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@ioe.ac.uk

Tel: (0)171 612 6682 Fax: (0)171 612 6686

Vol. 7 No. 1 March, 1995

First release of *MLn*

The successor program to *ML3*, *MLn*, is now available. This is a major revision which incorporates a number of improvements and innovations. The major ones are:

- The ability to handle data with any number of levels and parameters, within the capacity of the computer. For many problems, including those where there are complex cross-classifications, the number of levels often exceeds 3 and seven or eight levels are not uncommon. The user can now specify her own configuration at run time.
- A new version of the IGLS algorithm which speeds up many calculations and allows models of considerable flexibility to be fitted.
- A new MACRO language which efficiently automates sequences of instructions.
- Specially written MACROS to fit models with discrete responses, such as binary data, using improved estimation procedures. MACROS are also being prepared for the analysis of time series data, for data exploration using residuals and influence diagnostics, for fitting parametric and semiparametric survival/event history models, and for models where a variance is allowed to be a nonlinear function of explanatory variables.
- Matrix manipulation facilities which also give access to the matrices used in the estimation process.
- A facility to handle case weighting.

Future plans include a facility to carry out weighted analysis where there is informative non-response, procedures for handling efficiently missing data, bootstrapping and commands for the fitting of Bayesian models using Markov Chain Monte Carlo methods.

A full version and an inexpensive student version (up to 3 levels and 50K worksheet cells) are available. For details of ordering *MLn* or upgrading *ML3-E* to *MLn* contact *Mrs Milene Adaken, Finance Department of the Institute of Education. Tel +44 (0)171 612 6024. Fax +44 (0)171 612 6032. Email: mln.order@ioe.ac.uk.*

ML3/MLn Clinics in London 1995

Tuesday April 4
Tuesday May 9
Tuesday June 6
Tuesday July 11
Tuesday September 5
Tuesday October 3

Multilevel Models Project
11 Woburn Square, Second floor
London WC1A 0SN
contact Min Yang *for appointment at the project address*

Also In This Issue

aLCD analysis of large and complex datasets

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Multilevel Modelling Workshops & Courses

Workshop in Norwich: A three-day workshop to be conducted by Dr *Ian Langford* and Professor *Harvey Goldstein* will be held at University of East Anglia from 30th August to 1st September 1995. As a general introduction to the theory and practice of multilevel modelling using *MLn*, the workshop will cover topics such as basic principles, setting up two and higher level models, repeated measures, logistic and log-linear models, multivariate analysis and diagnostics. For further information please contact *Anné-List McDonald* at *Health Policy and Practice Unit, University of East Anglia, Norwich, NR4 7TJ. Tel +44 01603 593631, email a.cox@uea.ac.uk.*

Workshop in Glasgow: A worshop on the use of multilevel modelling in Public Health and Health Services Research will be held at the University of Glasgow from 27 to 29 September 1995. This workshop will give participants the chance to analyse personal data sets using *MLn* as well as following worked examples introducing a variety of models applied in the health field. Further details are available from *Alastair Leyland, Public Health Research Unit, University of Glasgow, 1 Lilybank Gardens, Glasgow G12 8RZ. Tel: +(0) 141 339 3118 Fax: +(0) 141 337 2776 E-mail: a.leyland@udcf.gla.ac.uk*

Workshop in Edinbough: A training workshop on Educational Data Analysis for Monitoring School Performance Using *MLn* will be held at the Centre for Educational Sociology, University of Edinburgh from 21st – 23rd*June* 1995, with Professor *J. Douglas Willms* as the instructor. The workshop will provide an understanding of the influences affecting school performance, and of the methods used to estimate school effectiveness. It will provide an introduction to multilevel statistical models which can be used to estimate 'added value' of schools. It will enable participants to interpret and report the findings of multilevel analyses. The intended audience are school and education authority administrators who are involved in the analysis of questionaire data or data describing pupil's test results. and educational researchers with an interest in the field of school and teacher effects.

For further details and booking form please contact *Mrs Marcia Wright, Centre for Educational Sociology, University of Edinburgh, 7 Buccleuch Place, Edinburgh EH8 9LW. Tel 0131 650 4186, Fax 0131 668 3263.*

A research half-day on multilevel models

This was held at the London School of Hygiene and Tropical Medicine in 5th January 1995, hosted by the Medical Statistical Unit and attended by about 31 researchers and research students from several universities.

A three-level analysis on quality of life data by *Heather Beacon* addressed the issues on multivariate repeated measures models, the appropriate choice of link function and missing data assumptions. The results were compared to two analogous univariate models, which led to a statistical discussion of the differences between the two models.

Standard linear models for continuous data assumed homoscedasticity of residuals. Sometime this can be achieved by variable transformation, but if not the problem is often ignored. *Bob Carpenter*'s presentation focused on how to actually model heteroscedastic residual variances rather than ignor them from two examples in a child growth study using *ML3*. The effect on the regression parameters and the potential for constructing normal ranges were discussed.

Based on data from a multicentre anti-atheroma study investigating the impact of cholesterol lowering on arterial width, *Paul Seed* presented results from three methods of analysis, an OLS model, ANOVA for repeated data, and a two-level models.

Finally, the talk by *Chris Frost* described the use of multilevel models in summarizing the relationship of saturated fat, polyunsaturated fat and dietary cholesterol levels with blood cholesterol levels in a meta-analysis of 105 metabolic ward studies. It focused on difficulties in fitting and interpreting results from both multiple regression models and multilevel models.

New Book Advance Information

Multilevel Statistical Models, the 2nd edition of Multilevel Models in Educational and Social Research by Professor *Harvey Goldstein*, is due to publish in March 1995. This book discusses, and expands on, the core set of established techniques, experience and software packages, which are becoming more commonly used in areas such as education, epidemiology, geography, child growth and household surveys.

