CHAPTER 13

Software for Multilevel Analysis

Jan de Leeuw
UCLA Department of Statistics, Los Angeles, CA, USA

Ita G.G. Kreft
Health and Human Services, CSLA, Los Angeles, CA, USA

13.1 INTRODUCTION

In this chapter we review some of the more important software programs and packages that are designed for, or can be used for, multilevel analysis. These programs differ in many respects. Some are parts of major statistics packages such as SAS or BMDP. Others are written in the macro language of a major package. And some are stand-alone special-purpose programs that can do nothing but multilevel analysis. We have been involved in a number of these comparisons before. The first (Kreft *et al.*, 1990), comparing HLM, ML3, VARCL, BMDP5-V and GENMOD, was published in Kreft *et al.* (1994). The second comparison (van der Leeden *et al.*, 1991), comparing HLM, ML3 and BMDP5-V on repeated measures data, was published in van der Leeden *et al.* (1996). We give both the reference to the internal report version and to the published version, because the unpublished version usually has much more material. Giving both references also shows the unfortunate time interval between the two, which is especially annoying in the case of software reviews. The reviews were summarised briefly in our book (Kreft and de Leeuw, 1998, Section 1.6).

Since our last publication on the subject, there have been many major changes. The program GENMOD, which was never easy to obtain, has more or less completely disappeared. VARCL, which was one of the leading contenders in the early 1990s, is no longer actively supported or developed, which means that it has rapidly lost ground. BMDP, as a company, went out of business, which had serious consequences for its software products. Programs such as HLM, written originally for DOS, were upgraded for Windows. ML3

Multilevel Modelling of Health Statistics Edited by A.H. Leyland and H. Goldstein

transformed to MLn, which transformed to MLwiN. And so on. Obviously, it is time for an update. In this update, we shall change our focus somewhat. We shall indicate what the programs can do, where you can get them, for how much money, on what systems they run, and how easy it is to use them. We do not emphasise computation speed, because this is hard to define, and not very relevant in most applications anyway. Computing time is usually infinitely small compared with the time needed to collect and clean the data.

13.1.1 Programs, macros and packages

Throughout this chapter, we shall use a number of classifications of the software we review. Of course the boundaries between the categories are somewhat fuzzy.

Research versus production Some software is written as a research tool, to analyse the examples in a research paper, or to analyse the data sets in a research project. Other software is written as a commercial product, or at least it is clearly intended for general use. Many programs start out in the research phase, and are subsequently promoted to the production phase. Many programs are still somewhere in between.

Stand-alone versus module There are multilevel programs that stand on their own, i.e. they are executables and they do not require other software (except the operating system, of course) to be present. Other programs are modules of an existing package. Usually they require the rest of the package to be present as well.

Program versus macro For programs written on top of an existing package, there are still two possibilities. Either the program is a module, existing in object code within the package, or the program is a macro, written in the scripting or extension language of the package. Macros are handled by an interpreter that is part of the package.

Another distinction that we make is between minor and major specialised programs. This is somewhat subjective, but the general idea is that a research program, usually with a rather primitive interface, is a minor specialised program. As soon as the authors start thinking about user-friendliness, making a nice graphical user interface, adding options that people have asked for, and going commercial, the minor program will become a major one. It will become bigger, have more possibilities, look better, and cost more.

13.1.2 Omissions

There are some programs that can be used to perform multilevel analysis, but that we do not discuss in detail. We just mention these programs briefly here, and we give the URL in case readers want to know more. (The `http://` part of the URLs have been omitted throughout.)

TERRACE This is multilevel research software, written by James Hilden-Minton for his PhD thesis. It is an add-on to the XLISP-STAT package, and it can be found (with manual) at
`www.stat.ucla.edu/consult/paid/nels/papers`

NLME The NLME software comprises a set of S-plus functions, methods and classes for the analysis of both linear and nonlinear mixed-effects models. It extends the linear and nonlinear modelling facilities available in release 3 of S-plus. It was written by José C. Pinheiro and Douglas M. Bates. It is available for Unix and Windows platforms from
`franz.stat.wisc.edu/pub/NLME`

BUGS This is a piece of computer software that permits the analysis of complex statistical models using Markov-chain Monte Carlo methods. The emphasis is on the Monte Carlo method, and a great variety of multilevel models can be analysed as well (see, e.g. Chapter 9 on institutional performance). For further details, see
`www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml`

Oswald This was developed by the Statistics Group at the University of Lancaster, and is a suite of S-plus functions for analysing longitudinal data. It includes mixed effects models, and many other possible options.
`www.maths.lancs.ac.uk/Software/Oswald/`

13.2 MAJOR SPECIALISED PROGRAMS

There are two major specialised programs for performing multilevel analysis – one from the UK and one from the USA.

13.2.1 MLwiN

The ML series of programs has a complicated history. The series was erected on top of the NANOSTAT program by Michael Healy. NANOSTAT is a general-purpose statistics program. The multilevel extensions started with ML2 in 1988, ML3 was introduced in 1990, and the final DOS program in the series was MLn, published in 1995. In 1998, they were all superseded by MLwiN. One can think of MLwiN as a separate program, but also as a graphical user interface on top of MLn. Throughout the project, most of the programming was done by Jon Rasbash, but clearly the program is the result of a team effort.

MLwiN contains the NANOSTAT package, so it can do a fair amount of data manipulation and general purpose statistics. This is all meant to assist in the multilevel analysis, and thus we still think of MLwiN as a specialised program.

In our report (Kreft *et al.*, 1990) we looked at ML2, and in the published version (Kreft *et al.*, 1994) at ML3. In a subsequent comparison, the internal

report version (van der Leeden *et al.*, 1991) looked at ML3, and the published version (van der Leeden *et al.*, 1996) at MLn.

We shall only discuss MLwiN in this chapter, since it includes all previous programs as special cases, and since most people seem to feel that putting a Windows interface on top of a DOS program is a step ahead.

Availability The Multilevel Models Project has three mirror homepages at
`www.ioe.ac.uk/multilevel/`
`www.medent.umontreal.ca/multilevel/`
`www.edfac.unimelb.edu.au/multilevel/`
The websites have information about the software, but also about the project and its activities. The MLwiN program has its own homepage at
`www.ioe.ac.uk/mlwin/`
The price of a single copy given on the MLwiN website is £500 ($900), although there is a 40% educational discount.

Documentation There is a large amount of documentation available, most of it is can be obtained from the multilevel project. Again, the various versions of the program lead to a somewhat bewildering variety of documents. Basically, there are three types of documents available for most versions of the program. The first document is the user's guide. For MLwiN, this is Goldstein *et al.* (1998). There is another type of user's guide, which is more introductory. For MLn, this is Woodhouse (1995). The second is the advanced manual, which discusses macros to fit more complicated models. This is Yang *et al.* (1998). Thirdly, there is the command reference. Since the command language for MLwiN is actually MLn, this is Rasbash and Woodhouse (1995).

In a sense, the most important document in the Multilevel Models Project is the book by Harvey Goldstein. Not surprisingly, this also exists in two very different editions. It developed with the project. The first edition (1987) covers basic multilevel analysis, with emphasis on applications in education. The second edition (1995) is much more statistical, and covers a large variety of extensions of the basic model and a much broader range of applications.

Model The basic model, in the case of two levels and p predictors, is

$$y_{ij} = \beta_{0ij}x_{0ij} + \beta_{1ij}x_{1ij} + \ldots + \beta_{pij}x_{pij}, \tag{13.1}$$

where

$$\beta_{sij} = \beta_s + u_{sj} + e_{sij}. \tag{13.2}$$

It is easy to see how this generalises to more than two levels. Observe that usually we have $x_{0ij} = 1$ for all i, j, i.e. the zero term in the regression corresponds to the intercept. Also observe that each regression coefficient has a fixed part and a random part, and the random part has random components for both levels.

Interface As indicated above, MLwiN provides (or maybe we should say 'is') a Windows interface on top of MLn. It is possible to use MLwiN very much

like MLn, because there is a command window in which MLn commands can be entered. This window also tracks the command history. In fact, menu commands are translated to MLn commands in the command window. There are many MLn and NANOSTAT commands that *cannot* be entered from the menus.[1]

An interesting feature of the program is the *equations window*. This is a specialised equations editor, in which the model can be defined in equation form by clicking and pasting. Obviously, there is no way to do this in DOS, so this defines a clear distinction between MLn and MLwiN. On the other hand, it is unclear if this is actually superior to written output that users themselves have to translate into a formula. It seems that if users know what they are doing, they can make the translation very easily themselves. If they don't know what they are doing, then they should not be using the program in the first place!

