MULTILEVEL MODELLING NEWSLETTER

The Multilevel Models Project:

Dept. of Mathematics, Statistics & Computing
Institute of Education, University of London

20 Bedford Way, London WC1H 0AL, ENGLAND
e-mail: temsmya@uk.ac.ioe

Telephone: 071-612-6682 Fax: 071-612-6686

Vol. 4 No. 2 **June, 1992**

Editorial

Dear readers,

We are grateful to everyone who has contributed to Multilevel Modelling Newsletter in past years. Aimed at sharing ideas and experiences of multilevel modelling and its application, it has reached researchers and data analysts in many fields including education, organizational behaviour, psychology, sociology, human growth and development, geography and the health sciences. The total circulation at present exceeds 500 from over 40 countries. To improve our work, your further opinions and contributions are invited. It can be news of your project, small articles about your research, reviews of books, critical papers, announcement of courses, workshops, seminars, meetings, and development in teaching and use of software. Where possible, articles should be submitted as an ASCII text file on floppy disk or email. Please send to me at the project address above. Thank you for your assistance!

Min Yang

ML3 Clinics in London 1992

Free for users of *ML3*/*ML3-E*

Tuesday July 7
Tuesday September 15
Tuesday October 13
Tuesday November 10
Tuesday December 8

10:00 am - 5:00 pm, Room 683
Institute of Education
20 Bedford Way
London, WC1H 0AL

Call Min Yang on 071 612 6682
email: temsmya@uk.ac.ioe

Also In This Issue

The Multilevel Models Project is supported by a grant from the Economic and Social Research Council, U.K.

Multilevel Software - Summary Comparison

	HLM	VARCL	ML3
Authors	Steve Raudenbush Tony Bryk Richard Congdon	Nick Longford	Jon Rasbash Bob Prosser Harvey Goldstein
Availability	Commercially	From author	Commercially
Manual	Adequate to run software	Adequate to run software	Detailed theory, commands for software + examples
Interface	Menu driven, interactive	Menu driven, limited interaction	command driven, interactive
Data manipulation facilities	Minimal but interfaces with SYSTAT	Minimal	Extensive, including simple statistics
MACRO facilities	No	No	Yes
Computational algorithm	EM	Fisher scoring	Iterative generalised least squares
Estimates	ML, REML	ML, Quasi-likelihood for binomial,poisson, gamma variables	ML, REML, General Quasi-likelihood
Output	Standard default set	Standard default	Can be fully tailored
Statistics	Hypothesis tests,likelihood statistic and standard error estimates	Likelihood statistic and standard error estimates	Hypothesis tests, confidence intervals, likelihood statistic and standard error estimates
Weighting	Yes	Yes	No
Ease of learning	Easy	Easy	Slow
Binary response data	No	Yes	Yes
General nonlinear models	No	No	Yes
Complex level 1 structures	Yes,but must be known up to a scalar constant	No	Yes
Linear constraints on parameters	On fixed effects	Covariances can be constrained to 0, variances to a non-negative value. Fixed coefficients to any value.	Yes
Multivariate multilevel models	No	No	Yes
Graphical facilities	None	None	High resolution colour plotting and printing
Implementations	PC(DOS), UNIX, VAX, IBM-CMS	PC (DOS), VAX, UNIX	PC (DOS), VAX
Size of problem handled in PC version	Runs in 640k under DOS. No limit on number of level 1 units.	Runs in 640k under DOS and can handle large problems.	Extended memory version available to handle large problems.
Speed of execution	Fast	Very fast	Moderate
Number of levels	3	3. Up to 9 for simple variance components models	3
On-line help	None	None	Basic command reference

See also *Kreft, etal*(1990), *Bryk,A.S.* and *Raudenbush, S*(1989), *Longford, N.* (1989) and *Prosser,R. etal*(1991). GENMOD, referred to by *Kreft etal* is not generally available now.

Multilevel Models: Selected References

Albander, J.M. & Goldstein, H. (1992). Multilevel statistical models in studies of periodontal diseases. Journal of Periodontology (in press).

Berkey, S., Laird, N.M., Valadian, I. & Gardner, J. 1989, The analysis of longitudinal growth data with covariates. In: Auxology 88; Perspectives in the Science of growth and development, J.M. Tanner. (ed.) London: Smith-Gordon, p. 31-39.

Bock, R.D.(ed.) (1989). Multilevel Analysis of Educational Data, New York: Academic Press,

Bondi, L. & Bradford, M. (1990). Applications of multilevel modelling to geography. Area 22:256-263.