This second edition aims to integrate existing methodological developments, within a consistent terminology and notation, and avoids undue statistical complexity. Methodological derivations are provided in the appendices and examples and diagrams are used to illustrate the applications of techniques. It explains a number of new developments, especially in the areas of discrete response data, time series models, random cross classifications, errors of measurement, missing data and nonlinear models. This book is suitable for postgraduate courses and as a general reference tool and as such is extensively referenced.

A review of the book will be carried in the next issue of the newsletter.

For ordering information contact *Deborah Rowe*, Marketing Department, Edward Arnold, Hodder Headline PLC, 338 Euston Road, London NW1 3BH, UK. Tel: 44 (0)171 873 6358. Fax: 44 (0)171 873 6325. E-mail: I36JXS@Hodder.mhs.compuserve.com.

Research Fellows and Visitors

In 1994 the Multilevel Models Project received 2 research fellowships through the ALCD programme of the ESRC. They are:

Dougal Hutchison, National Foundation for Educational Research, Slough, fellowship period October 1994 - March 1995. His study interest is in multilevel modelling of data with measurement errors.

Toby Lewis, University of East Anglia, fellowship period October 1994 - March 1995. He is interested in data diagnostics for multilevel models.

In 1995 three fellowships have been received for the period October 1995 - March 1996. The fellows are:

Ian Langford, School of Health and Social Work, University of East Anglia. His interest is in analysis of spatial data with hierarchical structures.

Dick Wiggins, Social Survey Research Unit, City University. He will be working on multilevel models with missing data.

Alistair Leyland, Public Health Research Unit, University of Glasgow. He will be working on the analysis of health service statistics.

Magdalena Mok from School of Education, Macquarie University in Australia visited the project in the Spring term 1995. Her interest focuses on modelling design, in particular, the effects of sample size in educational studies.

Some New References to Multilevel Modelling

Babiker, A., & Cuzick, J. (1994). A simple frailty model for family studies with covariates. *Statistics in Medicine*, **13**, 1679-92.

Berkey, C.S., Hoaglin, D.C., Mosteller, F., & Colditz, G.A. (1995). A random effects regression model for meta-analysis. *Statistics in Medicine*, **14**, 395-411.

Fitzmaurice, G.M., Laird, N.M., and Lipsitz, S.R. (1994). Analysing incomplete longitudinal binary responses: a likelihood based approach. *Biometrics*, **50**, 601-12.

Follmann, D. (1994). Modelling transitional and joint marginal distributions in repeated categorical data. *Statistics in Medicine*, **13**, 467-77.

Goldstein, H. (1994). Multilevel Cross-classified models. *Sociological methods & research*, **22**, 364-375.

Goldstein, H., & Healy, M.J.R. (1995). The graphical presentation of a collection of means. *J. Royal Statistical Society, A.*, **158**, 175-7.

Grender,J.M., & Johnson, W.D. (1994). Fitting multivariate polynomial growth curves in two-period crossover designs. *Statistics in Medicine*, **13**, 931-43.

Kuk, A.Y.C. (1995). Asymptotically unbiased estimation in generalised linear models with random effects. *J.Royal Statistical Society, B.*, **57**, 395-407.

Li, Z., & Begg, C.B. (1994). Random effects models for combining results from controlled and uncontrolled studies in a meta analysis. *J.American Statistical Association*, **89**, 1523-1527.

Lipsitz,S.R., Laird, N.M., & Harrington, D.P. (1994). Weighted least squares analysis of repeated categorical measurements with outcomes subject to nonresponse. *Biometrics*, **50**, 11-24.

Longford, N.T. (1994). Reliability of essay rating and score adjustment. *Journal of Educational and Behavioural Statistics*, **19**, 171-200.

McDonald, R.P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological methods & research*, **22**, 399-413.

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

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

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

Sanders, W.L., & Horn, S.P. (1994). The Tennessee value-added assessment system (TVAAS): mixed model methodology in educational assessment. *Journal of personnel evaluation in education*, **8**, 299-311.

Ten Have, T.R., & Chinchilli, V.M. (1994). Bayesian hierarchical analysis of within-units variances in repeated measures experiments. *Statistics in Medicine*, **13**, 1841-52.

Tian, J.J., Shukla, R., & Buncher, R. (1994). On prediction of future observation in growth model. *Statistics in Medicine*, **13**, 2205-18.

Wang-Clow, F., Lange, M., Laird, N.M., & Ware, J.H. (1995). A simulation study of estimators for rates of change in longitudinal studies with attrition. *Statistics in Medicine*, **14**, 283-97.

Yang. M., & Leung, S.S.F. (1994). Weight and Length growth of two Chinese infant groups and the seasonal effects on their growth. *Annals of Human Biology*, **21**, 547-62.

Zerbe,G.O., Wu, M.C., & Zucker, D.M. (1994). Studying the relationship between change and initial value in longitudinal studies. *Statistics in Medicine*, **13**, 759-68.

Please send us any multilevel modelling publications for inclusion in this sectin in the future issues.

Theory and Applications

Applications in genetic epidemiology: modelling correlations within nuclear families using multilevel modelling

Paul Burton

Given recent advances in DNA technology, and the improved scope of genetic analysis that this has permitted, there has been an increasing emphasis upon collecting biological data from families rather than individuals. Unfortunately, the implications that this has for analysis are not always recognised and at least some of the sophisticated **DNA** work that is currently being carried out is being marred by inappropriate analysis at the final stage, that is, when attempts are made to estimate the quantitative association between newly identified mutations and phenotype. When the original data have been collected from nuclear families, even though principal interest may be centred upon the *fixed effect* of a known mutation, it is clearly essential to take account of the phenotypic correlations - arising from the effect of unknown genes and environmental factors - between siblings, between children and their parents and between parents. Unfortunately, one of the problems facing those molecular geneticists who have limited statistical support is that the existing software that allows you to model these correlations is, by and large, rather difficult to use. For example, when the phenotype is a continuous Gaussian response, genetic epidemiologists might currently model such data using *Fisher* (Lange, 1993) or one of the newer *Generalized Estimating Equation* programs such as *GEE4* (Liang et al, 1992; Hanfelt, 1993), but these approaches generally require a working knowledge of a programming language such as *Fortran*, *C* or *Pascal.* At the same time, construction of a model based upon *Path Analysis* in e.g. *LISREL* (Neale and Cardon, 1992) can prove cumbersome when there are nuclear families with a variety of different structures within a data set (e.g. 1 and 2 parent families with a variety of different numbers of children).