Algorithm MLwiN uses the IGLS or the RIGLS algorithms first described respectively by Goldstein (1986) and Goldstein (1989b). The algorithms are block-relation algorithms. There are two blocks of parameters: the fixed regression coefficients and the variance/covariance components. The algorithm fixes the variance components at some initial value and maximises the likelihood over the fixed coefficients. This is just a generalised least squares problem. Then it fixes the coefficients at their current values and maximises the likelihood over the variance components, by solving another, more complicated, generalised least squares problem. The two optimisations are alternated until convergence. A concise description appears in Goldstein and Rasbash (1992).

It is not entirely clear from the documentation what happens in boundary cases when dispersion matrices become singular or even indefinite.

Extensions By 'extensions' we mean various options and additions that do not really belong to the core of the program, but that the authors have added because of user demand, competitive pressure or their research program. Some extensions are obviously more useful for the general public than others, but most of them are at least interesting enough to be mentioned.

In MLwiN, there are two 'levels' of extensions. The first are features that are part of the program core. They can be handled by using the menus, or the MLn language, but they are features that will not often be needed. The second level consists of true extensions, written in the macro language provided by the package, and these are add-ons that the user may or may not load.

The first class of extension is discussed in the user's guide (Goldstein *et al.*, 1998). This already covers a substantial number of procedures. Hierarchical generalised linear models with binomial or Poisson outcomes can be fitted. Markov-chain Monte Carlo methods are available to optimise complicated likelihoods or compute complicated posterior distributions. Parametric

[1] Editor's note: The current release of MLwiN, v1.10.000 at the time of going to press, has many more MLn/NANOSTAT comands included in the menus.

bootstrap methods are used for bias correction and for standard error computation.

These extensions may be exciting to some, but they do take MLwiN several steps in the direction of research software. Users have to take very many things for granted, and have to hope that the default values of the many parameters and tuning constants work in their case. It seems a bit too demanding to ask casual users to choose between the MQL and PQL methods for quasi-likelihood estimation (see Chapter 3 for a discussion of these estimation procedures), or to choose an appropriate burn-in period for their MCMC procedure. It is true that with these extensions, models can be fitted that could not be fitted by older versions of MLn. But, for most people, it becomes impossible to understand what is actually going on inside the program, and to explain why certain choices and not others were made.

The macro-based extensions of MLwiN allow for even more flexibility. Macros are available for fitting multicategory models (see Chapter 8), survival models (see Chapter 12), time-series models (see Chapter 2) and nonlinear models (see Chapters 3 and 4). While these extensions are undoubtedly useful, again the reservations in the previous paragraph apply.

For the average MLwiN user, the instructions in the manual are voodoo. There are references to the statistical literature, of course, but these references are in many cases too technical to be of much use.

This, of course, is a well-known dilemma. If applied researchers, with often quite limited technical expertise, want to fit very complicated models with very complicated algorithms then the documentation and the implementation should handle this very carefully. Both in Goldstein (1995) and in the various manuals, one often gets the idea that instead of carefully guiding users through complicated territory, they are invited to climb on the roller coaster, close their eyes, and enjoy the ride.

13.2.2 HLM

HLM's history is similar to MLwiN's. There was first an HLM for two-level models, then one for three-level models, then one that could also do generalised linear regression, and finally a Windows version. This last version, version 4 of HLM/2L and HLM/3L, is the one we review here. Of course HLM/2L does two-level analysis and HLM/3L three-level analysis. The programming was done by Richard Congdon; the HLM team also includes Stephen Raudenbush and Tony Bryk.

Availability The software is available from Scientific Software International in Chicago, where HLM has its homepage

`www.ssicentral.com/hlm/mainhlm.htm`

The price is $400 for the DOS version and $430 for the Windows version.

Documentation The HLM documentation consists of a user's manual (Bryk *et al.*, 1996) and a book by Bryk and Raudenbush (1991). In a sense, the book is

independent of the software, but since you get the book when you buy the software, and since the HLM program is used throughout the book, the two are really intimately related (even closer than MLwiN and Goldstein (1995)). The choice of name HLM can lead to unfortunate confusion of the model, the technique and the package (de Leeuw and Kreft, 1995).

The Windows interface to HLM is only documented in the on-line help, it seems.

Model The basic HLM model, for two-level data, is

$$y_{ij} = \beta_{0j}x_{0ij} + \beta_{1j}x_{1ij} + \ldots + \beta_{pj}x_{pij} + e_{ij}, \tag{13.3}$$

where

$$\beta_{sj} = \gamma_{s0}z_{j0} + \gamma_{s1}z_{j1} + \ldots + \gamma_{sh}z_{jh} + u_{sj}. \tag{13.4}$$

To make the comparison with MLwiN somewhat easier, we rewrite this as

$$y_{ij} = \gamma_{00}x_{0ij}z_{j0} + \ldots + \gamma_{ph}x_{pij}z_{jh} + u_{0j}x_{0ij} + \ldots + u_{pj}x_{pij} + e_{ij}. \tag{13.5}$$

There are two obvious differences between the programs.

1. In MLwiN, we can have random coefficients β_{sij} on the first level. This is not possible in HLM.
2. In HLM, the emphasis is on the cross-level interactions in the fixed part, which are products of a first-level and a second-level predictor. It is possible to have such variables in MLwiN, but they are not as central.

Interface Similar to what we saw in MLn, the newer versions of HLM now have a Windows interface. In addition, they have the interactive (question-and-answer) and batchfile interfaces from the older versions. In Bryk *et al.* (1996), the authors remark that most PC users will prefer the Windows interface – but it is pretty clear from the rest of the book that they themselves do not.

The Windows version of HLM also comes with an equation editor, similar to the one in MLwiN.

Algorithm By default, HLM/2L uses REML estimation, while HLM/3L uses FIML. Nevertheless, both programs can actually do both forms of estimation. Older versions of HLM relied on the EM algorithm, which can sometimes be hopelessly slow. It is now possible to speed up convergence by switching to Fisher scoring.

Extensions There are a number of interesting extension in HLM. First, there is the V-known option (where the variance components are supposed to be known), which can be used in meta-analysis. Secondly, there are hierarchical generalised linear models, fitted by using penalised quasi-likelihood or generalised estimation equations. In particular, Poisson, Bernoulli and binomial models can be fitted. Thirdly, a form of plausible value analysis using multiple imputation is available for two-level models. Although the scope of HLM is

obviously much smaller than that of MLwiN, this restriction of generality makes the program easier to use. Generally, the documentation is more user-oriented, the number of choices the user can make is more limited, and often the authors of the program have already made many of the choices.

13.3 MINOR SPECIALISED PROGRAMS

13.3.1 VARCL

The program VARCL started out in the mid-1980s as one of the major contenders. It was Longford's research software, used in the path-breaking paper by Aitkin and Longford (1986) and in his book (Longford, 1993). Since 1990, Longford has moved to using S-plus for research software development, and nothing has happened with VARCL. It is still a DOS program.

Availability The program is sold by ProGAMMA in Groningen, The Netherlands, but it is in a very remote section of their catalogue:
`www.gamma.rug.nl`
The price is $375 ($250 educational).

Documentation The distribution comes with a 100–page user manual (Longford, 1990), with an additional 100 pages of example runs. The manual follows the interactive interface closely, but 20 pages are used for explaining some of the technical background.

Model There are two different versions of VARCL available. The first is VARCL3, which handles three-level models, the second is VARCL9, which handles random intercept models with up to nine levels. The model in VARCL3 is

$$y_{ijh} = \beta_{0jh}x_{0ijh} + \beta_{1jh}x_{1ijh} + \ldots + \beta_{pjh}x_{pijh} + e_{ijh}, \tag{13.6}$$

where

$$\beta_{sjh} = \beta_s + u_{sh} + u_{sjh}. \tag{13.7}$$

Again, we use the notation of MLwiN, and we assume that $x_{0ijh} = 1$ for all i, j, h.