Bryk, A.S. & Raudenbush, S.W. 1989, Toward a more appropriate conceptualisation of Research on School Effects: a three level hierarchical linear model. In: Multilevel Analysis of Educational data, R.D. Bock. (ed.) New York: Academic Press.

Bryk, A.S. & Raudenbush, S.W. (1992). Hierarchical Linear Models, Newbury Park, California:Sage.

Chi, E.M. & Reinsel, G.C. (1989). Models for longitudinal data with random effects and AR(1) errors. J.Amer.Statist.Assocn. 84:452-459.

Creemers, B., Peters, T. & Reynolds, D. (1989). School effectiveness and school improvement, Amsterdam:Swets and Zeitlinger.

DiPrete, T.A. & Grusky, D.B. (1990). The multilevel analysis of trends with repeated cross sectional data. Sociological methodology 20:337-368.

Geary, D.N. (1989). Modelling the covariance structure of repeated measurements. Biometrics 45:1183-1195.

Goldstein, H. (1989). Restricted Unbiassed Iterative Generalised Least Squares Estimation. Biometrika 76:622-623.

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

Goldstein, H. (1991). Multilevel modelling of survey data. The Statistician 40:235-244.

Goldstein, H. (1992). The analysis of randomly cross classified multilevel data. Sociological Methodology (in press).

Goldstein, H. & Rasbash, J. (1992). Efficient computational procedures for the estimation of parameters in multilevel models based on iterative generalised least squares. Computational Statistics and Data Analysis. 13:63-71.

Goldstein, H. & Silver, R. 1989, Multilevel and Multivariate models in survey analysis. In: Analysis of Data from Complex Surveys, edited by D. Holt, C.J. Skinner and T.M.F. Smith. Chichester: J.Wiley.

Jones, K. & Moon, G. (1991). Multilevel assessment of immunization uptake as a performance measure in general practice. British Medical Journal 303:28-30.

Jones, R.H. & Boadi-Boateng, F. (1991). Unequally spaced longitudinal data with AR(1) serial correlation. Biometrics 47:1 61-175.

Kreft, I.G.G. & Kim, K-S. (1991). Statistical Software Reviews. Appl.Statist. 40:343-354.

Kreft,I.G.G., De leeuw,J. & Kim, K-S. (1990). Comparing four different statistical packages for hierarchical linear regression: GENMOD,HLM,ML2,VARCL. UCLA Centre for research on evaluation,, Los Angeles, California, USA.

Laird, N.M. & Louis, T.A. (1989). Empirical Bayes ranking methods. J.Educ.Stats. 14:29-46.

Lee, S. (1990). Multilevel analysis of structural equation models. Biometrika 77:763-772.

Liang, K-Y., Zeger, S.L. & Qaqish, B. (1992). Multivariate regression analyses for categorical data. JRSSB 54:3-40.

Lindstrom, M.J. & Bates, D.M. (1990). Nonlinear mixed effects models for repeated measures data. Biometrics 46:673-687.

Lipsitz, S.R., Laird, N.M. & Harrington, D.P. (1991). Generalised estimating equations for correlated binary data: using the odds ratio as a measure of association. Biometrika 78:153-160 .

Lockheed, M.E. & Longford, N.T. 1991, School effects on mathematics achievement gain in Thailand. In: Schools, classrooms and pupils, S.W. Raudenbush and J.D. Willms. (ed.) New York: Academic Press.

Longford, N.T. 1989, Fisher scoring algorithm for variance component analysis of data with multilevel structure. In: Multilevel Analysis of Educational Data, edited by R.D. Bock. New York: Academic Press.

McDonald, R.P. & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. Brit.J.Math.Statist.psychol. 42:215-232.

Miller, M.E. & Landis, J.R. (1991). Generalised variance component models for clustered categorical response variables. Biometrics 47:33-44.

Muthen, B. & Satorra, A. 1989, Multilevel Aspects of Varying Parameters in Structural Models. In: Multilevel Analysis of Educational Data, R.D. Bock. (ed.) New York: Academic Press.

Pan, H.Q. & Goldstein, H. (1992). A 2-level cross sectional growth model using grafted polynomials. Annals of Human Biology (in press).

Paterson, L. & Goldstein, H. (1991). New statistical methods for analysing social structures: an introduction to multilevel models. Brit.Educ.Res.J. 17:387-394.

Petkau, A.J. & Sitter, R.R (1989). Models for quantal response experiments over time. Biometrics 45:1299-1308.

Plewis, I. (1991). Using multilevel models to link educational progress with curriculum coverage. In: Schools, classrooms and pupils, S.W. Raudenbush & J.D. Willms. (ed.) New York: Academic Press.

Prosser, R., Rasbash, J. & Goldstein,H. (1991) Data Analysis with ML3: London, Institute of Education.