In view of the relative ease of utilisation of *ML3*, it therefore seemed useful to determine whether there was a way to model nuclear family correlations using *Multilevel Modelling.* Initially problems were encountered because of the non-nested nature of the correlation structure. However, it then became clear that the problem could be parameterised as a three level structure. Consider the following data set:

FAM ILY	GENERa tion	INDIVI Dual	PARE NT	MUTA TION	CONFOU ND	PHENOtype
1	1	1	1	0	7.36	-1.43
1	1	1	1	1	8.85	0.92
1	2	3	0	1	0.71	0.38
1	2	4	0	1	7.72	-1.88
1	2	5	0	1	6.55	-2.22
2	1	1	1	0	8.34	0.34
2	1	2	1	0	3.43	-4.16
2	2	3	0	1	3.11	-1.81
2	2	4	0	0	0.01	-4.92
2	2	5	0	1	7.45	-0.14
.	.	.	.	.	.	.
.	.	.	.	.	.	.

Interest centres upon the fixed effect of MUTATION on the *Gaussian* response variable PHENOtype having adjusted for the fixed effect of CONFOUND and for the different correlations of PHENOtype between siblings, between parents and their offspring and between parents. The GENERation variable divides each family into two generational subgroups, parents and children.

In order to fit an approriate model:

(i) Declare 'PHENO' to be the response variable y.

(ii) Declare 'CONS' (a vector of 1s), 'MUTATION', 'CONFOUND' and 'PARENT' as explanatory variables, written as x_0, x_1, x_2 and x_3 respectively.

(iii) Select 'FAMILY', 'GENER' and 'INDIVID' as the identifiers for levels 3, 2 and 1, to be indicated by k, j and i respectively.

(iv) Specify a variance term for 'CONS' at all three levels and covariance terms between 'CONS' and 'PARENT' at levels 2 and 1.

The fitted model can be written as follows,

$$y_{ijk} = \beta_{0jk}x_0 + \beta_1(x_1)_{ijk} + \beta_2(x_2)_{ijk} + \beta_{3jk}(x_3)_{ijk} + e_{ijk}$$

$$e_{ijk} = e_{0ijk} + e_{3ijk}(x_3)_{ijk}$$

$$\beta_{0jk} = \beta_0 + v_{0k} + u_{0jk}$$

$$\beta_{3jk} = \beta_3 + u_{3jk}$$

where v_k refers to the random residuals of a particular parameter at family level, u_{jk} refers to those at level 2 and e_{ijk} are those at level 1. The additional subscript for a term indicates the residuals of the parameter by that number.

In table 1 are listed parameters to be estimated for the model by means of the model specification in *ML3*. The variance terms of variable PARENT at both levels 1 and 2 are constrained to be zero. The reason for this is detailed later in the paper.

Table 1 Parameters both in fixed and random parts of the model

Parameter	Estimate
<u>Fixed</u>	
CONSTANT	β_0
MUTATION	β_1
CONFOUND	β_2
PARENT	β_3
<u>Random</u>	
<u>Level 3</u>	
CONS/CONS	$\sigma^2_{v_0}$
<u>Level 2</u>	
CONS/CONS	$\sigma^2_{u_0}$
CONS/PARENT	$\sigma_{u_{03}}$
<u>Level 1</u>	
CONS/CONS	$\sigma^2_{e_0}$
CONS/PARENT	$\sigma_{e_{03}}$

Having fitted the model one obtains estimates for the fixed and random coefficients as detailed in table 1. The required variances, covariances and correlations may then be estimated as follows:

Variances

Child: $\sigma^2_{v_0} + \sigma^2_{u_0} + \sigma^2_{e_0}$

Parent: $\sigma^2_{v_0} + \sigma^2_{u_0} + 2\sigma_{u_{03}} + \sigma^2_{e_0} + 2\sigma_{e_{03}}$

Covariances (within a family)

Child:Child $\sigma^2_{v_0} + \sigma^2_{u_0}$

Child:Parent $\sigma^2_{v_0}$

Parent:Parent $\sigma^2_{v_0} + \sigma^2_{u_0} + 2\sigma_{u_{03}}$

Correlations (within a family)

Child:Child $\left(\sigma^2_{v_0} + \sigma^2_{u_0}\right)/\left(\sigma^2_{v_0} + \sigma^2_{u_0} + \sigma^2_{e_0}\right)$

Child:Parent $\sigma^2_{v_0}/\left\{\left(\sigma^2_{v_0} + \sigma^2_{u_0} + \sigma^2_{e_0}\right) \times \left(\sigma^2_{v_0} + \sigma^2_{u_0} + 2\sigma_{u_{03}} + \sigma^2_{e_0} + 2\sigma_{e_{03}}\right)\right\}^{1/2}$

Parent:Parent $\left(\sigma^2_{v_0} + \sigma^2_{u_0} + 2\sigma_{u_{03}}\right)/\left(\sigma^2_{v_0} + \sigma^2_{u_0} + 2\sigma_{u_{03}} + \sigma^2_{e_0} + 2\sigma_{e_{03}}\right)$

Standard errors for these variances, covariances and correlations may then be obtained using the estimated variance-covariance matrix for the random effects (in *ML3* column C97) and the *Delta Method*. These calculations are all carried out using an *ML3* macro which operates on any model in which the random stucture is equivalent to that in table 1. The macro may be obtained from the author at email address: *paulb@ichr.uwa.edu.au*.