In VARCL9, the model is (for four levels, as an example)

$$y_{ijhg} = \alpha_{jhg} + \beta_1 x_{1ijhg} + \ldots + \beta_p x_{pijhg} + e_{ijhg}, \tag{13.8}$$

where

$$\alpha_{jhg} = \alpha + u_g + u_{hg} + u_{jhg}. \tag{13.9}$$

Here we have singled out the intercept more explicitly, by using α for the corresponding regression coefficient.

We see that VARCL does not have the emphasis on cross-level interactions (or on two-level specification) that HLM has. It also does not have the facility for random coefficients at the first level, which does exist in MLwiN.

Interface VARCL has a command-line interface. The programs asks a large number of questions, and it uses the information provided by the user to build up a setup file that describes the analysis. It first builds up information about the maximal model, which is the largest model (in terms of the number of variables) that can be fitted in a session. Then additional questions are used to exclude variables from the maximal model to define the model to be fitted. Constraints on the parameters can be defined, by setting them to fixed values. The model is shown, and, if the user likes it, the program computes the estimates. Then the interface asks if additional models will be fitted, and if so, which variables are to be removed and added, and which parameters have to be constrained or freed.

If you know from experience which questions the program will ask, and which answers you are going to give, then you can put the answers in a batch file and take standard input from that file.

Algorithm The program uses the scoring method on the reduced form. It is based on the fact that within-group means, variances and covariances are jointly sufficient for the model parameters. Thus, if we compute the within-group statistics for the maximal model, we can forget about the original data. Since these within-group statistics can be computed in the input loop, the number of observations that VARCL can handle is infinitely large, as long as they come in a finite number of groups. This is different from MLwiN, which keeps the data in core for the whole session. It is similar to HLM, which uses the within-group statistics to compute separate regressions for each group (if the groups are big enough). Both VARCL and HLM use within-group statistics to compute initial estimates of the variance and covariance components.

In our experience, documented in our previous review papers, the scoring algorithm is both fast and reliable.

Extensions VARCL was also the first multilevel program that could deal with hierarchical generalised linear models. Early on in the interactive questions and answers session, the user can choose the error model to be normal, binomial, Poisson or gamma. Again quasi-likelihood methods are used to fit these models, using the procedures designed by Longford (1988) for this purpose.

13.3.2 MIXFOO

The MIXFOO program is special – because it does not exist. It is the generic name for a whole series of multilevel programs. At the MIXFOO website, we find the following: 'The statistical research presented in this homepage is based on the collaborative effort of Donald Hedeker and Robert D. Gibbons of the University of Illinois at Chicago. The computer programs were written by Don Hedeker with interfaces written by Dave Patterson (Discerning Systems, Inc.).'

The names of all programs in the series start with MIX. There are DOS programs MIXOR, MIXREG, MIXGSUR, MIXNO and MIXPREG, while MIXOR and MIXREG also exist with Windows interfaces. The website also

has a SAS-IML macro that fit a random intercept version of MIXREG, and SPSS-MATRIX macros for random intercept versions of both MIXOR and MIXREG. All this is freely available from the website.

We give here a brief indication of what these programs do. MIXREG fits the linear multilevel model, but it allows for various forms of autocorrelation between the first-level disturbances. MIXOR adds ordinal multicategory outcomes. MIXGSUR fits grouped-time survival data, MIXNO fits nominal multicategory data, and MIXPREG handles multilevel Poisson regression.

Availability Software and manuals can be downloaded from
`www.uic.edu/~hedeker/mix.html`
All programs, macros and manuals are free. There are PowerMac and Sun/Solaris version of the software at
`www.stat.ucla.edu/~deleeuw/mixfoo`
It must be emphasised that this is quite unique – all other programs (except VARCL and MLA) only exist for DOS or Windows.

Documentation The two core programs, MIXOR and MIXREG, are described in Hedeker and Gibbons (1996a) and Hedeker and Gibbons (1996b). These are published versions of the manuals; the manuals themselves are included in the software distribution. The theory is described in great detail in Hedeker (1989) and for MIXOR in the article by Hedeker and Gibbons (1994). MIXGSUR, MIXNO and MIXPREG also have a manuals on the website. The theory of MIXGSUR is described in the technical report by by Hedeker *et al.* (1996). More generally, there is a list of both theoretical and applied articles using MIXFOO at
`www.uic.edu/~hedeker/works.html`

Model For MIXREG, the model is

$$y_{ij} = \alpha_0 w_{0ij} + \ldots + \alpha_p w_{pij} + \beta_{0j} x_{0ij} + \ldots + \beta_{qj} x_{qij} + e_{ij}, \tag{13.10}$$

with

$$\beta_{sj} = \beta_s + u_{sj}. \tag{13.11}$$

This is a straightforward two-level random coefficient model, very similar to what we have in VARCL. The unique aspect is that we do not assume that

$$\text{cov}(e_{ij}, e_{kj}) = \delta^{ik}\sigma^2, \tag{13.12}$$

but we allow for a much more general parametric first-level error covariance structure. More specifically, e_{ij} the can have an autoregressive AR(1), moving-average MA(1), or autoregressive moving-average ARMA(1,1) covariance structure, a general stationary autocorrelation structure, and even a special non-stationary one.

For MIXOR, the model is the same, except that we do not observe the y_{ij}; rather, we observe a multicategory version generated by a threshold model. Thus there are unknown cut-off points (assumed to be the same for all

variables). If y_{ij} is below the first cut-off, we observe a '1'; if it is between the first and second cut-offs, we observe a '2'; and so on. Thus the model is basically the same as before, but there are 'missing data' because we only know that y_{ij} is in a particular interval, but we don't know where it is in that interval. More precisely, the model is

$$\eta_{ij} = \alpha_0 w_{0ij} + \ldots + \alpha_p w_{pij} + \beta_{0j} x_{0ij} + \ldots + \beta_{qj} x_{qij} + e_{ij}, \tag{13.13}$$

where η_{ij} is the unobserved continuous response, and

$$y_{ij} = \begin{cases} 1 & \text{if } k_0 \leq \eta_{ij} < k_1, \\ 2 & \text{if } k_1 \leq \eta_{ij} < k_2, \\ r & \text{if } k_{r-1} \leq \eta_{ij} < k_r, \end{cases} \tag{13.14}$$

where $k_0 = -\infty$ and $k_r = +\infty$.

Interface All programs exist as batch versions, using command files. MIXOR and MIXREG also have interactive DOS versions, in which the command file is constructed from menu commands issued by the user. Finally, MIXREG and MIXOR are available as Windows programs.

Algorithm MIXREG uses a combination of the EM and the scoring algorithm, in much the same way as, for instance, HLM. For MIXOR there are additional complications, because multidimensional integrals must be evaluated to compute the likelihood and its derivatives. MIXOR approximates these integrals by using Gauss–Hermite quadrature.

Extensions If we consider MIXREG and MIXOR to be the basic components then MIXGSUR, MIXNO and MIXPREG are extensions. But of course drawing the line in this way is rather arbitrary. It is better to think of the whole set of programs as a modular alternative to programs such as HLM and MLwiN, which cover about the same amount of territory in a single program.

13.3.3 MLA

MLA is a batch program, running under DOS. It was written by Frank Busing, Rien van der Leeden, and Eric Meijer of the University of Leiden, The Netherlands. It differs from similar programs because it has extensive simulation possibilities built in (notably the bootstrap and the jackknife), and it has various ordinary least squares estimation methods as options.

Availability Software and manual can be downloaded from
`www.fsw.leidenuniv.nl/www/w3_ment/medewerkers/BUSING/MLA.HTM`
The program is free.

Documentation The software distribution contains a 70–page Postscript manual (Busing *et al.*, 1994). It is a bit wordy, because it tries to spell out all details, especially the technical ones. The ultimate example is the 20–page

Technical Appendix A, which gives in painstaking detail the derivations of the formulae for the likelihood function, its first and second derivatives, and their expectations. It illustrates the effect TEX has on the mind of an individual who has just escaped from the dungeons of WYSIWYG. The user's guide portion of the manual is quite clear, however.

The authors have informed us that the manual is out of date, because many options have been added to MLA in the meantime. The program can now make histograms (of bootstrap results, for instance) and scatterplots (of residuals), and there are different bootstrap-based confidence intervals. Also, permutation tests for testing intraclass correlation are available.

Model The model is the same as in (13.3) and (13.4).

Interface MLA requires the user to create a parameter file, and then the program is started from the DOS command line as

$$\text{mla} < \text{inputfile} > \text{outputfile}$$

Not much of an interface – but it does the job.