Raudenbush, S.W. (1989). The analysis of longitudinal multilevel data. Int.J.Ed.Research 13:721-740.

Riddell, A,B. (1989). An alternative approach to the study of school effectiveness in third world countries. Comparative Education Review 33:481-497.

Thall, P.F. & Vail, S.C. (1990). Some covariance models for longitudinal count data with overdispersion. Biometrics 46:65 7-671.

Willms, J.D. & Raudenbush, S.W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. JEM (In Press)

Willms, J.D. & Raudenbush, S.W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. JEM 26:209-232.

Zeger, S.L. & Karim, M.R. (1991). Generalised linear models with random effects; a Gibbs sampling approach. J.Amer.Statist.Assocn. 86:79-86.

Workshop Announcement

Workshop on three-level modelling: London, 1-4 December 1992. The four day session is based around the *ML3* program package, lectures, seminars, discussions and participants' presentation. The main topics of the workshop will cover the following:

> variance component models, random coefficients models, complex level 1 variation, residual analysis, repeated measures data analysis, multivariate models, Logit or binary data models

Seminars related to the application of multilevel approach will be given by *Linsday Paterson* from Centre for Educational Sociology in Edinburgh and *Kelvyn Jones* from Department of Geography, Portsmouth Polytechnic. For more details, information and booking, please contact *Min Yang*.

A Guide to ML3 for New Users

In working with users new to multilevel modelling the multilevel models project has found that there are certain problems which recur. To help to acquire a sound grasp of the fundamental ideas and so avoid those problems, a new guide for *ML3* users is available now from the Project.

This booklet describes the principles of multilevel modelling and demonstrate them in a simple analysis of an educational data set. An introduction to the *ML3* package follows after some basic ideas in the context of the data set are explained.

Please write to the Multilevel Models Project if you are *ML3/ML3-E* user and wish to have a copy.

APPLICATIONS

Multilevel Models for Comparing Schools

Harvey Goldstein, Jon Rasbash, Min Yang, Geoffrey Woodhouse, Huiqi Pan
Desmond Nuttall, Sally Thomas

This article is based upon a paper read to the European Conference on Educational Research, University of Twente, June 1992.

Introduction

The important paper of Aitkin and Longford (1986) established that the minimal requirement for valid institutional comparisons was an analysis based upon individual level data which adjusted for intake differences and used efficient techniques of multilevel modelling.

In that paper and the discussion on it, several outstanding problems were raised. The purely technical problems of carrying out the estimation for large data sets have been solved fairly effectively by the development of computer programs such as VARCL, HLM, and ML3. The remaining problems are concerned with the existence of suitable measurements which can be used to adjust for intake, and any other relevant, differences, the multivariate nature of school outcomes and the kinds of interpretations which can be made of results. The present paper addresses some of these issues.

Data

The data are examination results from 5748 students in 66 schools in six Inner London Education Authorities. These students had data on their General Certificate of Secondary Examination (GCSE) grades in Mathematics, and English together with a total score for all the subjects taken in that examination. A full description of the data and the scoring system used is given in Nuttall et al(1989). For Mathematics and English, a score ranging from 0 (no grade awarded) to 7 (grade A) was used in the analysis and for the total score, the scale ranged from 0 to 70. These students also had scores on a common reading test taken when they were 11 years old and were graded also into three categories on the basis of a verbal reasoning (VR) test at 11 years.

Several exploratory models were fitted and the following table gives estimates for the model found to give the most satisfactory fit, using ML3 which allows for complex level 1 variation.

In the fixed part of the model the table shows the average effects of the explanatory factors fitted jointly. The effect of school gender is small and there seems to be a small advantage for those attending Roman Catholic schools. Girls do better than boys, and as expected there are large differences between those in the different verbal reasoning categories and a strong relationship with *LRT*, which is quadratic.

Fixed Part

	Estimate	S.E.
Intercept	-0.53	
LRT	0.37	0.02
LRT^2	0.035	0.008
VR1 – VR3	0.70	0.04
VR2 – VR3	0.31	0.03
Girls – Boys	0.13	0.03
Girls – Mixedschool	0.07	0.06
Boys – Mixedschool	0.09	0.07
C.E. – Stateschool	-0.04	0.13
R.C. – Stateschool	0.20	0.06
Other – StateSchool	0.12	0.16

Random - between schools: Covariances (correlations)

	Intercept	*LRT*	*VR1 – VR3*
Intercept	0.055		
LRT	0.012(0.75)	0.0046	
VR1 – VR3	0.013(0.40)	0.0091(0.97)	0.019

Random - between students

	Intercept	*LRT*
Intercept	0.55	
LRT	0.046	0

The level 1 (between students) variance is thus a linear function of *LRT* score, given by;

$Variance = 0.55 + 0.092 \times LRT$

A test for the sets of random parameters yields significance levels of 0.01 or less.