Table 2 details the results obtained from the analysis of a simulated data set (all relevant parameter values known) consisting of 10,000 families each containing 2 adults and 3 children. The simulations were carried out in Splus in a six stage process. (a) Initial parental responses (P_{j1}, P_{j2}) in family j were generated as random *N(0,2)*. (b) Initial responses for the three children in family j were specified as $1/2(P_{j1} + P_{j2})$. (c) A random effect common to all members of family j was generated as *N(0,1)* and added to the initial response in each individual. (d) A random effect common to all children in family j was generated as *N(0,4)* and added to the response in each child. (e) A random perturbation *N(0,3)* was added to the response in each individual. (f) Fixed effect vectors were generated, multiplied by their chosen coefficients and the resultant quantities added to the response in each individual. In this manner (conditional upon the fitting of an appropriate model to address the fixed effects) the following components of variance were simulated (the subscribed letters refer to the simulation stage at which each quantity was introduced):

Variances

Child = $1_{(a,b)} + 1_{(c)} + 4_{(d)} + 3_{(e)} = 9$

Parent = $2_{(a)} + 1_{(c)} + 3_{(e)} = 6$

Covariances (within a family)

Child:Child = $1_{(a,b)} + 1_{(c)} + 4_{(d)} = 6$

Child:Parent = $1_{(a,b)} + 1_{(c)} = 2$

Parent:Parent = $1_{(c)} = 1$

These values correspond to the simulated variances, covariances, and hence to the correlations, specified in the *Simulation* column of table 2. The coefficients chosen for the fixed effects are specified in the same column.

Table 2 Model estimation

Parameter	Simulation	Model Estimate	Model *S.E.*	Empirical *Mean*[a]	Empirical *S.E.*[a]
Fixed effects					
CONS	0.0	0.044	0.037	0.045	0.039
MUTATION	1.0	0.975	0.024	0.977	0.023
CONFOUND	0.2	0.196	0.0033	0.196	0.0033
PARENT	0.0	−0.021[b]	0.026	-0.022	0.022
Variances					
Child	9.00	9.01	0.101	8.96	0.104
Parent	6.00	6.04	0.061	6.01	0.063
Covariances					
Child:Child	6.00	6.01	0.100	5.97	0.107
Child:Parent	2.00	1.98	0.053	1.96	0.047
Parent:Parent	1.00	1.00	0.061	0.98	0.063
Correlations					
Child:Child	0.67	0.67	0.0045	0.66	0.0051
Child:Parent	0.27	0.27	0.0061	0.27	0.0054
Parent:Parent	0.17	0.17	0.0097	0.16	0.0100

a Empirical mean estimate and standard error of the mean based upon 50 simulations of 200 families.

b The parameter estimate for PARENT could be constrained to 0.

In every instance (see table 2) the model-based standard error is very similar to the empirical standard error of the mean estimate obtained from 50 simulations of 200 families. Furthermore, all of the estimated parameter values are within 1.28 standard errors of the true simulated values and are similar to the empirical means from the 50 simulations.

For theoretical reasons (in genetic epidemiology), the covariance between two children should normally equal or exceed the covariance between a child and a parent which should normally equal or exceed the covariance between two parents. In consequence, it is generally necessary for the CONS/PARENT term $\sigma_{u_{03}}$ in table 1 to be negative and it must therefore be parameterised as a covariance rather than as a variance.

If the covariance term $\sigma_{e_{03}}$ at level 1 in table 1 is excluded, twice the covariance term at level 2 ($2\sigma_{u_{03}}$) has to simultaneously estimate the difference between the variance of a child and the variance of a parent *and* the difference between the covariance of a parent with a parent and the covariance of a child with a child. This can lead to a poorly fitting model and is not to be recommended.

One problem that can occur is that one or more of the variance parameters may be constrained to zero. This will happen, for example, if the observed child:child covariance is *less* than the observed child:parent covariance; for then $\sigma^2_{u_0}$ (CONStant at level 2) must take a negative value to properly model the data. This will not generally happen in a data set which is well behaved genetic epidemiologically, however it is a problem that does sometimes occur in practice, particularly in small data sets. When it does arise, the problem may be consequent upon a misspecification of the fixed effects model and this can sometimes be remedied. However, if no explanation can be found and the constraint proves impossible to circumvent, the parameter estimates from the constrained model are likely to be misleading and in such a case it is advisable to proceed to an alternative analysis using, for example, *GEE4*.¶

This proposed approach can clearly be generalized to situations with any number of fixed covariates. It works equally well when there are a variety of different nuclear family sizes and compositions. Missing data are not a problem provided the mechanism of information loss is uninformative. It would also be straightforward to analyse a standard twin study (dizygous and monozygous twins reared together or apart) using *ML3* and once *MLn*, which deals with more than 3 levels becomes available (see announcement in the Newsletter), more complex pedigree structures will also be able to be modelled. We are currently investigating the use of *ML3* in the modelling of nuclear family correlations when the phenotype is binary.

Although this type of analysis is principally recommended for use in situations where the correlation structure is effectively a nuisance, it can also be used when the correlation structure is of primary interest which is often the case in genetic epidemiology. The *ML3* macro (see above) also generates estimates of σ_A^2 (additive genetic variance), σ_D^2 (dominance genetic variance), σ_{CS}^2 (variance arising from common sibling environment) and σ_E^2 (variance arising from unshared environment) which are of interest to genetic epidemiologists in their own right and can, when it is theoretically appropriate, be used to estimate heritability. (Note that σ_D^2 and σ_{CS}^2 are completely confounded in a study based upon nuclear families alone and cannot therefore be interpreted simultaneously).

