score for each patient (Y_{ij}).

Analysis. Using simulated data we estimated intercepts and slopes as outcomes hierarchical linear models in HLM/2L and in SAS Proc Mixed and compared the estimates and tests of fixed effects, covariance components and individual random effects. For comparison purposes, we also fit two general linear models to the simulated data using SAS Proc GLM. In the first model we pooled the patient-level data and performed an analysis relating the patient-level outcome, Y_{ij}, to the patient-level covariate, $X_{ij,}$ (centred at $\bar{X}..$) and the physician-level covariate, W_j (centred at $\bar{W}.$). In the second model, we included $(J-1)$ dummy variables indicating individual physician practices and related the patient-level outcome, Y_{ij}, to the patient-level covariate, X_{ij}, and the $J-1$ dummy variables.

Results. In general, HLM/2L and SAS Proc Mixed produced similar estimates of fixed effects, covariance components, and individual random effects (see Appendix), which were very close to the parameters modelled in our simulation. In the pooled analysis, in which we combined all level 1 units (patients) and ignored the level 2 units (physicians), the estimates of the fixed effects were very similar to those produced by HLM/2L and SAS Proc Mixed. However, the standard errors of the fixed effects were much smaller, in some cases as much as 50 per cent smaller, compared to the standard errors from HLM/2L and SAS Proc Mixed. This could lead to inflated type I error rates. In the model in which we included $(J-1)$ dummy variables to reflect the level 2 units (physician practices), it was not possible to model the physician-level covariate, W. The unique effect of the jth physician is captured in an estimate of the individual physician mean and it is not possible to separate the effect due to W, the physician's years in medical practice.

We used simulation techniques to generate samples of data for analysis by various packages and procedures. We did not perform a Monte Carlo simulation in which we varied parameters and effects systematically, therefore we do not draw general conclusions from these analyses. We simply conclude that HLM/2L and SAS Proc Mixed produce similar results based on the configurations we considered. SAS Proc GLM, which is often used in similar applications, is problematic in that standard errors of fixed effects are too small and these models do not allow for adequate specification of effects (for example, the physician-level covariate W, in the model in which we specified individual physician practices using dummy variables). The use of the 'repeated' option in SAS Proc GLM would remedy the problem with the standard errors, but would still leave us with inadequate specification of effects.

7. CONCLUSION

Hierarchical modelling techniques are important to explicitly take into account the multi-level structure of data. In hierarchical models, we can explore the nature and extent of relationships within level 2 units and among level 2 units. Estimates based on data in which level 1 units (for example, patients) are aggregated across level 2 units (for example, physicians) may not be appropriate if there is substantial heterogeneity among the level 2 units. Further, ordinary regression techniques produce estimates of standard errors which are too small, resulting in inflated type I error rates and misleadingly tight confidence intervals.

As a final note, users should be cautious in interpreting results of significance tests (tests for covariance components and individual random effects in particular) when the number of level 2 units is small ($J < 30$) or the data are extremely unbalanced. More research needs to be done to determine the robustness of such tests in the presence of small samples and unbalanced data.

APPENDIX

Comparison of estimates from HLM/2L, SAS Proc Mixed and GLM using simulated data balanced design (number of level 1 units (n_j) constant across level 2 units (J)): large level 1 and level 2 samples, $J = 100$, $n_j = 50$.

Package	HLM/2L	SAS Proc Mixed	SAS Proc GLM (pooled analysis)	SAS Proc GLM (dummy variable to indicate level 2 unit)
		Fixed effects		
True†	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
$\gamma_{00} = 50{\cdot}0$	51·45 (0·795)	49·97 (0·774)	49·80 (0·417)	54·35 (4·253)
$\gamma_{01} = 1{\cdot}5$	1·45 (0.131)	1·49 (0·127)	1·55 (0·069)	
$\gamma_{10} = -4{\cdot}0$	− 4·21 (0·288)	− 4·27 (0·289)	− 4·27 (0·048)	− 4·23 (0·050)
$\gamma_{11} = 0{\cdot}25$	0·18 (0·047)	0·18 (0·047)	0·18 (0·008)	

APPENDIX. (Continued)