Turning to the between school variation, we see that the relationship between exam score and *LRT* varies as does the difference between verbal reasoning categories 1 and 3, with high positive correlations. At the student level the variance increases with increasing *LRT* score, so that the estimated variance for an *LRT* score of -2 is 0.37 and for one of +2 is 0.73. The proper specification of the level one variance is important in order to increase precision and to enable complex variation to be fitted at level 2 and above.

To illustrate some of the implications of this model, we can form particular extreme combinations of school residuals. For each school we can estimate the 'effect' for students with an *LRT* score of -2, the approximate lower 2.5th percentile, and in verbal reasoning group 2 or 3, and also the estimated 'effect' for a student at the approximate upper 2.5th percentile and in verbal reasoning group 1. These are plotted against each other in the following figure.

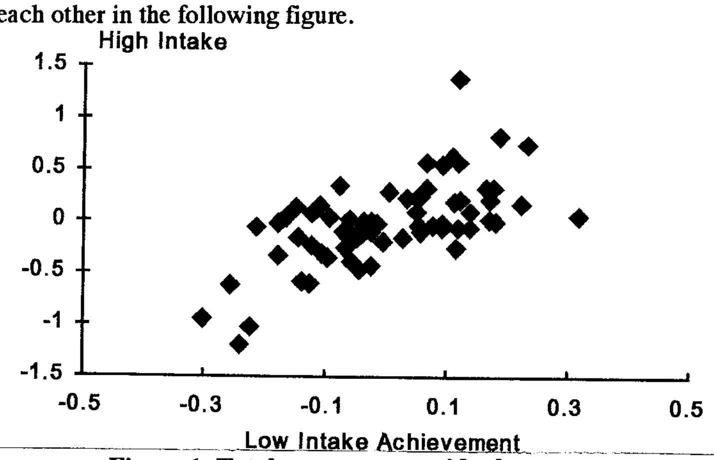


Figure 1. Total exam score residuals

As pointed out above, there is a positive correlation between the school 'effect' for the low and high scoring intake achievers. Nevertheless, there are some schools with below average values for the low achievers who have above average values for the high achievers, and vice-versa. This emphasises the point that schools appear to be differentially effective for different kinds of students.

Because the residuals are estimated they have a sampling variation, and this enables us to construct confidence intervals for them. In figure 2 are shown approximate 95% C.I. for the estimates of the intercept residual, that is the school 'effect' estimated at the mean *LRT* score for those in verbal reasoning groups 2 and 3. It should also be noted that these intervals are 'conservative' since they are calculated separately for each residual, and are based upon the estimated standard error which in general will be an underestimate of the true standard error. As can be seen, there is a very considerable overlap of intervals, which suggests that, apart from the extremes, it is not possible statistically to discriminate easily between schools. This has important implications for the use of such estimates, when comparing schools.

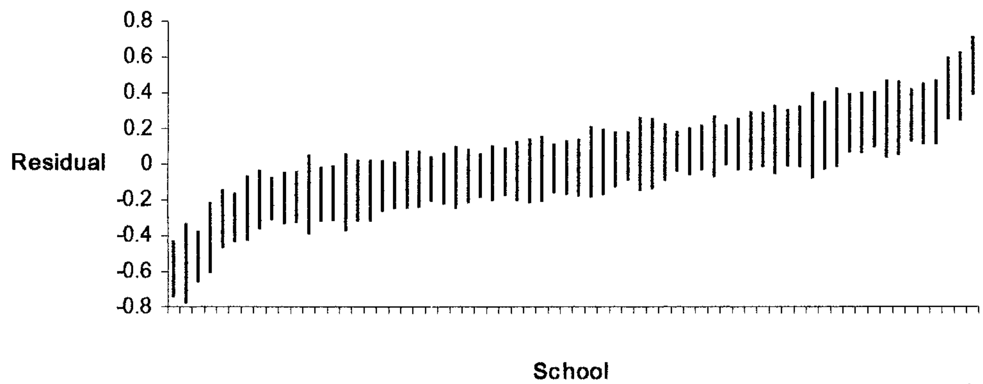


Figure 2. Total exam score residuals confidence intervals

Discussion

The principal aim of this analysis has been to show how differences between schools in examination results vary by intake achievement and in addition, the analysis explores the extent to which schools can be compared based on residual estimates of 'effectiveness'. A further analysis, not described here, explores the relationships among different examination subjects.