References

Hanfelt J. *GEE4 documentation*. Johns Hopkins University, Baltimore, 1993.

Lange K, Boehnke M, Weeks D. *Programs for pedigree analysis: documentation for Fisher version 2.1*. UCLA, Los Angeles, 1993.

Liang K-Y, Zeger SL, Qaqish B. *Multivariate regression analyses for categorical data*. J R Statist Soc series B 1992;54:3-40.

Neale MC, Cardon LR. *Methodology for genetic studies of twins and families*. Kluwer Academic Publishers, Dordrecht, 1992.

¶ Editors' note: *MLn* will allow negative variance estimates.

Additive, Multiplicative and Generalized Relative Risk Multilevel Models

Kevin Pickering & Andrew Pickles

Introduction

The choice of link function for a binary response model is rarely given much thought, and can sometimes seem a rather esoteric question. It is none the less important. To provide some insight into the problem, we begin with a heuristic example.

Figure 1 illustrates diagrammatically two ways in which two risk factors X_1 and X_2 might act in generating a positive value for a binary response. Example response measures include high attainment in some test, ill-health or, as in our later example, the taking up of smoking. In the one-stage model there is a baseline rate that applies in the absence of the two risk factors and this is represented by the centre route or arrow. There are also two other routes, each associated with a particular risk factor. The diagram suggests that in general the rate of positive response will depend upon the addition of rates through the baseline route and those risk factor routes to which an individual is exposed. In the two stage model, there is an intermediate stage that we do not directly observe. At each stage there is a baseline route and a single risk factor route. Risk factor X_1 provides a route to the intermediate stage only and has no effect thereafter. Correspondingly, although others may be exposed to risk factor X_2, the effect of X_2 is restricted to those who have reached the intermediate stage. The two risk factors now act multiplicatively and, in comparison to the one-stage model, the two-stage multiplicative model implies a degree of synergism in the effects of risk factors. Thus if X_1 increases the baseline rate by a factor of 4, while X_2 increases it 3 times, when both are present the one stage, additive model will imply a rate 7 times the baseline while the two-stage multiplicative model will imply a rate 12 times the baseline. If one or both risk factors operated on both steps of the two stage process then the effects would combine somewhere between additivity and multiplication.

Largely because of its elegant statistical properties the common link function used in analysing binary response data is the logistic function. Implicit in the use of this link function is that the effects of explanatory variables combine multiplicatively, as in the two stage model. The alternative additive model is obtained with an identity link function.

Breslow and Storer (1985) suggested a generalized link function that ranged from the additive to the multiplicative (and beyond) by varying the exponent in a power transformation. Using this model one can assess the suitability of additive and multiplicative models, and try to search for an 'optimal' model that best explains the data. As suggested by the heuristic example, such a search may be instructive about the structure of the process being analysed. We have implemented a multi-level form of this *generalized relative risk* (GRR) model in *ML3*.

The dataset

The data used in the analysis are from an OPCS study of children's smoking behaviour. Children from 32 secondary schools completed questionnaires in three waves between 1986 and 1988, corresponding to their second, third and fourth years at school (Goddard, 1990). The response variable used in the analyses was whether the child was a smoker or not at the third wave and hence is binary. We fitted two factors; whether one of the parents smoked, and whether the child had tried a cigarette by the first wave. Including only children with complete data for the variables of interest and those with both parents living at home gave us a reduced dataset of 2926 children.

The model

The GRR model suggested by Breslow and Storer is

$$logit(\pi) = \alpha + \frac{(1+x\beta)^{\lambda}-1}{\lambda} \qquad \lambda \neq 0$$

$$logit(\pi) = \alpha + \log(1+x\beta) \qquad \lambda = 0$$

This model gives a logistic model when the value of the power parameter, λ, is 1, and is equivalent to a model with an identity link function when $\lambda = 0$.

Because the children in the study are grouped within schools, a two-level hierarchical design was required to account for the variation between schools and children. The random effects for schools operate on the logit rather than the probability itself, giving a penalized quasilikelihood model (PQL) (Breslow & Clayton 1993). In this example the model is:

$$y_{ij} = \mu_{ij} + e_{ij}$$

where

$$logit(\mu_{ij}) = \alpha + u_j + \frac{(1+x_{ij}\beta)^{\lambda}-1}{\lambda} \qquad \lambda \neq 0$$

$$logit(\mu_{ij}) = \alpha + u_j + \log(1+x_{ij}\beta) \; \lambda = 0$$

This model was fitted in *ML3* using a macro for the generalized multilevel relative risk (GMRR) model. These macros follow the generalized linear modelling procedure suggested by Goldstein (1991), that uses a weighted iterated GLS algorithm to fit a specified variance function and link function.

Results

The parameter estimates for the models fitted are shown in Table 1. In the additive model the interaction term was positive and significant, but in the multiplicative model the interaction was non-significant. These two observations are consistent and suggest that there was synergy between the two risk factors. Since these models were saturated they gave the same level of fit as given by the criterion, and the same predicted probability (obtained by cross-tabulation of the predicted values including random effects) as shown in Table 2.