	Covariance components			
True†	Estimate	Estimate	Estimate	Estimate
$\sigma^2 = 15^2 = 225$	222·77	222·77		
$\tau_{00} = 8^2 = 64{\cdot}0$	58·70	55·18	MSE = 865·42	MSE = 903·19
$\tau_{11} = 3^2 = 9{\cdot}0$	8·24	8·24		
$\tau_{01} = -5{\cdot}0$	−8·37	−6·40		
	Random coefficients			
First 5 level 2 units (j)	$\hat{\beta}^*_{0j}$ $\hat{\beta}^*_{1j}$	$\hat{\beta}^*_{0j}$ $\hat{\beta}^*_{1j}$	–	$\hat{\beta}_{0j}$
1	54·42	53·30	–	52·04
	−4·71	−4·70		
2	23·15	22·54	–	18·46
	−2·53	−2·55		
3	59·41	58·47	–	58·65
	−3·95	−3·94		
4	45·69	44·56	–	44·38
	−4·73	−4·73		
5	43·02	42·14	–	41·38
	−3·68	−3·69		

†Simulation models:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(W_j - \bar{W}.) + \upsilon_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(W_j - \bar{W}.) + \upsilon_{1j} \qquad \upsilon_{0j}, \upsilon_{1j} \sim \mathrm{BVN}(0, 0, \tau_{00}, \tau_{11}, \rho)$$

where

$$\rho = \frac{\tau_{01}}{\sqrt{(\tau_{00}, \tau_{11})}}.$$

Comparison of estimates from HLM/2L, SAS Proc Mixed and GLM using simulated data balanced design (number of level 1 units (n_j) constant across level 2 units (J)): small level 2 sample and large level 1 samples, $J = 10$, $n_j = 50$.

Package	HLM/2L	SAS Proc Mixed	SAS Proc GLM (pooled analysis)	SAS Proc GLM (dummy variable to indicate level 2 unit)
		Fixed effects		
True†	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
$\gamma_{00} = 50{\cdot}0$	56·12 (3·634)	54·16 (4·313)	55·19 (1·531)	78·93 (6·012)
$\gamma_{01} = 1{\cdot}5$	0·83 (0.824)	0·76 (0·826)	0·76 (0·293)	
$\gamma_{10} = -4{\cdot}0$	−2·53 (0·891)	−4·94 (1·051)	−4·92 (0·170)	−2·65 (0·214)
$\gamma_{11} = 0{\cdot}25$	0.86 (0.202)	0·85 (0·201)	0·82 (0·033)	

APPENDIX. (Continued)

	Covariance components			
True†	Estimate	Estimate	Estimate	Estimate
$\sigma^2 = 15^2 = 225$	212·07	212·07		
$\tau_{00} = 8^2 = 64{\cdot}0$	126·31	128·19	MSE = 828·75	MSE = 1806·30
$\tau_{11} = 3^2 = 9{\cdot}0$	7·83	7·82		
$\tau_{01} = -5{\cdot}0$	− 11·91	− 12·44		
	Random coefficients			
First 5 level 2 units (j)	$\hat{\beta}^*_{0j}$ $\hat{\beta}^*_{1j}$	$\hat{\beta}^*_{0j}$ $\hat{\beta}^*_{1j}$	–	$\hat{\beta}_{0j}$
1	40·67	41·77	–	55·31
	− 13·60	− 13·61		
2	59·93	60·13	–	60·36
	− 2·52	− 2·52		
3	59·29	59·24	–	61·28
	0·71	0·71		
4	55·84	55·87	–	55·98
	− 0·34	− 0·34		
5	68·44	68·76	–	69·82
	− 3·96	− 3·96		

†Simulation models:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}(W_j - \bar{W}.) + \upsilon_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}(W_j - \bar{W}.) + \upsilon_{1j} \qquad \upsilon_{0j}, \upsilon_{1j} \sim \mathrm{BVN}(0, 0, \tau_{00}, \tau_{11}, \rho)$$

where

$$\rho = \frac{\tau_{01}}{\sqrt{(\tau_{00}, \tau_{11})}}.$$

Comparison of estimates from HLM/2L, SAS Proc Mixed and GLM using simulated data balanced design (number of level 1 units (n_j) constant across level 2 units (J)): large level 2 sample and small level 1 samples, $J = 100$, $n_j = 10$.

Package	HLM/2L	SAS Proc Mixed	SAS Proc GLM (pooled analysis)	SAS Proc GLM (dummy variable to indicate level 2 unit)
	Fixed effects			
True†	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
$\gamma_{00} = 50{\cdot}0$	47·06 (0·926)	49·41 (0·944)	49·84 (0·905)	66·97 (9·532)
$\gamma_{01} = 1{\cdot}5$	1·73 (0.164)	1·60 (0·168)	1·54 (0·162)	
$\gamma_{10} = -4{\cdot}0$	− 4·11 (0·276)	− 4·06 (0·276)	− 4·01 (0·105)	− 4·08 (0·118)
$\gamma_{11} = 0{\cdot}25$	0.28 (0.049)	0·28 (0·049)	0·28 (0·019)	