It is clear that there is no single dimension along which schools differ. The ordering of school effects depends on the intake achievements of students. It is clear that the uncertainty attached to individual school estimates, at least based upon a single year's data, is such that fine distinctions and detailed rank orderings are statistically invalid. This has implications for published 'league tables' whether or not these are adjusted for intake achievement and whether or not multilevel modelling has been used. At best, a study of residuals differentiated by intake achievement and by subject, can suffice as a screening device and as feedback to individual schools about potential problems.

Perhaps the most promising line for further research lies not in studying schools residual estimates, but in collecting high quality data, over time, which will enable us to study the factors associated with successes and failures and to discover those factors which can account for the observed variation between schools along its several dimensions.

APPLICATIONS

Adjusting for measurement unreliability in multilevel modelling

Min Yang, Geoffrey Woodhouse, Harvey Goldstein, Huiqi Pan and Jon Rasbash

Introduction

Most measurements in educational research are subject to error in the sense that a repetition of the measurement process did not produce an identical result. For example cognitive outcomes such as scores on standardized tests can be affected by the atmosphere and environment of the schools and classes where these outcomes are measured, by the administration of the tests and by differences in the way examination regulations are interpreted. Non-cognitive outcomes such as children's behaviour, self-concept and attitudes to school can be more unreliable owing to the differential effects of interviewer skill, questionnaire design, collaboration between parents and school teachers, school or class organization, and school policy towards the investigation itself.

It is well known that the use of such measurements in the analysis of educational effects, without taking measurement errors into account, can lead to mistaken causal inferences. For example *Goldstein* (1979) shows, in an analysis of the educational attainment of children aged 11, how a conclusion can be reversed when a correction is introduced for measurement error. *Fuller* (1987) gives a comprehensive account of methods for dealing with errors of measurement in regression models but observes that few statistical studies appear to use such procedures. *Plewis* (1985) reviews the traditional methods of correcting for measurement error proposed by *Degracie and Fuller* (1972), *Jöreskog* (1970), and explores the effects of different methods on the conclusions obtained.

These studies are based on classical single-level regression models. Most educational research data, however, have a hierarchical structure and are most efficiently analysed by means of multilevel models. In a paper presenting to the first European Conference on Educational Research in 1992, a preliminary explanation of multilevel measurement error models is carried out. Here we summary the main results. Interested readers can obtain the full paper from the first author.

Basic Theory

In multilevel modelling, the procedure of adjusting for unreliability at different levels is not as straightforward as that in a single level modelling. In particular, the estimates of the random parameters, the variances and covariances at each level, are also affected. *Goldstein* (1986) shows how to make adjustment when variables defined at level 1 are subject to measurement error.

In the present paper, we consider the further problem when explanatory variables defined at higher levels have errors of measurement. In particular, suppose we wished to use the 'contextual' variable which is the average of the prior achievements of all the students in a school, we now have a model

$$Y_{ij} = \alpha + \beta X_{ij} + \gamma \overline{X}_{.j} + u_j + e_{ij} \qquad (1)$$

Where $\overline{X}_{.j}$ is the school mean and u_j, e_{ij} are the level 2 and level 1 random residual terms respectively. If X is measured with error, then $\overline{X}_{.j}$ has error inherited X. If X is measured without error but only n_j of the students in the *j-th* school had been sampled out of N_j in the school total, still $\overline{X}_{.j}$ is measured with error in terms of sampling. This situation is in fact very common in practice, e.g. in household surveys.

Let the true mean in the *j-th* school be $\overline{x}_{.j} = \sum x_j/N_j$ and sample mean as $\overline{X}_{.j} = \sum X_j/n_j$. Following the standard definition of reliability, we can define the level 2 or school level reliability, for the *j-th* school, as,

$$R_{2j} = Var(\overline{x}_{.j})/Var(\overline{X}_{.j}) = R_1 \times n_j/N_j \qquad (2)$$

where n_j/N_j is the sampling fraction in the *j-th* school, and R_1 is the reliability at pupil level. Therefore measurement error at higher levels is a function of the sampling fraction and the level 1 reliability. It is the quantity used to adjust the estimator of γ in equation (1) to get a consistent estimator.

Data Set

The data set we use is from the Junior School Project (JSP). The JSP pupils attended 50 primary schools, selected randomly from the 636 primary schools that were maintained by the Inner London Education Authority at the start of the project. Numerous measurements were made on the pupils during the four years of the study, including measurements of background and of educational outcomes in cognitive and non-cognitive domains.