Table 1. Parameter Estimates for the Fitted Models (standard error in parenthesis)

Saturated Models

	Add.	Multip.	λ = 0.6
Fixed	Estimates (SE)	Estimates (SE)	Estimates (SE)
Baseline	-2.637 (0.145)	-2.637 (0.145)	-2.637 (0.145)
Parent smokes	0.819 (0.304)	0.598 (0.167)	0.667 (0.205)
Tried smoking by wave 1	5.156 (0.985)	1.817 (0.160)	2.418 (0.262)
interaction	3.374 (1.117)	-0.079 (0.212)	0.224 (0.293)
Random			
School	0.173 (0.066)	0.173 (0.066)	0.173 (0.066)
Criterion	1810.31	1810.31	1810.31

Models without interaction

	Add.	Multip.	λ = 0.6
Fixed	Estimates (SE)	Estimates (SE)	Estimates (SE)
Baseline	-2.677 (0.146)	-2.611 (0.126)	-2.681 (0.135)
Parent smokes	1.029 (0.331)	0.550 (0.104)	0.770 (0.159)
Tried smoking by wave 1	7.049 (1.093)	1.817 (0.160)	2.561 (0.186)
Random			
School	0.163 (0.063)	0.174 (0.066)	0.171 (0.065)
Criterion	1824.49	1813.26	1808.24

To find an optimal link function, the interaction between terms was dropped and the value of the power parameter that minimized a goodness-of-fit criterion derived for multivariate normal data and minimized by *ML3* gave the plot shown in Diagram 2. The optimal model had a value of the power parameter near 0.6, representing a link function intermediate between additive and multiplicative. However, the half-awake reader will have noticed that the criterion value for the optimal model indicates that this model gave a better fit than the saturated model! An explanation for this is being investigated.

Table 2. Predicted Probabilities for Fitted Models

Parent smokes	No	No	No	No
Tried smoking by wave 1	Yes	Yes	Yes	Yes
Observed	0.0684	0.3151	0.1163	0.4252
Predicted				
Saturated model	0.0684	0.3151	0.1163	0.4252
Models without interaction				
Multiplicative	0.0700	0.3313	0.1140	0.4280
$\lambda = 0.6$	0.0657	0.3243	0.1204	0.4182
additive	0.0657	0.3642	0.1236	0.3859

Discussion

The non-significant interaction in the multiplicative model suggests that the two factors of interest may combine in a two-stage process. Perhaps the increased risk arises from an early willingness to try smoking, that is exacerbated by the ongoing availability of cigarettes that having a smoking parent presents.

We should, however, emphasize two things. Firstly, the additive linear model with an interaction and the multiplicative/logistic model without an interaction both gave very similar fitted values. The presence of a significant interaction in one and not in the other should not result in a different interpretation of the data; they merely represent the same finding but with a somewhat different emphasis. Secondly, one and two stage models are just one of several ways in which additive or multiplicative effects may come about. Other derivations, such as threshold models, are also possible (Rutter and Pickles, 1991).

The model we estimated included just a single school random effect within a PQL approach. To avoid the possibility of negative predicted rates these are not directly subject to the power transformation (see earlier equations) but are placed within the GRR link function as though in a standard logistic link. Thus random effects associated with a more complex multi-level structure will always be combined multiplicatively, regardless of what the value of power parameter implies for the combining of fixed effects. This asymmetry in the treatment of random and fixed effects appears to be inevitable when using PQL with link functions that lack natural 0-1 bounds. In MQL the random effects act outside the link function and thus combine additively, regardless of how the fixed effects are combined within the link function.

The GMRR PQL and MQL model macros are available from the first author. The work has been funded under the ESRC's research programme into the Analysis of Large and Complex Datasets (ALCD).

References

Breslow, N.E. & Clayton, D.G. (1993). Approximate inference in Generalized linear mixed models. *J. Am. Statist. Assoc.* **88**, 9-25.

Breslow, N.E. & Storer, B.E. (1985). general relative risk functions for case-control studies. *Am. J. Epidem.* **122**, 149-162.

Goddard, E. (1991). *Why Children Start Smoking.* HMSO, London.

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

Rutter, M. & Pickles, A.R. (1991). Person-environment interactions: concepts, mechanisms, and implications for data analysis. In T.D. Wachs & R. Plomin (eds), *Conceptualization and Measurement of Organism-Environment Interaction*, American Psychological Association, Washington, PP.105-141.

Use of Multi-Level Modelling in Procedures for Maximising Between-School Comparability of Final Year School-Based Assessments

Peter W. Hill and Kenneth J. Rowe

At the Year 12 level in the State of Victoria, as in all other Australian state education systems, there is widespread recognition of the value of ensuring that final assessments are based on a mix of subject-based examinations and structured school assessments. In particular, school-based assessment is seen as a means of allowing assessment of the broadest possible range of valued outcomes and of improving the validity of final assessments. The use of school assessments within a high stakes environment does, however, raise significant issues regarding the comparability of those assessments.

In the past, these concerns were addressed largely through 'statistical moderation', a process whereby final assessments were derived from a weighted combination (usually 50:50) of school assessments and external examination results, after the location and spread of the school assessments had been adjusted to the location and spread of the respective examination results of students within that school. Statistical moderation had a number of drawbacks, however, of which the following were the most serious:

- Examination scores were not always seen to be valid moderator variables, particularly in cases where school assessment has been specifically designed to measure outcomes which were not or could not be assessed through external examination.
- It encourage schools to focus all their efforts on maximising scores on the examination to the neglect of school assessment. There was no incentive for the school to put significant effort into the school assessment, since, regardless of the actual standard of performance, students' assessments was automatically adjusted to coincide with their performance on the examination.

With the introduction in 1991 of the new Year 12 Victorian Certificate of Education (VCE), the mix of external examinations and school-based assessment was maintained, but statistical moderation was abandoned in favour of a new approach to comparability involving three key elements:

1. *Common Assessment Tasks* To provide greater structure to school-based assessment, common assessment tasks were introduced. For example, one of the common assessment tasks that students do as part of English requires them to present a folio of selected pieces of writing. The Writing Folio is completed according to specifications which describe the purpose and details of the task, the conditions under which it will be completed and the procedures the school will follow to authenticate that the completed work is indeed that of the student. For most subjects there are two school-assessed common assessment tasks.
2. *Criteria for the award of marks* To ensure that teachers mark common assessment tasks consistently, detailed attention was given to the development of criteria to be used to assess student work. These criteria were incorporated into rating scales for each common assessment task which were empirically evaluated.
3. *Verification of schools' assessments* To check on whether teachers in schools consistently apply the criteria, a system of verification based on checking of samples of work from each school by an external reviewer was introduced. If the verification process identified discrepancies between a school's assessments and those of the external reviewer, the external reviewer had the capacity to make changes to that school's initial assessments.