Algorithm MLA uses the BFGS algorithm, which is a general-purpose quasi-Newton optimisation algorithm, to maximise either FIML or REML. Alternatively, it can also use EM. In order to make sure that the level-2 dispersion matrix is positive-semidefinite, two different parameterisations are available that ensure this.

Extensions MLA is somewhat unremarkable as a multilevel program (although the parameterisations of the dispersion matrix and the use of BFGS are unique). However, it is remarkable because it supports a wide variety of simulation analyses. This tends to suggest that it is research software, but for practitioners the confidence information will also be useful.

Here is a list of the unique features taken from the website:

- different kinds of simulations (bootstrap, jackknife and permutation),
- different methods of bootstrap simulation (cases, parametric and error),
- different types of residual estimation (raw and shrunken),
- different cases resampling schemes (level 1, level 2 and both),
- balanced resampling schemes (balanced bootstrap),
- linking of residual levels,
- distribution plots (histograms) for parameters, standard errors and t-value.

13.4 MODULES IN A MAJOR PACKAGE

13.4.1 BMDP5-V

BMDP was the first major statistical package. It has always had an excellent reputation, because its approximately 40 programs were developed in close

cooperation with excellent statisticians. But BMDP has a stormy recent past. The company was bought by SPSS in 1996. It is difficult to find BMDP on the main SPSS website at `http://www.spss.com`. It turns out that many of the BMDP products have been declared 'dead' by SPSS, and the only product still sold seems to be BMDP Classic for DOS. See
`www.spss.com/software/science/BMDP`

Some additional research suggests that SPSS plans to integrate BMDP in its SYSTAT product. In corporate-speak: 'I would like to welcome BMDP customers to the SYSTAT family.' In Europe, the situation is different. One can buy BMDP from Statistical Solutions in Cork, Ireland, and their web page clearly indicates they consider it to be one of their major products.

Availability BMDP can be ordered from
`www.statsol.ie/bmdp.html`
The price is £1295 (£1100 academic; $895 and $695 in the Americas), but of course this is for the whole package (all 40+ routines, of which BMDP5-V is only a single one).

Documentation The theory behind BMDP5-V is explained in Jennrich and Schluchter (1986). Very useful information about both the methodology and the program is in Schluchter (1988). There is also a chapter on 5V in the BMDP user's manual (Dixon, 1992) and the BMDP user's digest (BMDP Statistical Software, 1992). And, last but not least, there is a detailed 'how-to-use' book with many examples (Dixon and Merdian, 1992) (with a 5.25 floppy in the back!).

Model BMDP5-V is described in the documentation as a program for unbalanced repeated measures models with structured covariance matrices. This means that the emphasis is not on general hierarchies, but specifically on repeated measures, i.e. on repetitions nested within subjects.

To maintain the same notation as we have used before, we rewrite the model as

$$y_{ij} = x_{1ij}\beta_1 + \ldots + x_{pij}\beta_p + e_{ij}, \tag{13.15}$$

where the index j now stands for subjects, and i stands for the repeated measures (nested within subjects). Thus it seems there are no random coefficients. BMDP5-V allows a large number of possible choices for the within-subject covariance matrices $\mathbf{\Sigma}_j$ which has elements

$$\sigma_{ii'j} = E(e_{ij}, e_{i'j}). \tag{13.16}$$

We briefly describe the available options, in matrix notation:

$$\mathbf{\Sigma}_j = \sigma^2\mathbf{I}, \tag{13.17}$$

$$\mathbf{\Sigma}_j = \omega^2\mathbf{e}\mathbf{e}^{\mathrm{T}} + \sigma^2\mathbf{I}, \tag{13.18}$$

$$\mathbf{\Sigma}_j = \mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}} + \sigma^2\mathbf{I}, \tag{13.19}$$

$$\boldsymbol{\Sigma}_j = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Psi}, \tag{13.20}$$

$$\boldsymbol{\Sigma}_j = \theta_1 \mathbf{G}_1 + \ldots + \theta_r \mathbf{G}_r, \tag{13.21}$$

$$\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}. \tag{13.22}$$

Moreover, there are two options that can best be defined in elementwise notation:

$$\sigma_{ii'j} = \sigma^2 \rho^{|i-i'|}, \tag{13.23}$$

$$\sigma_{ii'j} = \theta_r, \quad \text{where} \quad r = |i - i'| + 1. \tag{13.24}$$

Clearly, this covers many of our well-known friends from earlier sections.

Model (13.17) defines independent observations (within subjects; there is always between subject independence), while model (13.18) (sometimes described as compound symmetry) is in fact the random intercept model. Model (13.19) is the random-coefficient model, which allows for random slopes as well. Thus random coefficients, which seemed to be missing in the BMDP5-V specification, enter again through the back door. Model (13.20) says that within-subject covariance matrices have a factor analysis structure, while (13.21) describes a general linear covariance structure, and (13.22) allows for any matrix (except that it must be the same over subjects). The final two models allow for autoregressive structures, similar to those we have seen in MIXREG. BMDP5-V also allows you to define your very own structure, provided you write a FORTRAN program to compute the structure and its derivatives. Really enterprising users could fit a repeated measures model in which the within-subject covariance structures satisfy a LISREL model, for instance.

Interface BMDP is a very modular package. Each of the 44 programs can basically be used in a stand-alone way. This makes it quite different from SAS, which is much more integrated. Modularity produces a great deal of efficiency. On the other hand, BMDP is very, very DOS. It comes with BMDP/DYNAMIC, a somewhat optimistically named data and program managing module. BMDP/DYNAMIC has the familiar DOS-type pseudo-menus and pseudo-windows. It is fairly lightweight in terms of overhead.

Algorithm A variety of algorithms are available. If you choose the FIML criterion, you can use Fisher scoring, Newton–Raphson or generalised EM. In the case of REML, generalized EM and quasi-scoring are available.

Extensions Some of the options for the within-subject covariance matrix are rather esoteric, and could be considered extensions. Certainly options for which you have to write the code yourself qualify as such.

13.4.2 PROC MIXED

SAS has many modules (or PROCs) to fit linear models and generalised linear models. Until fairly recently, however, there was nothing that could compete

with BMDP5-V. But now there is – and with a vengeance. PROC MIXED is the most general program of those we discuss, and it is also embedded in the totality of SAS (or whatever parts of SAS you happen to have installed).

Availability You can obviously order the package from SAS Institute. Start at `www.sas.com/service/techsup/faq/stat_proc/mixeproc.html`
Because SAS is such a huge and complicated product, it does not make much sense to give a price. Many users will simply add PROC MIXED to their PC or mainframe SAS setup. Many people probably already have it available without actually knowing this.

Documentation The documentation for PROC MIXED is very extensive. It consists of a chapter in SAS (1992). This is written in the familiar format of the SAS manuals. It has almost 100 pages, set up like a user's guide. There is not much theory, and not many examples, but the instructions and options are explained in detail.

There is also an excellent and extensive book by Littell *et al.* (1996). It has chapters on randomised blocks, repeated measures, multilevel models, analysis of covariance, spatial models, heteroscedastic models and best linear unbiased prediction. There are also chapters on generalised linear mixed models and nonlinear mixed models, where the SAS-IML macros GLIMMIX and NLINMIX are explained. The book contains many examples, worked out in detail.

For people who have started out with a program such as HLM, which has a much simpler interface and a much smaller menu of models, the use of PROC MIXED is explained expertly by Singer (1998).

In terms of documentation, PROC MIXED is the clear winner in our comparison. In most other respects, it is not.

Model The model in PROC MIXED is very similar to the model in BMDP5-V. It may even be true, historically, that BMDP5-V inspired PROC MIXED. The main difference, however, is that the PROC MIXED specification does not have levels or repeated measures. They start from the general mixed model, which is

$$y_i = \beta_1 x_{1i} + \ldots + \beta_r x_{ri} + u_1 z_{1i} + \ldots + u_s z_{si} + e_i. \qquad (13.25)$$

The random effects u have a covariance matrix $\boldsymbol{\Sigma}$ and the errors e have a covariance matrix $\boldsymbol{\Delta}$. As in BMDP5-V, we can specify additional parametric structure in these covariance matrices. PROC MIXED allows for uncorrelated, compound symmetry, unstructured, autoregressive, and spatial (or a block-diagonal version of these – which is how the levels come in again). There is no factor-analytic or general linear covariance structure.