APPENDIX. (Continued)

	Covariance components			
True†	Estimate	Estimate	Estimate	Estimate
$\sigma^2 = 15^2 = 225$	220·92	220·92		
$\tau_{00} = 8^2 = 64{\cdot}0$	60·39	63·57	MSE = 814·59	MSE = 906·11
$\tau_{11} = 3^2 = 9{\cdot}0$	7·26	7·26		
$\tau_{01} = -5{\cdot}0$	− 1·34	− 4·99		
	Random coefficients			
First 5 level 2 units (j)	$\hat{\beta}^*_{0j}$ $\hat{\beta}^*_{1j}$	$\hat{\beta}^*_{0j}$ $\hat{\beta}^*_{1j}$	–	$\hat{\beta}_{0j}$
1	29·82	32·49	–	24·69
	− 5·32	− 5·32		
2	31·81	34·57	–	32·72
	− 5·49	− 5·49		
3	30·98	34·05	–	26·73
	− 6·10	− 6·10		
4	52·69	52·85	–	70·12
	− 0·33	− 0·33		
5	32·43	36·11	–	32·40
	− 7·33	− 7·33		

†Simulation models:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}(W_j - \bar{W}.) + \upsilon_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}(W_j - \bar{W}.) + \upsilon_{1j} \qquad \upsilon_{0j}, \upsilon_{1j} \sim \mathrm{BVN}(0, 0, \tau_{00}, \tau_{11}, \rho)$$

where

$$\rho = \frac{\tau_{01}}{\sqrt{(\tau_{00}, \tau_{11})}}.$$

Comparison of estimates from HLM/2L, SAS Proc Mixed and GLM using simulated data unbalanced design (number of level 1 units (n_j) range from 10 to 50 in increments of 10) large level 2 sample and various level 1 samples, $J = 100$, $n_j = 10\ (10)\ 50$.

Package	HLM/2L	SAS Proc Mixed	SAS Proc GLM (pooled analysis)	SAS Proc GLM (dummy variable to indicate level 2 unit)
		Fixed effects		
True†	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
$\gamma_{00} = 50{\cdot}0$	48·40 (0·940)	49·42 (0·946)	49·34 (0·582)	37·13 (4·490)
$\gamma_{01} = 1{\cdot}5$	1·66 (0.171)	1·66 (0·171)	1·59 (0·105)	
$\gamma_{10} = -4{\cdot}0$	− 4·25 (0·299)	− 4·13 (0·301)	− 4·03 (0·067)	− 4·13 (0·068)
$\gamma_{11} = 0{\cdot}25$	0.20 (0.054)	0·20 (0·054)	0·17 (0·012)	

APPENDIX. (Continued)

	Covariance components			
True†	Estimate	Estimate	Estimate	Estimate
$\sigma^2 = 15^2 = 225$	221·73	221·73		
$\tau_{00} = 8^2 = 64{\cdot}0$	78·12	78·13	MSE = 1001·22	MSE = 1007·45
$\tau_{11} = 3^2 = 9{\cdot}0$	8·79	8·78		
$\tau_{01} = -5{\cdot}0$	− 6·65	− 6·64		
	Random coefficients			
First 5 level 2 units (j)	$\hat{\beta}^*_{0j}$ $\hat{\beta}^*_{1j}$	$\hat{\beta}^*_{0j}$ $\hat{\beta}^*_{1j}$	–	$\hat{\beta}_{0j}$
1 ($n_1 = 10$)	60·07	60·07	–	63·04
	− 2·06	− 2·05		
2 ($n_2 = 20$)	54·91	54·91	–	55·38
	− 5·51	− 5·51		
3 ($n_2 = 30$)	66·55	66·55	–	65·18
	− 2·24	− 2·24		
4 ($n_2 = 40$)	27·01	27·01	–	25·64
	− 4·95	− 4·95		
5 ($n_2 = 50$)	47·15	47·15	–	59·49
	− 9·08	− 9·08		

†Simulation models:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}..) + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}(W_j - \bar{W}.) + \upsilon_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}(W_j - \bar{W}.) + \upsilon_{1j} \qquad \upsilon_{0j}, \upsilon_{1j} \sim \mathrm{BVN}(0, 0, \tau_{00}, \tau_{11}, \rho)$$

where

$$\rho = \frac{\tau_{01}}{\sqrt{(\tau_{00}, \tau_{11})}}.$$

SAS Proc Mixed code for two-level hierarchical models.

Variable name	Description
Y	The dependent variable measured on the ith level 1 unit (for example, patient) nested within the jth level 2 unit (for example, physician)
X	The level 1 (patient) covariate or predictor
W	The level 2 (physician) covariate or predictor
phys	Level 2 identifier (for example physician identification number)

X and W variables can be modelled in their original, untransformed metric or centred (about respective grand means, or X about respective group means).

Model. One-way ANOVA with random effects:

```
proc mixed covtest;
        class phys;
        model y = /s;
        random int/s;
run;
```

Model 2. Means as outcomes regression model:

```
proc mixed covtest;
        class phys;
        model y = w/s;
        random int/s;
run;
```

Model 3. One-way ANCOVA with random effects:

```
proc mixed covtest;
        class phys;
        model y = x/s;
        random int/type = un subject = phys s;
run;
```

Model 4. Non-randomly varying slopes model:

```
proc mixed covtest;
        class phys;
        model y = x w x*w/s;
        random int/type = un subject = phys s;
run;
```

Model 5. Random coefficients regression model:

```
proc mixed covtest;
        class phys;
        model y = x/s;
        random int x/type = un subject = phys s;
run;
```

Model 6. Intercepts and slopes as outcomes model:

```
proc mixed covtest;
        class phys;
        model y = x w x*w/s;
        random int x/type = un subject = phys s;
run;
```


REFERENCES

1. Bryk, A. S., Raudenbush, S. W. and Congdon, R. T. HLM *Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs*, Scientific Software International Inc., Chicago, IL, 1996.
2. Littell, R. C., Milliken, G. A., Stroup, W. W. and Wolfinger, R. D. '*SAS*® *System for Mixed Models*, SAS Institute Inc., Cary, NC, 1996.
3. Greenfield, S., Kaplan, S. H., Silliman, R. S., Sullivan, L., Manning, W., D'Agostino, R., Singer, D. and Nathan, D. M. 'The uses of outcomes research for medical effectiveness, quality of care, and reimbursement in Type II diabetes', *Diabetes Care*, **17**, Suppl 32–39 (1994).
4. Bryk, A. S. Raudenbush, S. W. *Hierarchical Linear Models*: *Applications and Data Analysis Methods*, Sage Publications, Newbury Park, California, 1992.
5. Gatsonis, C. A., Epstein, A. M., Newhouse, J. P., Normand, S. L. and McNeil, B. J. 'Variations in the utilization of coronary angiography for elderly patients with acute myocardial infarction', *Medical Care*, **33**, 625–642 (1995).
6. Skene, A. M. and Wakefield, J. C. 'Hierarchical models for multicentre binary response studies', *Statistics in Medicine*, **9**, 919–929 (1990).
7. Wong, G. Y. and Mason, W. M. 'The hierarchical logistic regression model for multilevel analysis', *Journal of the American Statistical Association*, **80**, 513–524 (1985).
8. Christiansen, C. L. and Morris, C. N. 'Hierarchical Poisson regression modeling', *Journal of the American Statistical Association*, **92**, 618–632 (1997).
9. SAS Institute Inc., *SAS/STAT*® *Software*: *Changes and Enhancements through Release 6.11*, SAS Institute, Cary, NC, 1996.
10. Searle, S. R., Casella, G. and McCulloch, C. E. *Variance Components*, Wiley, New York, 1992.
11. Liang, K. Y. and Zeger, S. 'Longitudinal data analysis using generalized linear models', *Biometrika*, **73**, 13–22 (1986).
12. Zeger, S. L., Liang, K. Y. and Albert, P. S. 'Models for longitudinal data: A generalized estimating equation approach', *Biometrics*, **44**, 1049–1060 (1988).
13. Morris, C. N. 'Parametric empirical Bayes inference: theory and applications', *Journal of the American Statistical Association*, **78**, 47–65 (1983).
14. Kreft, I. G., De Leeuw, J. and Van Der Leeden, R. 'Review of five multilevel analysis programs: BMDP-5V, GENMOD, HLM, ML3, VARCL', *American Statistician*, **48**, 324–335 (1994).
15. Kaplan, S. H., Sullivan, L. M., Spetter, D., Dukes, K. A., Khan, A. and Greenfield, S. 'Gender and patterns of physician-patient communication', *in* Falik, M. M. and Collins, K. S. (eds), *Women's Health*: *The Commonwealth Fund Survey*, The Johns Hopkins University Press, Baltimore, MD, 1996.
16. Hall, J. A. and Dornan, M. C. 'Patient sociodemographic characteristics as predictors of satisfaction with medical care: A meta analysis', *Social Science and Medicine*, **30**, 811–818 (1990).