The third element in the above arrangements, namely the verification process, proved cumbersome and inefficient. Following reports by Brown and Ball (1992) and Hill, Brown & Masters (1993), a new system for undertaking comparability checks of school-based assessments was introduced. The new system uses *ML3-E* (Prosser, Rasbash & Goldstein, 1993) to undertake a series of statistical checks on the reasonableness of schools' assessments and on the basis of these checks schools' assessments are either confirmed or subject to follow-up by external reviewers.

Mid-way through the academic year, all Year 12 (final year) students take a *General Achievement Test (GAT)*. Details regarding the *GAT* are published by the Board (Board of Studies, 1994a). Following marking of students' responses, three sub-scores are computed for each student, namely written communication (*G1*), mathematics, science, technology (*G2*), and humanities, arts, social sciences (*G3*). These three sub-scores are used to construct composite scores which best predict students' results on each school-assessed task. A series of statistical tests are then used to check on the reasonableness of schools' numerical assessments of students. Where the assessments of a school are within the expected range given their weighted *GAT* scores, the school's assessments are confirmed without adjustment. Where there are substantial discrepancies, two external markers review the school's assessments of students' work. As a result of this re-marking, a school's assessments may be confirmed, adjusted upwards or adjusted downwards.

Details of the statistical checks undertaken to identify schools for follow-up are contained in a technical bulletin published by the Board (Board of Studies, 1994b). The aim of the statistical checks is to identify unexpected patterns of results given students' scores on the *GAT*. In particular, the checks aim to identify schools with unexpectedly high or low scores on the school-assessed common assessment task, or scores that are unexpectedly bunched together, or spread out. Schools with less than five students are removed from the statistical checking procedure and subject to automatic follow-up, as are schools for which more than 20 per cent of students do not have usable *GAT* scores.

For each school-assessed task, a weighted combination of the three *GAT* scores (*G1*, *G2* and *G3*) is computed as:

$$x_{ijk} = b_{1k}G_{1ijk} + b_{2k}G_{2ijk} + b_{3k}G_{3ijk} \qquad (1)$$

where x_{ijk} is the weighted composite *GAT* score for each student i, in school j, on common assessment task k, and b_{1k}, b_{2k}, b_{3k} are level 1 (student) regression coefficients obtained using a two-level (student and school) regression model to regress scores on the school-assessed task (y_{ijk}) on the three *GAT* scores for the population of students for whom data are available. This means that for each school assessed task, scores on the *GAT* are combined differently and in a way which maximises the correlation with the school assessed task. For example, in the case of the Writing Folio in English, the above procedure led in 1994 to the following weights being used to combine the three *GAT* scores:

Written expression (*G1*)	0.314
Mathematics, science, technology (*G2*)	0.065
Humanities, arts, social sciences (*G3*)	0.281

These weights reflect the nature of the abilities underlying the Writing Folio. Different patterns of weights were obtained for other school assessed tasks in other subjects.

A check is made on the correlation of the predicted school assessment (obtained from the weighted combination of the three *GAT* scores and adjusted for differences in intercepts among schools) with the actual score provided by the school (y_{ijk}). If the correlation between these two scores falls below 0.45, no further statistical checks are undertaken and all work for that task is subject to follow-up. In the case of the Writing Folio in English, a value of r=0.705 was obtained in 1994.

The first of the statistical tests involves estimating school-level residuals obtained from fitting a two-level regression model to the data which assumes that the observed scores of individual students as assessed by the school can be predicted by a knowledge of their expected level of performance as indicated by a weighted combination of their scores on the *GAT*, and of the identify of the school providing the initial assessment. The model can be written out in two parts. First, the relationship between the school-assessed task and the weighted *GAT* score can be expressed as:

$$y_{ijk} = a_{jk} + b_k x_{ijk} + e_{ijk} \qquad (2)$$

defining:

y_{ijk} as the score on the school-assessed task for student i in school j for task k;

a_{jk} as the intercept for school j, task k;

b_k as the slope of the regression line for predicting the y_{ij} for task k;

x_{ijk} as the weighted *GAT* score for student i, in school j, for task k and

e_{ijk} as the residual or unique contribution of student ijk.

Second, the relationships between the scores at the level of schools can be expressed as

$$a_{jk} = a_k + u_{jk} \qquad (3)$$

defining:

a_k as the mean of means for all schools (constant term) on task k; and

u_{jk} as the residual or unique contribution of school j on task k, beyond that explained by the constant term a_k.

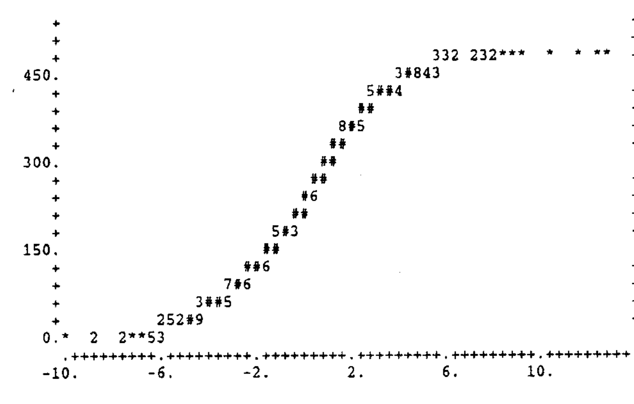
Combining the above into a single equation gives:

$$y_{ijk} = a_k + b_k x_{ijk} + u_{jk} + e_{ijk} \qquad (4)$$

This model is fitted to the data for each school-assessed task. A check is made on the reasonableness of the assumption that the coefficient b_k should be fixed at the same value for all schools. In 1994, this assumption was found to hold across the full range of school-assessed tasks.