We consider the data to have a two-level structure, with pupils at level 1 and schools at level 2. For our illustrative purposes we consider only a variance components model,

$$y_{ij} = \alpha + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \gamma \overline{x}_{1.j} + u_j + e_{ij} \qquad (3)$$

Where, for the *i*th pupil in the *j*th school, y_{ij} is the pupil's reading score in year 5. x_{1ij} is the pupil's reading score in year 3. x_{2ij} is the pupil's family socio-economic status, coded 1 if the father is employed in non-manual work, 0 otherwise, and $\bar{x}_{1,j}$ is the mean year-3 reading score for pupils in school j, as estimated from the sample.

The u_j and e_{ij} are level 2 and 1 residuals. Parameters to be estimated are the variances σ_e^2, σ_u^2, the fixed quantities α, β_1, β_2 and γ. Both x_{1ij} and $\bar{x}_{1,j}$ are subject to measurement error. x_{2ij} is assumed to be known without error.

After deletion of cases with missing data, 1075 pupils out of 2287 in 48 schools remain. Sampling fractions for each school vary from 0.16 to 1.00.

Results

The estimates resulting from four versions of the model described by equation (3) are shown in the table. Version A makes no adjustment at all. Version B adjusts for measurement error in x_{1ij} (reliability 0.80 assumed) but not for measurement error in $\bar{x}_{1,j}$, that is, it adjusts for measurement error at level 1 only. Version C adjusts for measurement error at level 2 only, that is, it assumes a reliability of 1.00 for x_{1ij} but adjusts for the sampling error in $\bar{x}_{1,j}$. Version D adjusts for measurement error at levels 1 and 2, assuming a reliability of 0.80 for x_{1ij}.

In version B, the estimate of the coefficient of x_{1ij} has increased in absolute value in comparison with the estimate of version A. The relationship with year 3 score has increased, and the standard error of this estimate has increased by a factor of approximately 2.8.

Parameter estimates of four versions of the model (standard errors in parentheses)

Fixed Part	Version A	Version B
$\hat{\alpha}$	40.9 (0.94)	41.5 (0.94)
$\hat{\beta}_1$	0.68 (0.02)	0.89 (0.05)
$\hat{\gamma}$	-0.02 (0.08)	-0.22 (0.11)
$\hat{\beta}_2$	4.48 (1.11)	2.21 (1.08)
Random		
$\hat{\sigma}_e^2$	148.7 (8.4)	74.8 (10.7)
$\hat{\sigma}_u^2$	30.3 (6.1)	29.9 (6.1)
Intra-school	0.17	0.29

	Version C	Version D
$\hat{\alpha}$	40.9 (1.00)	41.3 (1.33)
$\hat{\beta}_1$	0.68 (0.02)	0.90 (0.06)
$\hat{\gamma}$	-0.05 (0.27)	-0.60 (0.54)
$\hat{\beta}_2$	4.48 (1.11)	2.23 (1.09)
Random		
$\hat{\sigma}_e^2$	148.8 (8.4)	74.9 (10.9)
$\hat{\sigma}_u^2$	30.3 (6.1)	25.4 (9.2)
Intra-school	0.17	0.25

The variation among pupils is reduced by 50%, which is therefore the amount explained by the assumed measurement error of x_{1ij}. The variation among schools does not change appreciably from version A, but represents a considerably higher proportion of the total variation in version B.

The adjusted estimate of the effect on progress of a pupil's social class is half the unadjusted estimate.

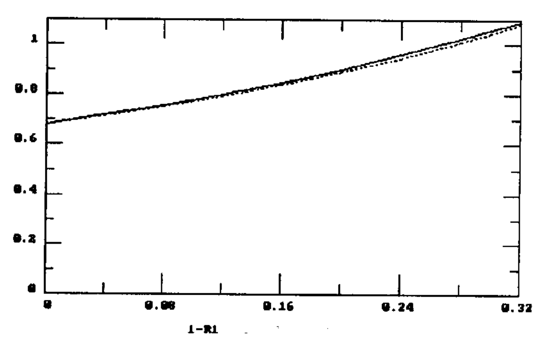
The school mean year-3 score is now estimated to have a significant effect on pupils' progress. Since x_{1ij} and $\bar{x}_{1,j}$ are negatively correlated, a substantial increase in the coefficient of x_{1ij} would be expected to be accompanied by a decrease in the coefficient of $\bar{x}_{1,j}$. The standard error of this estimate is likely to be underestimated because of failure to adjust for measurement error in $\bar{x}_{1,j}$.

In version C, the only appreciable change from the parameter estimates of version A is in the coefficient of the level-2 variable, $\bar{x}_{1,j}$. This estimate increases in absolute value by a factor of approximately 2.5. Its standard error increases by a factor of nearly 3.

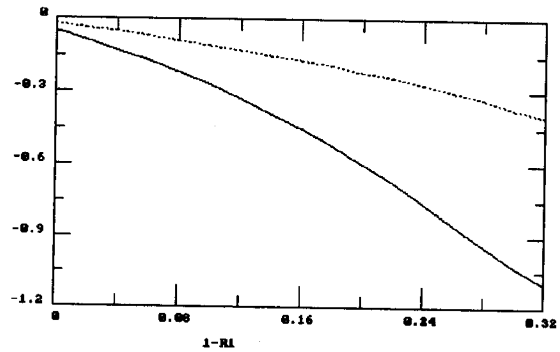
When adjustment is made at both levels, as in version D, the precision of all the parameter estimates, both fixed and random, is reduced. The estimates themselves are similar to those of version B, the main exception being the coefficient of $\bar{x}_{1,j}$ whose absolute value is very much increased, as is its standard error. The size of this estimate is now such that it can scarcely be ignored, but the estimate is so imprecise that any predictions based on it must be given very wide confidence intervals.

The analysis so far has been based on the assumption that the reliability of x_{1ij} is 0.80. Good estimates of reliability are rare in practice and we have explored the effect of varying this assumption by letting R_1 vary from 1.00 to 0.68, then running version B and D of the model.

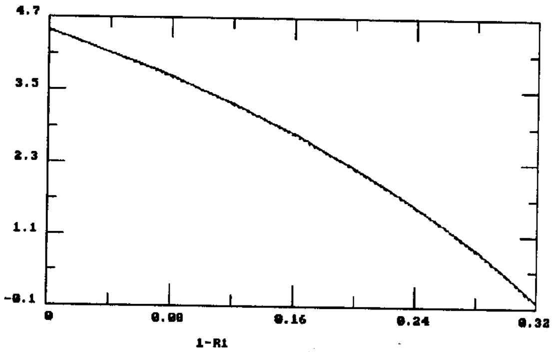
The fixed parameter estimates are plotted in Figure 1, in which the broken lines refer to version B and the solid lines to version D. In both versions the estimates of β_1 and γ increase in absolute value as R_1 decreases. The effect on γ is very much more marked when adjustment is made at both levels. Estimates of β_2, the coefficient of the social-class variable, decrease, eventually becoming small and negative.



1.1 Estimates of β_1



1.2 Estimates of γ



1.3 Estimates of β_2

Figure 1. Variation in fixed parameter estimates as R_1 decrease from 1.00 to 0.68.

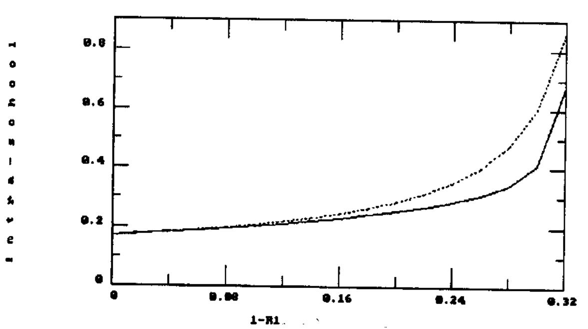


Figure 2 Intra-school correlation as R_1 decrease from 1.00 to 0.68.

With the random parameters, as R_1 decreases the level-1 estimates $\hat{\sigma}_e^2$ in both versions of the model decrease. They are halved as R_1 changes from 1.00 to 0.80, and continue to decrease proportionable at an increasing rate as R_1 moves from 0.80 to 0.68. At the same time the standard errors of these estimates increase substantially, eventually exceeding the estimates themselves.

The two versions differ in their estimates of the level-2 variance σ_u^2, in particular as R_1 decreases below 0.80. Whereas in version B the estimates decrease only gradually with very slight increases in standard errors, when adjustment is made for measurement error at level 2 the level-2 variance decreases at an accelerating rate, like the level-1 variance. As with the level-1 variance, the standard error of the level-2 variance increases rapidly and eventually exceeds the estimate itself. The effects of decreasing R_1 on the intra-school correlation are compared for the two versions in Figure 2. This statistic increases sharply for values of R_1 below about 0.75.

Discussion

In a single-level model, adjustment for measurement error in an explanatory variable generally has the effect of increasing the size of the estimate of the associated coefficient. Our analysis has confirmed a similar effect in a multilevel model, but in this case the correction is not equivalent to dividing by the reliability of the explanatory variable. As in the single-level case, adjustment for measurement error can have substantial effects on other parameter estimates also.

Standard errors of both fixed and random parameter estimates increase when adjustment is made. The amount of the increase depends on the level at which adjustment is carried out. If adjustment is carried out at level 1 only, the standard errors of level-2 parameters increase only slightly; adjustment at both levels can lead to much larger increases in level-2 standard errors. A similar effect is seen in the case of the intercept.

The estimates of the random parameters are reduced, while their standard errors increase, at the levels at which adjustment for measurement error is carried out. As the extent of measurement error is allowed to increase it accounts for more and more of the observed variation. If adjustment is carried out only at level 1, the level-2 variance is over-estimated, as is its precision, leading to unjustified conclusions about school differences.

Analyses of school outcomes often use school-level data based on a sample. Examples include the proportion of pupils eligible for free meals, the proportion of girls in the school, the proportion of pupils from specific ethnic backgrounds etc. Such variables are subject to sampling error, even if at the individual level they are known without error. These errors are most conveniently treated in a multilevel model as measurement errors at the appropriate level.

In a household survey, where we wish to model the hierarchical population structure, we will typically sample only a small percentage of households in an area. If we wish to use area-level variables, based upon aggregating the sample household characteristics, then these aggregate-level variables will tend to have low reliabilities. The same situation can arise in the modelling of electoral behaviour, where we may wish to analyze the influence of ward characteristics using ward-level variables based upon a sample of votes within wards. The method described here enables us to adjust for unreliability for both level-1 and higher-level explanatory variables in multilevel modelling if suitable estimates of measurement errors for these variables are available.

RESEARCH AT WISCONSIN MEDICAL SCHOOL

The Wisconsin Medical School is applying multilevel modelling in a variety of medical studies. Here are some examples.

Heart Disease Prevention in Primary Care. *Mcbride,P.E., Brown,R.L., Solberg,L., Plane, M.B., Underbakke, G. and Schrott, H.*: The systematizing of primary care prevention services has the potential for improving physicians' implementation of national guide-lines for coronary heart disease (CHD) prevention. This five-year study will provide extensive descriptive information regarding current prevention practices of primary care physicians and will evaluate two interventions designed to improve recognition and management of patients who have high blood cholesterol, hypertension, history of smoking, or family history of premature CHD. The data are structured as that patients are nested under physicians, and physicians are nested under clinics. The proposed analytic strategy is based on current advances in multilevel models theory and application, which has several advantages in this research. First, it allows us to assess treatment effects while controlling for potential contaminating variables. Second, it will prove a more appropriate modeling of the situation at hand. Third, the analytic approach incorporates the fact that units within natural groups share common features, which provides a more realistic portrayal of the effects of grouping.

Management, Staff Burnout and Severely Mentally Ill Client Satisfaction. *Schulz, R., Greenley, J., Greenberg, J., Brown, R.L. and Mckee, D.:* Client satisfaction with the services they receive from their mental health care provides is a primary outcome objective in the care of the severely mentally ill (SMI). This project is to advance knowledge to see if certain management practices can raise staff satisfaction and lower burnout and, in turn, elevate client with satisfaction. To analyze the relationship among management, staff and client satisfaction, multilevel models approach will be used. The data are including the satisfaction of approximately 444 SMI clients, 52 mental health staff members who manage these clients and management practices in the 23 organizations in which these managers serve their clients.

Variation in a medical faculty's decisions to transfuse: Implications for modifying blood product utilization. *Brown, R., Brown, R.L., Edwards, J.A. and Nutz, J.F.:* Previous studies have identified disconcerting differences in transfusion practices among physicians. This study investigates decision making on transfusion by a medical school faculty. One hundred fourteen physicians indicated their propensity to transfuse 24 hypothetical patients who varied systemically in four clinical cues. Physicians' decisions to transfuse in the 24 vignettes were analyzed using factorial survey methods. ML3 software was used to construct a two-level hierarchical linear model of vignette effects within and between the physicians. Although the physicians agreed on the order of importance of the clinical cues, they varied significantly in the importance they empirically attributed to the cues and in their thresholds for transfusion . Aside from a slightly higher propensity to transfuse by general internists, differences in decision making were not attributable to physician-based variables. The wide variability in decision making among physicians suggests that quality assurance programs and guide-lines only listing the cues that ought to influence transfusion decisions, have low potential to achieve blood product conservation.

The use of three-level hierarchical linear modelling in cancer prevention research. *Brown, R. L.:* In cancer prevention research, many variables are measured from a variety of sources (e.g., patient, physician and treatment variables), which are hierarchical in nature. Multilevel models analysis allows to model simultaneously these three hierarchical levels of information. Two three-level models are demonstrated on the analysis of a cancer prevention dataset.

Address for details:
Dr Roger L. Brown (Coordinator of Research)
Director-Biostatistics Unit, Research Division
Department of Family Medicine
University of Wisconsin Medical School
Madison, WI 53715