It is interesting that HLM and MLA fit true multilevel models, in the sense that they specify their regression model at multiple levels. MIXFOO and MLwiN have multiple levels, but only a single regression equation. The same is true for BMDP5-V, where the levels are defined by repeated measures. In PROC MIXED the levels have disappeared, and they have to be introduced by

suitably arranging the input and parameter files. This is precisely the reason why a paper such as Singer (1998) is necessary for some groups of users.

Interface The interface to PROC MIXED is the familiar SAS interface – or more precisely any one of the familiar SAS interfaces. As in the case of BMDP, all these interfaces are rather desperate attempts to get away from the mainframe or DOS heritage. Most of the Windows interface is used to construct batch files with options and instructions, which are then submitted to the old familiar SAS engine in the background.

Algorithm PROC MIXED can use both REML and FIML estimation, and it maximises the likelihood by a combination of Fisher scoring and Newton–Raphson. The default is to use only Newton–Raphson, but optionally one can start with some Fisher scoring iterations.

Extensions The two major extensions of PROC MIXED are both written in the macro language SAS-IML, by Russ Wolfinger. They are discussed in detail in Littell *et al.* (1996, Chapters 11 and 12). Generalised linear mixed models can be fitted with the GLIMMIX macro. This incorporates binomial models with logit and probit links and Poisson (count) models with the log link. The macros uses a joint quasi-likelihood method. Nonlinear mixed models can be fitted with the NLINMIX macro. This macro actually has three options, depending on whether first-order, second-order or marginal approximations are used. This allows one to fit a really large range of models such as nonlinear growth curve models.

It is probably obvious that using such macros cannot possibly be very efficient. Using SAS to fit your multilevel models is already quite a stretch, because you have to carry along the enormous overhead of the SAS system. A class of students, each using SAS on a single CPU, will show you what this means. On top of that, the use of macros and the SAS-IML interpreter adds another layer of inefficiency. It's a little bit like using a big truck to pick up some groceries at your local supermarket.

There is a trade-off, of course. Many people have SAS installed and are quite familiar with its interface and macro language. If they need to fit a multilevel model once in a while, it probably does not make sense for them to learn a new language or to buy a new software package. If you are already driving around in a truck, you might as well pick up your groceries.

13.5 CONCLUSIONS

We have already commented throughout the text on some of the differences. There are multilevel, random coefficient and mixed linear model programs. There are package modules and stand-alone programs. There are research and production programs. One obvious characteristic they have in common is that they are almost all DOS, trying to become Windows.

We do not really want to get too deeply into the discussion of the relative advantages of putting a GUI on top of a DOS program. Clearly, there are some definite advantages. Graphics output becomes much more attractive and more flexible, and the equations window of MLwiN is a useful addition, which would be impossible under DOS. There are also some disadvantages, which are the same as for other programs with the same history such as SAS, SPSS and BMDP. We are forced to leave the safe domain of objective and impartial software comparison here, in order to voice an opinion. The GUI of all these programs is ugly as sin. This is partly due to the inherent ugliness of the Windows environment, but also to the fact that the interface is an afterthought. Its only job is to prepare parameter and data files, which are then passed to the DOS programs running in the background. One does not need a GUI to prepare parameter files – in fact, this is not the natural way to prepare them at all. If you are a command line program at heart, then you should acknowledge that, and you should not pretend that a layer of Windows make-up can change this.

A command line interface that builds up the parameter file by asking questions, that is the interface for VARCL and that was also the interface for older versions of HLM has some advantages over the Windows interfaces. It is obviously more structured, and the user does not have to search around in the menus for the appropriate commands. Strange and impossible combinations of options are more easily avoided. It is more foolproof. On the other hand, it rapidly becomes very boring to go through the same sequence of questions every time, and more experienced users will prefer to type the answers ahead of time into a command file and use the program in batch mode (which makes the 'interface' similar to the batch versions of MIXFOO or MLA).

If we summarise our findings, there is little doubt that MLwiN is the most comprehensive program for multilevel analysis. Of course, it is not exactly fair to compare MLwiN with the whole SAS or even the BMDP system. Those are more comprehensive, and they can also be used to perform multilevel analysis. HLM obviously has also developed into a mature and excellent product. Its basic design is more modest than MLwiN, it lacks the macro language, the graphics, and the NANOSTAT statistics and data handling. MLA is somewhat limited, but it can do many useful things, some of which are impossible or very hard to do in the other packages.

Probably for most academic users, the price of the packages is not really a problem. The fact that MLA is free and that MLwiN costs money does not seem to be a major problem. We must say, however, that the suit of MIXFOO programs by Hedeker is really a bargain. One can do ordinary multilevel analysis, multilevel survival analysis, categorical and ordinal outcomes, and Poisson multilevel analysis – and all this in a uniform and modular way. The programs are fast, have little overhead, are versatile, well documented and completely free. Moreover, binaries are available for Power-Macs and for Sun Solaris machines. For most of the MIXFOO programs, one needs to write batch setup files, but, as we have indicated above, that is not really a major disadvantage. Market forces have conspired to make us believe that it is,

and that graphical interfaces are the be-all and end-all of software, but most quality statistical and numerical software does not have and does not need such interfaces. Editing an existing text file is usually faster than wading through not very intuitive menus.

References

Aitkin MA, Longford NT (1986) Statistical modeling issues in school effectiveness studies (with discussion). *Journal of the Royal Statistical Society A* **149**, 1–43.

Albert J, Chib S (1997) Bayesian tests and model diagnostics in conditionally independent hierarchical models. *Journal of the American Statistical Association* **92**, 916–25.

Amemiya T (1981) Qualitative response models: A survey. *Journal of Economic Literature* **19**, 483–536.

Belsley DA, Kuh E, Welsch RE (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, New York.

Bentham G, Langford IH (1996) Climate change and the incidence of food poisoning in England and Wales. *International Journal of Biometeorology* **39**, 81–86.

Bentham G, Langford IH and Day RJ (2000) Weekly incidence of food poisoning and ambient temperatures in the hot summers of 1989 and 1990. *International Journal of Biometeorology* (submitted).

Bernardinelli L, Montomoli C (1992) Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statistics in Medicine* **11**, 983–1007.

Bernardo JM, Smith AFM (1994). B*ayesian Theory*. Wiley, Chichester.

Besag J, York J, Mollié A (1991) Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**, 1–59.

Best NG, Spiegelhalter D, Thomas A, Brayne CEG (1996) Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society A* **159**, 323–42.

BMDP Statistical Software (1992) BMDP *User's Digest*. University of California Press.

Breslow NE, Clayton DG (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

Bryk AS, Raudenbush SW (1992) *Hierachical Linear Models*. Sage, Newbury Park, CA.

Bryk AS, Raudenbush S, Congdon R (1996) *Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs*. Scientific Software International, Chicago.

Bull JM, Riley GD, Rasbash J, Goldstein H (1999) Parallel implementation of a multilevel modelling package. *Computational Statistics and Data Analysis* **31**, 457–74.

Busing FMTA, Meier E, van der Leeden R (1994) MLA software for multilevel analysis of data with two levels. Technical Report, Department of Psychometrics and Research Methodology, University of Leiden.

Cavalli-Sforza LL, Menozzi P, Piazza, A (1994). *The History and Geography of Human Genes*. Princeton University Press.

Chatterjee S, Hadi AS (1988) *Sensitivity Analysis in Linear Regression*. Wiley, New York.

Clayton D (1988) The analysis of event history data: a review of progress and outstanding problems. *Statistics in Medicine* **7**, 819–41.

Clayton, D, Hills, M (1993). *Statistical Models in Epidemiology*. Oxford University Press.

Clayton, D, Kaldor J (1987) Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671–81.

Clayton D, Rasbash J (1999) Estimation in large crossed random effects models by data augmentation. *Journal of the Royal Statistical Society* A **162**, 425–36.

Cochran WG (1977) *Sampling Techniques*, 3rd edn. Wiley, New York.
Cohen J (1992) A power primer. *Psychological Bulletin* **112**, 155–9.
Cohen M (1998) Determining sample sizes for surveys with data analyzed by hierarchical linear models. *Journal of Official Statistics* **14**, 267–275.
Collett D (1991) *Modelling Binary Data*. Chapman & Hall, London.
Cook RD, Weisberg S (1982, 1995) *Residuals and Influence in Regression*. Chapman & Hall, London.
Copas J (1999) What works? Selectivity models and meta-analysis. *Journal of the Royal Statistical Society A* **162**, *95–109*.
Cox DR, Oakes D (1984) *Analysis of Survival Data*. Chapman & Hall, London.
Cox DR, Snell EJ (1989) *Analysis of Binary Data, 2nd edn. Chapman & Hall, London.*
Curtis SL, Diamond I, McDonald JW (1993) Birth interval and family effects on postneonatal mortality in Brazil. *Demography* **30**, 33–43.
Davis P, Gribben B (1995). Rational prescribing and interpractitioner variation: a multilevel approach. *International Journal of Technology Assessment in Health Care* **11**, 428–442.
de Courcy-Wheeler RHB, Wolfe CDA, Spencer M, Goodman JDS, Gamsu HR (1995). Use of the CRIB (clinical risk index for babies) score in prediction of neonatal mortality and morbidity. *Archives of Disease in Childhood* **73**, *32–6*.
de Leeuw J, Kreft IGG (1995) Questioning multilevel models. *Journal of Educational and Behavioral Statistics* **20**, 171–90.
DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–88.
Diggle PJ (1988) An approach to the analysis of repeated measurements. *Biometrics* **44**, 959–971.
Dixon WJ (1992) BMDP *Statistical Software Manual*. University of California Press.
Dixon WJ, Merdian K (1992) *ANOVA and Regression with* BMDP5–V. Dixon Statistical Associates, Los Angeles.
Dobson AJ (1991) *An Introduction to Generalized Linear Models*. Chapman & Hall, London.
Donner A (1986) A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *International Statistical Review* **54**, 67–82.
DuBois RW, Brook RH, Rogers WH (1987) Adjusted hospital death rates: A potential screen for quality of medical care. *American Journal of Public Health* **77**, 1162–7.
Duncan C (1997) Applying mixed multivariate multilevel models in geographical research. In: Westert GP, Verhoeff RN (eds) *Places and People: Multilevel modelling in Geographical Research*. The Royal Dutch Geographical Society, Utrecht.
Duncan C, Jones K, Moon G (1998) Context, composition and heterogeneity: using multilevel models in health research. *Social Science and Medicine* **46**, 97–117.
Echoard R, Clayton D (1998) Multilevel modelling of conception in artificial insemination by donor. *Statistics in Medicine* **17**, 1137–56.
Elliott P, Martuzzi M, Shaddick G (1995) Spatial statistical methods in environmental epidemiology: a critique. *Statistical Methods in Medical Research* **4**, 137–59.
English D (1992) Geographical epidemiology and ecological studies. In: Elliott P, Cuzick J, English D, Stern R (eds) *Geographical and Environmental Epidemiology: Methods for Small-Area Studies. Oxford University Press, 3–13.*
Epstein A (1995) Performance reports on quality-prototypes, problems and prospects. *New England Journal of Medicine* **333**, 57–61.
Flay BR, Brannon BR, Johnson CA, Hansen WB, Ulene AL, Whitney-Saltiel DA, Gleason LR, Sussman S, Gavin M, Glowacz KM, Sobol DF Spiegel DC (1989) The television, school and family smoking cessation and prevention project: I. theoretical basis and program development. *Preventative Medicine* **17**, 585–607.
Flay BR, Miller TQ, Hedeker D, Siddiqui O, Britton CF, Brannon BR, Johnson CA, Hansen WB, Sussman S, Dent C (1995) The television school and family smoking

prevention and cessation project: VIII. Student outcomes and mediating variables. *Preventive Medicine*, **24**, 29–40.
Fuller P, Picciotto A, Davies M, McKenzie SA (1998) Cough and sleep in inner city children. *European Respiratory Journal* **12**, 426–431.
Gatsonis C, Normand S-L, Liu C, Morris C (1993) Geographical variation of procedure utilization: a hierarchical model approach. *Medical Care (Suppl)* **31**, YS55–9.
Gelfand AE and Smith AFM (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Society* **85**, 398–409.
Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **6**, 721–41.
Gilks WR, Richardson S, Spiegelhalter DJ (1996), *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.
Goldstein H (1979) *The Design and Analysis of Longitudinal Studies*. Academic Press, London.
Goldstein H (1986) Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* **73**, 43–56.
Goldstein H (1987) *Multilevel Models in Educational and Social Research*. Griffin, London.
Goldstein H (1989a) Efficient prediction models for adult height. In: Tanner JM (ed.) *Auxology 88; Perspectives in the Science of Growth and Development*. Smith Gordon, London.
Goldstein H (1989b) Restricted (unbiased) iterative generalized 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 (1995) *Multilevel Statistical Models, 2nd edn. Edward Arnold, London Wiley, New York.*
Goldstein H, Rasbash J (1992) Efficient computational procedures for the estimation of parameters in multilevel models based on iterative generalized least squares. *Computational Statistics and Data Analysis* **13**, 63–71.
Goldstein H, Rasbash J (1996) Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society A* **159**, 505–13.
Goldstein H, Spiegelhalter DJ (1996) League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society A* **159**, 385–443.
Goldstein H, Rasbash J, Yang M, Woodhouse G, Pan H, Nuttall D (1993) A multilevel analysis of school examination results. *Oxford Review of Education* **19**, 425–33.
Goldstein H, Healy MJR, Rasbash, J (1994) Multilevel time series models with applications to repeated measures data. *Statistics in Medicine* **13**, 1643–55.
Goldstein H, Rasbash J, Plewis I, Draper D, Browne W, Yang M, Woodhouse G, Healy M (1998). *A User's Guide to MLwiN*. University of London.
Goldstein H, Yang M, Omar R, Turner R, Thompson SG (1999) Meta-analysis using multilevel models with an application to the study of class size effects. *Applied Statistics* **49**, 399–412.
Grizzle JC, Allen DM (1969) An analysis of growth and dose response curves. *Biometrics* **25**, 357–61.
Groenewegen P (1997) Dealing with micro-macro relations: a heuristic approach with examples from health services research. In: Westert GP, Verhoeff RN (eds) *Places and People: Multilevel Modelling in Geographical Research*. The Royal Dutch Geographical Society, Utrecht, 9–18.
Guilkey DK, Murphy JL (1993) Estimation and testing in the random effects probit model. *Journal of Econometrics* **59**, 301–17.
Gustafsson JE, Stahl PE (1997) *STREAMS User's Guide*. Multivariate Ware, Mölndal, Sweden.

Harrison GA, Brush G (1990) On correlations between adjacent velocities and accelerations in longitudinal growth data. *Annals of Human Biology* **17**, 55–7.

Hedeker D (1989) Random regression models with autocorrelated errors. PhD Thesis, University of Chicago.

Hedeker D, Gibbons RD (1994) A random effects ordinal regression model for multilevel analysis. *Biometrics* **50**, 933–44.

Hedeker D, Gibbons RD (1996a) MIXOR. A computer program for mixed-effects ordinal probit and logistic regression analysis. *Computer Methods and Programs in Biomedicine* **49**, 157–76.

Hedeker D, Gibbons RD (1996b) MIXREG. A computer program for mixed-effects regression analysis with autocorrelated errors. *Computer Methods and Programs in Biomedicine* **49**, 229–52.

Hedeker D, Gibbons RD, Flay BR (1994) Random-effects regression models for clustered data: with an example from smoking prevention research. *Journal of Consulting and Clinical Psychology* **62**, 757–65.

Hedeker D, Siddiqui O, Hu FB (1996) Random-effects regression analysis of correlated grouped-time survival data. Technical Report, University of Illinois, Chicago.

Hedges L, Vevea JL (1996) Estimating effect size under publication bias: small sample properties and robustness of a random effects selection model. *Journal of Educational and Behavioural Statistics* **21**, 299–332.

Hill PW, Goldstein H (1998) Multilevel modelling of educational data with cross-classification and missing identification of units. *Journal of Educational and Behavioral Statistics* **23**, 117–28.

Hocking RR (1996) *Methods and Applications of Linear Models: Regression and the Analysis of Variance*. Wiley, New York.

Hox JJ (1995) *Applied Multilevel Analysis*. TT-Publikaties, Amsterdam.

Jencks SF, Daley J, Draper D, Thomas N, Lenhart G, Walker J (1988) Interpreting hospital mortality data. The role of clinical risk adjustment. *Journal of the American Medical Association* **260**, 3611–16.

Jennrich R, Schluchter M (1986) Unbalanced repeated measures models with structured covariance matrices. *Biometrics* **42**, 805–20.

Jones K (1997) Multilevel approaches to modelling contextuality: from nuisance to substance in the analysis of voting behaviour. In: Westert GP, Verhoeff RN (eds) *Places and People: Multilevel Modelling in Geographical Research*. The Royal Dutch Geographical Society, Utrecht 19–43.

Jones K, Moon G, Clegg, A (1991) Ecological and individual effects in childhood immunisation uptake: a multi-level approach. *Social Science and Medicine* **33**, 501–8.

Kenward MG (1998) Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. *Statistics in Medicine* **17**, 2723–32.

Knaus WA, Draper EA, Wagner DP, Zimmerman JE (1985) APACHE II: A Severity of Disease Classification System. *Critical Care Medicine* **13**, 818–29.

Kraemer HC, Thiemann S (1987) *How many subjects? Statistical power analysis in research*. Sage, London.

Kreft IGG, de Leeuw J (1998) Introducing Multilevel Modelling. Sage, London.

Kreft IGG, de Leeuw J, Kim K-S (1990) Comparing four different statistical packages for hierarchical linear regression: GENMOD, HLM, ML2, VARCL. UCLA Statistics Preprint 50.

Kreft IGG, de Leeuw J, van der Leeden R (1994) Review of five multilevel analysis programs: BMDP-5V, GENMOD, HLM, ML3, VARCL. *American Statistician* **48**, 324–35.

Kunin CM, Johansen KS, Worning AM, Daschner FD (1990) Report of a symposium on use and abuse of antibiotics worldwide. *Reviews of Infectious Diseases* **12**, 12–19.

Lange N, Ryan L (1989) Assessing normality in random effects models. *Annals of Statistics* **17**, 624–42.

Langford IH (1994) Using empirical Bayes estimates in the geographical analysis of disease risk. *Area* **26.2**, 142–9.

Langford IH, Bentham G (1997) A multi-level model of sudden infant death syndrome in England and Wales. *Environment and Planning A* **29**, 629–40.

Langford I, Lewis, T (1998) Outliers in multilevel data. *Journal of the Royal Statistical Society A* **161**, 121–60.

Langford IH, Bentham G, McDonald A-L (1998) Multilevel modelling of geographically aggregated health data: a case study on malignant melanoma mortality and UV exposure in the European Community: a multi-level modelling approach. *Statistics in Medicine* **17**, 41–58.

Langford IH, Leyland AH, Rasbash J, Goldstein H (1999a) Multilevel modelling of the geographical distributions of rare diseases. *Applied Statistics* **48**, 253–68.

Langford IH, Leyland AH, Rasbash J, Goldstein H, Day RJ, McDonald A-L (1999b) Multilevel modelling of area-based health data. In: Lawson A, Biggeri A, Böhning D, Lesaffre E, Viel J-F, Bertolini R (eds) *Disease Mapping and Risk Assessment for Public Health*. Wiley, Chichester, 217–28.

Lawson AB, Biggeri A, Williams FLR. (1999a) A review of modelling approaches in health risk assessment around putative sources. In: Lawson A, Biggeri A, Böhning D, Lesaffre E, Viel J-F, Bertolini R (eds) *Disease Mapping and Risk Assessment for Public Health*. Wiley, Chichester, 231–45.

Lawson AB, Böhning D, Biggeri A, Lesaffre E, Viel J-F. (1999b) Disease mapping and its uses. In: Lawson A, Biggeri A, Böhning D, Lesaffre E, Viel J-F, Bertolini R (eds) *Disease Mapping and Risk Assessment for Public Health*. Wiley, Chichester, 3–13.

Leyland A H, Boddy F A (1998) League tables and acute myocardial infarction. *Lancet* **351**, 555–8.

Leyland AH, Langford IH, Rasbash J, Goldstein H (1998) Spatio-temporal modelling of mortality data for multiple causes of death. In: *Proceedings of the Biometrics Section of the American Statistical Association*, 10–15.

Leyland AH, Langford IH, Rasbash J, Goldstein H (2000) Multivariate spatial models for event data. *Statistics in Medicine* **19**, 2469–2478.

Li XS (1997) multilevel statistical modelling for discrete responses in epidemiological field study. PhD Dissertation, School of Public Health, West China University of Medical Sciences.

Lindsey JK (1999) Relationships among sample size, model selection and likelihood regions, and scientifically important differences. *Journal of the Royal Statistical Society D* **48**, 401–12.

Littell RC, Milliken GA, Stroup WW, Wolfinger RD (1996) *SAS System for Mixed Models*. SAS Institute, Cary, NC.

Little RJA (1995). Modeling the dropout mechanism in repeated measures studies. *Journal of the American Statistical Association* **90**, 1112–21.

Longford NT (1988) A quasi-likelihood adaptation for variance component analysis. In: *Proceedings of the Section on Statistical Computing of the ASA*, 137–42.

Longford NT (1990) VARCL Software for variance component analysis of data with nested random effects (maximum likelihood). Technical Report, Educational Testing Service, Princeton.

Longford NT (1993) *Random Coefficient Models*. Oxford University Press.

McCullagh P, Nelder JA (1994). *Generalized Linear Models*. Chapman & Hall, London.

McCulloch C, Leyland AH, Boddy FA (1997) Understanding the relationship between length of stay and readmission rates for hospitals (progress report for the King's Fund Comparative Database Initiative). University of Glasgow, Public Health Research Unit.

McDonald RP (1994) The bilevel reticular action model for path analysis with latent variables. *Sociological Methods and Research* **22**, 399–413.

McKee M, Hunter D (1995). Mortality league tables: Do they inform or mislead? *Quality in Health Care* **4**, 5–12.

McLeod A, Ross P, Mitchell S, Tay D, Hunter L, Hall A, Paton J, Mutch L (1996) Respiratory health in a total very low birthweight cohort and their classroom controls. *Archives of Disease in Childhood* **74**, 188–94.

Maddala GS (1996). *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge University Press.

Marshall EC Spiegelhalter DJ (1998) League tables of in vitro fertilisation clinics: How confident can we be about the rankings? *British Medical Journal* **317**, 1701–04

Modigliani F, Ando A (1963) The "life-cycle" hypothesis of saving: aggregate implications and tests. *American Economic Review* **53**, 55–84.

Moerbeek M, van Breukelen GJP, Berger MPF (2000a) Design issues for experiments in multilevel populations. *Journal of Educational and Behavioural Statistics* **25** (in press).

Moerbeek M, van Breukelen GJP, Berger MPF (2000b) On the design of experiments with binary responses in multilevel populations. *The Statistician* (in press).

Mok M (1995) Sample size requirements for 2-level designs in educational research. *Multilevel Modelling Newsletter* **7(2)**, 11–15 [available from http://www.ioe.ac.uk/multilevel/workpap.html].

Mollié A (1996) Bayesian mapping of disease. In: Gilks WR, Richardson S, Spiegelhalter DJ (eds). *Markov Chain Monte Carlo In Practice.* Chapman & Hall, Boca Raton FL, 360–79.

Mollié A (1999) Bayesian and empirical Bayes approaches to disease mapping. In: Lawson A, Biggeri A, Böhning D, Lesaffre E, Viel J-F, Bertolini R (eds) *Disease Mapping and Risk Assessment for Public Health.* Wiley, Chichester 15–29.

Morris CN, Christiansen CL (1996). Hierarchical models for ranking and for identifying extremes, with applications. In: Bernardo JO, Berger JO, Dawid, AP, AFM (eds) *Bayesian Statistics 5.* Oxford University Press, 277–97.

Muthén BO (1994) Multilevel covariance structure analysis. *Sociological Methods and Research* **22**, 376–98.

Muthén LK, Muthén BO (1998) *Mplus User's Guide.* Muthén & Muthén, Los Angeles.

New York State Department of Health (1995) *Coronary Artery Bypass Surgery in New York State, 1991–1993, New York State Department of Health, Albany.*

New York State Department of Health (1996) *Coronary Arteny Bypass Surgery in New York State, 1992–1994,* New York State Department of Health, Albary.

NHS Executive (1995) *The NHS Performance Guide 1994–1995.* NHS Executive, Leeds.

Normand S-L T, Glickman ME, Ryan TJ (1995). Modelling mortality rates for elderly heart attack patients: profiling hospitals in the co-operative cardiovascular project. In: Gatsonis C, Hodges J, Kass RN, Singpurwalla N (eds) Case Studies in Bayesian Statistics. Springer-Verlag, New York, 435–56.

O'Donnell O, Propper C, Upward, R. (1993). United Kindom. In: Van Doorslaer E, Wagstaff A, Rutten F (eds) *Equity in the Finance and Delivery of Health Care: An International Perspective* Oxford University Press, 237–261.

OPCS (Office of Population Censuses and Surveys) (1989, 1990) *Communicable Disease Statistics: Series MB12.* HMSO, London.

Palmer MJ, Phillips BF, Smith GT (1991) Application of nonlinear models with random coefficients to growth data. *Biometrics* **47**, 623–35.

Pan HQ, Goldstein H (1998) Multilevel repeated measures growth modelling using extended spline functions. *Statistics in Medicine* **17**, 2755–70.

Pfefferman D, Skinner CJ, Holmes DJ, Goldstein H, Rasbash J (1998) Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society B* **60**, 23–40.

Plewis I (1985). *Analysing Change.* Wiley, Chichester.

Plewis I (1993) Reading progress. In: Woodhouse G (ed.) *A Guide to ML3 for New Users* Multilevel Models Project, Institute of Education, London.

Pocock SJ (1983) *Clinical Trials: A Practical Approach.* Wiley, Chichester.

Ramsey JB (1969) Tests for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society B* **31**, 350–71.

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

Rasbash J, Browne W, Goldstein H, Yang M et al. (1999a) *A user's guide to MLwiN*, 2nd edn. Institute of Education, London.

Rasbash J, Healy M, Browne W, Cameron B, Charlton C (1999b) *MLwiN v1.01*. Multilevel Models Project, Institute of Education, London.

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

Raudenbush SW (1997) Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods* **2**, 173–185.

Rice N, Leyland A (1996) Multilevel models: applications to health data. *Journal of Health Services Research and Policy* **1**, 154–64.

Richardson S, Green PJ (1997) On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society B* **59**, 731–92.

Ripley B (1987) *Stochastic Simulation*. Wiley, New York.

Rodriguez G, Goldman N (1995) An assessment of estimation procedure for multilevel models with binary responses. *Journal of the Royal Statistical Society A* **158**, 73–89.

Rothman KJ, Boice JD (1979) *Epidemiological Analysis with a Programmable Calculator*. NIH Publication 79–1649. US Government Printing Office, Washington, DC.

Rowan KM, Kerr JH, McPherson K, Short A, Vessey MP (1993) Intensive care society's APACHE II study in Britain and Ireland – II: Outcome comparisons of intensive care units after adjustment for case mix by the American APACHE II method. *British Medical Journal* **307**, 977–81.

Rubin DB (1976) Inference and missing data. *Biometrika* **63**, 581–92.

Rubin DB (1987) *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York.

SAS (1992) SAS/STAT software: changes and enhancements. Technical report, SAS Institute, Cary, NC.

Schluchter M (1988) BMDP5–V – unbalanced repeated measures models with structured covariance matrices. Technical Report 86, BMDP Statistical Software, Los Angeles.

Schneider EC, Epstein AM (1996) Influence of cardiac-surgery performance reports on referral practices and access to care. *New England Journal of Medicine* **335**, 251–6.

Scott A, Shiell A (1997a) Do fee descriptors influence treatment choices in general practice? A multilevel discrete choice model. *Journal of Health Economics* **16**, 323–42.

Scott A, Shiell A (1997b) Analysing the effect of competition on general practitioners' behaviour using a multilevel modelling framework. *Health Economics* **6**, 577–88.

Scottish Office (1995) *Clinical Outcome Indicators – 1994*. Clinical Resource and Audit Group, Edinburgh.

Singer JD (1998) Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics* **23**, 323–55.

Skinner CJ, Holt D, Smith TMF (eds) (1989) *Analysis of Complex Surveys*. Wiley, New York.

Smans M, Muir CS, Boyle P (1992) *Atlas of Cancer Mortality in the European Economic Community*, International Agency for Research or Cancer, Lyon.

Snijders TAB, Bosker RJ (1993) Standard errors and sample sizes for two-level research. *Journal of Educational Statistics* **18**, 237–59.

Snijders TAB, Bosker RJ (1999) *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling. Sage, London.*

Spiegelhalter DJ (1998) Bayesian graphical modelling: a case study in monitoring health outcomes. *Applied Statistics* **47**, 115–34.

Spiegelhalter DJ, Thomas A, Best NG, Gilks WR (1995) BUGS: Bayesian inference using Gibbs sampling, version 0.50. Technical Report, MRC Biostatistics Unit, Cambridge.

Steele F, Diamond I, Amin S (1996) Immunisation uptake in rural Bangladesh: a multilevel analysis. *Journal of the Royal Statistical Society A* **159**, 289–99.

Steen Carlsson K, Lyttkens CH (1997) *Health Care Consumption and Individual Consumption Possibilities*. Department of Community Medicine, University of Malmö.

Stern H, Cressie N (1999) Inference for extremes in disease mapping. In: Lawson A, Biggeri A, Böhning D, Lesaffre E, Viel J-F, Bertolini R (eds) *Disease Mapping and Risk Assessment for Public Health*. Wiley, Chichester, 63–84.

Taylor A (1994) Appendix: Sample characteristics, attrition and weighting. In: Buck N, Gershuny J, Rose D, Scott J (eds) *Changing Households: The British Household Panel Survey 1990–92*. ESRC Research Centre on Micro-Social Change, colchester, Essex.

Thomas N, Longford NT, Rolph JE (1994) Empirical Bayes methods for estimating hospital specific mortality rates. *Statistics in Medicine* **13**, 889–903.

Thompson, SG, Pocock SJ (1991) Can meta-analyses be trusted? *Lancet* **338**, 1127–30.

van der Leeden R, Vrijburg K, de Leeuw J (1991) A review of two different approaches for the analysis of growth data using longitudinal mixed linear models: Comparing hierarchical linear regression (ML3, HLM) and repeated measures design with structured covariance matrices (BMDP-5V). UCLA Statistics Preprint 98.

van der Leeden R, Vrijburg K, de Leeuw J (1996) A review of two different approaches for the analysis of growth data using longitudinal mixed linear models. *Computational Statistics and Data Analysis* **21**, 583–605.

Van Doorslaer E, Wagstaff A, Rutten, F (eds) (1993) *Equity in the Finance and Delivery of Health Care: An International Perspective*. Oxford University Press.

Wagstaff A, Van Doorslaer, E. (1993) Equity in the finance and delivery of health care: concepts and definitions. In: Equity in the Finance and Delivery of Health Care: An international perspective. Van Doorslaer E, Wagstaff A, Rutten F (eds). Oxford University Press, 7–19.

Wakefield J, Elliott P (1999) Issues in the statistical analysis of small area health data. *Statistics in Medicine* **18**, 2377–99.

WHO (1991a) *Technical bases for the WHO Recommendations on the Management of Pneumonia in Children at First-Level Health Facilities*. WHO/ARI/91.20, Geneva.

WHO (1991b) *Programme for Control of ARI. Case Management of the Young Child with an ARI*. WHO, Geneva.

Woodhouse G (1995) *A Guide to MLn for New Users*. Multilevel Models Project, Institute of Education, London.

Woodhouse G, Yang M, Goldstein H, Rasbash J (1996) Adjusting for measurement error in multilevel analysis. *Journal of the Royal Statistical Society A* **159**, 201–12.

Wright DB (1997) Extra-binomial variation in multilevel logistic models with sparse structures. *British Journal of Mathematical and Statistical Psychology* **50**, 21–9.

Yang M (1997) Multilevel models for multiple category responses – a simulation, *Multilevel Modelling Newsletter, Multilevel Models Project, Institute of Education*, Vol. 9, No. 1, 9–15.

Yang M, Goldstein H, Heath A (2000) Multilevel models for repeated binary outcomes: attitudes and voting over the electoral cycle. *Journal of the Royal Statistical Society Series A* **163**, 49–62.

Yang M, Rasbash J, Goldstein H (1998) *MLwiN Macros for Advanced Multilevel Modelling*. Institute of Education, London.

Yang M, Rasbash J, Goldstein H, Barbosa M (1999) *MLwiN Macros for Advanced Multilevel Modelling*. Institute of Education, London.