Having fitted the two-level regression model of equation (4) to the data for a particular school-assessed task, the focus of interest is on the estimate of the residual term for each school ($\hat{u}_{jk}$), because this indicates how discrepant a school's scores are from there which would be predicted by the model. These residuals are standardised by dividing them by the square root of their respective comparative variances and schools are sorted on the basis of the magnitude of these standardised school-level residuals. Large negative residuals indicate schools with assessments below expectation and large positive residuals indicate schools with higher than expected assessments.

Cut-offs are determined by visual inspection of a plot of ranked residuals as shown below for the Writing Folio in English. This plot shows ranked residuals by residuals for 497 schools. Schools with unexpected patterns of results are deemed to be those located in the top and bottom tails of the graph where the plot flattens out into a horizontal line.



Thereafter a supplementary statistical test is applied to ascertain whether the spread of scores provided by the school for a given task differs significantly from the spread of the corresponding weighted *GAT* scores. The test used is a test for equality of variances, given by the formula:

$$vd_{jk} = \frac{\left(s_{y_{jk}}^2 - s_{x_{jk}}^2\right)\sqrt{n_{jk} - 2}}{2s_{y_{jk}}\sqrt{1 - r_{jk}^2}} \qquad (5)$$

defining:

vd_{jk} as the test statistic for equality of variances for school j, task k;

$s_{y_{jk}}^2$ as the variance of the scores on school assessment for school j, task k;

$s_{x_{jk}}^2$ as the variance of the weighted *GAT* score for school j, task k;

n_{jk} as the number of students in school j with both a school assessment and a *GAT* score for task k; and

r_{jk} as the correlation coefficient between the school assessment and the weighted *GAT* score within school j, for task k.

This test, was programmed using the utility commands within *ML3-E*. Only a small number of schools are typically identified by this test and these are added to those identified through the estimating and plotting of school-level standardised residuals.

In 1994, the new procedure resulted in between 15 to 20 per cent of schools being followed up for each common assessment task. In the case of the Writing Folio in English, a decision was made to follow up 92 out of 513 schools. This means that the assessments as provided by 82 per cent of schools were confirmed, while the assessments of the remaining 18 per cent of schools were subject to the follow-up by external reviewers.

The new statistical checking arrangements have significantly reduced the workload and costs associated with the previous verification procedures and enable a sharper focus on a smaller number of schools where there is *a priori* evidence that the application of resource-intensive moderation processes is justified. Under the previous approach there was follow-up of samples of student work in *all* schools. Under the new approach, follow-up by external reviewers is invoked only when there is evidence of potential discrepancies. In practice, the new procedures led in 1994 to about half the number of pieces of student work requiring re-marking by external markers of the corresponding figure for 1993.

ML3-E is used as the main software package in the entire statistical checking process, although supplemented by other packages for the purposes of generating specialist management reports. A large number of reports are generated and special-purposes files created, including graphical displays and tables. The analysis of a single common assessment task using *ML3-E* involves 263 lines of commands, of which 177 have been incorporated into 15 macros. The process is thus not fully automated as in statistical moderation and there is need for judgements to be made throughout the process by persons fully competent in the statistical procedures employed.

Despite the size of many of the data files (for example, the data file for the Writing Folio in English contained data for some 41,000 students in 513 schools) and the time required to compute standardised residuals, it is possible to process the entire results for the State on three desk-top computers within a matter of days.

Further details regarding the above approach to establishing comparability of Year 12 school-based assessments are contained in HIll, Brown, Rowe and Turner (1994).

References

Board of Studies (1994a). *More About the GAT: The General Achievement Test 1994*. Melbourne: Board of Studies.

Board of Studies (1994b). *General Achievement Test Technical Bulletin*. Melbourne: Board of Studies.

Brown, T., & Ball, S. (1992). *A report on the VCE verification process*. Melbourne: Victorian Curriculum and Assessment Board.

Hill, P.W., Brown, T., & Masters, G.N. (1993). *Fair and authentic school assessment: Advice to the Board of Studies on verification, scaling, and reporting of results within the VCE*. Melbourne: Board of Secondary Education.

Hill, P.W., Brown, T., Rowe, K.J., & Turner, R. (1994). *Establishing comparability of Year 12 school-based assessments*. Paper presented at the 1993 annual conference of the Australian Association for Research in Education, Fremantle, Western Australia, November 22-25, 1993.

Prosser, R., Rasbash, J., & Goldstein, H. (1993). *ML3-E software for three-level analysis (Version 2.3)*. Multilevel Models Project, Institute of Education, University of London.

Contributors

We are grateful to the following who provided articles or information for this issue.

Paul Burton, Institute for Child Health Research, Princess Margaret Hospital for Children, Roberts Road, Subiaco, WA 6008, Western Australia.

Linda Croxford, Centre for Educational Sociology, University of Edinburgh, 7 Buccleuch Place, Edinburgh EH8 9LW, UK.

Peter W. Hill & Kenneth J. Rowe, Centre for Applied Educational Research, The University of Melbourne, Parkville, Victoria 3052, Australia.

Ian Langford, Health Policy and Practice Unit, University of East Anglia, Norwich NR4 7TJ, UK.

Alistair Leyland, Public Health Research Unit, University of Glasgow, 1 Lilybank Gardens, Glasgow G12 8RZ, UK.

Kevin Pickering & Andrew Pickles, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK.

Simon Thompson, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK.