26

Software for Fitting Multilevel Models

Andrzej T. Gałecki and Brady T. West

University of Michigan, USA

26.1 INTRODUCTION

Today's applied statistician or research analyst has the luxury of working with a variety of powerful statistical software procedures that enable users to efficiently fit many of the multilevel models covered in this volume. Developments in software for multilevel models, including linear (LMMs), generalized (GLMMs), and non-linear (NLMMs) mixed-effects models, are propelled both by advances in statistical methodology and technological progress in meeting computational demands.

With this chapter, we aim to provide readers with some history and an overview of the current software procedures that are available for fitting these models, and the abilities of the software to accommodate the many analytic aspects that accompany this broad class of models.

This chapter begins in Section 26.2, with some background information and an overview of important historical developments leading to the development of the powerful software available today. An overview of currently available software is given in Section 26.3. Section 26.4 discusses implementation details for the various procedures. Sections 26.5 and 26.6 conclude the chapter by discussing available software for fitting more advanced models, and directions for future developments.

26.2 BACKGROUND

For many years, software for the analysis of clustered and/or longitudinal data relied on simplifying model assumptions and employed straightforward computational algorithms. In the case of linear models, software implementing repeated measures analysis of variance (rmANOVA) models was the only option for longitudinal data. Fitting these models required data sets with very rigid structures: balanced counts of observations between groups and over time, no missing data, homogeneity of variance components, and no time-dependent covariates. For non-linear models, a very popular two-stage approach could be implemented using standard non-linear regression software, followed by simple ANOVA-type analysis of subject-specific estimates obtained in the first stage. This overly simplistic approach was based on the premise that subject-specific estimates obtained from individual non-linear fits were exactly equal to true values of subject-specific parameters. This approach suffered from similar shortcomings to those exhibited by rmANOVA; for example, time-varying covariates were not allowed. Moreover, overall variability of the dependent variable was not properly partitioned into between- and within-cluster/subject variation, leading to biased estimates of variance components.

A noteworthy exception in this regard was the NONMEM procedure, developed at the University of California for non-linear mixed-effects models dating back to 1980 (Beal and Sheiner, 1980). At its early stages, NONMEM employed an estimation method referred to as extended least squares (Sheiner et al., 1972; Beal and Sheiner, 1982), which was not widely accepted outside of the pharmaceutical community.

With the advent of personal computers, the progress in developing new software was greatly accelerated. In the mid 1980s, software programs for fitting mixed-effects models diverged into three broad groups, roughly ordered with respect to increasing levels of generality and computational difficulty: LMMs, GLMMs, and NLMMs.

In this section, we first focus on frequentist approaches to model fitting, followed by Bayesian approaches.

26.2.1 Frequentist Approaches

From the estimation point of view, we will first focus on the likelihood-based approach for mixed-effects models. The common feature of procedures using this approach is that the objective function, i.e., *marginal likelihood*, is expressed in the form of an integral over random effects, which has important implications for how the computations are performed.

In what follows, we briefly present discussions of approaches for fitting LMMs, GLMMs, and NLMMs, with references to methodological literature highlighting important computational and software developments.

Linear mixed-effects models

Over the years, the list of available software for fitting LMMs has become very impressive. We start our presentation with software for fitting LMMs, as specified by Laird and Ware (1982), for at least three reasons. First, these are the simplest types of mixed-effects models. Second, in many cases, they are the primary choice for the analysis of longitudinal and/or clustered data because of their numerical stability, substantive theoretical advances, and the availability of the software. Third, several aspects of LMMs are shared by and, to varying degrees, often adopted by more advanced models. By considering LMMs first, we set the stage for the more advanced classes of models described in the remainder of this chapter.

In their seminal paper, Laird and Ware (1982) introduced the specification of LMMs for longitudinal data and demonstrated how to obtain likelihood-based estimates using the EM algorithm (Dempster et al., 1977). In contrast to previous approaches, Laird–Ware's specification allowed for time-dependent covariates, missing values of the dependent variable, and structured variance-covariance matrices.

In the late 1980s, the 5V function (Schluchter, 1988) for two-level models for longitudinal and repeated measures data was introduced in BMDP, a general purpose statistical package. Origins of the two major software packages for fitting multilevel models, i.e., HLM (Bryk et al., 1988) and MLwiN, with ML2 (Rasbash et al., 1989), ML3 (Prosser et al., 1991), and MLn being predecessors of the latter, also date back to that time. These programs became the software of choice for analysts for many years to come.

The popularity of BMDP/5V diminished after SAS/PROC MIXED was introduced in 1992. Originally, the stand-alone software VARCL (Longford, 1990) was also one of the major packages, but its development, similar to that of BMDP/5V, has been discontinued. Additional software packages developed in later years are described in Section 26.3.

Nowadays, it is difficult to imagine a general purpose statistical software package without a module or function for fitting LMMs.

Other developments have involved faster algorithms based on sparse matrices (George and Liu, 1981) and are designed for large-scale problems encountered in fields such as genetics. Software employing this type of methodology includes ASReml (under development since 1992), SAS/PROC HPMIXED (SAS Institute Inc., 2010) and the `lmer` function in the R/**lme4** package (Bates et al., 2011), first released in 2005. We note that for the latter two programs the increased execution speed comes at the expense of a limited choice of variance-covariance structures for random effects and errors.

With respect to estimation, the key feature of LMMs is that the marginal likelihood can be expressed in a closed form, which essentially allows the software to directly employ standard optimization algorithms, such as Newton–Raphson (N–R) and Fisher scoring (Jennrich and Schluchter, 1986), to obtain maximum likelihood estimates.

Generalized linear mixed-effects models

In contrast to LMMs, the key computational difficulty for GLMMs is that, in general, the integral used in the marginal likelihood representation does not have a closed-form solution. Consequently, computational methods need to circumvent this difficulty, most often by approximating the likelihood before proceeding to the likelihood optimization. The choice of the approximation method may have considerable implications for the properties of the parameter estimates, and we therefore elaborate on this issue in more detail.

First, we note that a variety of approximations of the marginal likelihood have been proposed in the literature. One group of approximations, referred to as *deterministic*, provide an analytic formula for approximating the likelihood. Selected examples of deterministic approximations include:

- Non-adaptive Gaussian-Hermite quadrature (Stroud and Secrest, 1966);
- Laplace (Tierney and Kadane, 1986) approximation of the integrand function;
- Quasi-likelihood approaches, e.g., marginal (MQL) in Goldstein (1991) and Breslow and Clayton (1993), and penalized (PQL) in Breslow and Clayton (1993) and Wolfinger and O'Connell (1993);
- Adaptive Gaussian quadrature (AGQ) in Liu and Pierce (1994) and Pinheiro and Bates (1995a).

Approximations based on non-adaptive Gauss-Hermite quadrature are computationally intensive. This type of approximation was implemented for the first time for selected examples of two-level models in EGRET (Mauritsen, 1990).

Approximations of the likelihood based on the Laplace and quasi-likelihood methods are simple to implement but, at the same time, their early versions were fairly crude. Rodriguez and Goldman (1995) report that MQL-based estimates of fixed effects and variance components are subject to substantial downward bias when the random effects are relatively large. A similar observation was made for the PQL method in Breslow and Clayton (1993).

The Laplace approximation in Tierney and Kadane (1986) was implemented in early versions of the HLM software. Quasi-likelihood methods have been implemented in several software procedures, including earlier versions of SAS/PROC GLIMMIX, HLM, MLwiN, and others.

To alleviate the shortcomings of the aforementioned early implementations, the Laplace approximation was refined in Raudenbush et al. (2000). Similarly, the quasi-likelihood methods were also improved by using second-order terms in the Taylor expansion (Rodriguez and Goldman, 1995; Goldstein and Rasbash, 1996). Subsequently, these refinements were implemented in the HLM and MLwiN software.

AGQ, a substantial improvement over regular non-adaptive Gaussian-Hermite, was implemented for the first time in SAS/PROC NLMIXED (Wolfinger, 1999) and later in the `gllamm` command in Stata (Rabe-Hesketh et al., 2004). It is well known that by choosing one quadrature point AGQ becomes Laplace's approximation. By increasing the number of quadrature points, the accuracy of the approximation can be increased at the expense of longer execution times.

Alternative likelihood approximations are based on Monte Carlo integration methods and are referred to as *stochastic approximations* or, equivalently, *simulated maximum likelihood methods*. They employ, for example, Monte Carlo type or importance sampling based algorithms to simulate the likelihood function (Pinheiro and Bates, 1995a). On the one hand, stochastic approximations introduce another source of error, i.e., error due to sampling variance; on the other hand, with a large number of simulations, this error becomes negligible. Moreover, sampling variance can be used to assess the accuracy of the stochastic approximation.

When fitting GLMMs, approximation of the marginal likelihood is followed by an optimization step, most commonly the Expectation–Maximization (EM) algorithm (Dempster et al., 1977), N–R, or Fisher scoring. Each of the aforementioned approximation methods of the marginal likelihood can be combined with one of the optimization methods, leading to a large number of choices in finding optimal values for parameter estimates.

Other approaches, such as hierarchical likelihood, or h-likelihood (Lee and Nelder, 1996, 2001), attempt to avoid integration of the likelihood over random effects altogether by employing an objective function that is different from the marginal likelihood. This approach has been implemented, for example, in the **hglm** package (Alam et al., 2010) in the R software.

Nonlinear mixed-effects models

With respect to likelihood optimization, NLMMs with Gaussian errors share a similar challenge to that encountered for GLMMs, i.e., an intractable integral in the definition of the marginal likelihood. Approximations similar to those used for GLMMs have been implemented in modern software procedures.

The most common approximations are based on a linear Taylor expansion of the model mean function around current estimates of the fixed effects and/or empirical Bayesian (EB) estimates of the random effects. This linearization-based approach is conceptually similar to the one used in the quasi-likelihood approach for GLMMs.

The first-order (FO) linearization proposed in Beal and Sheiner (1982) was implemented, for example, in NONMEM. It performs in a satisfactory way if random effects enter the model in a linear fashion. In this approximation, the Taylor expansion is performed with EB estimates of the random effects set to zero.

Conditional first-order (FOCE) linearization, drawing on a method presented by Lindstrom and Bates (1990), is an improvement over the FO method in a similar way that PQL is an enhancement of MQL. In contrast to the FO method, it involves expansion of the model function around parameter estimates, including the current values of the EB estimates. This method is carried out in NONMEM. It is also the basis of the estimation algorithm implemented in the `nlme` function of the R/**nlme** package, and in the SAS NLINMIX macro (Littell et al., 1996, 2006).

Similar to developments for GLMMs, more recent approaches involve stochastic approximations of the likelihood function and are implemented, for example, in the Monolix software (Lavielle, 2008).

26.2.2 Bayesian Approaches

The Bayesian framework is especially well suited for multilevel models due to their hierarchical structure. The key feature in model specification, distinguishing from the frequentist approach, is the need for a prior distribution for the model parameters. Bayesian inference itself is based on the posterior distribution of the parameters given data. In general, the posterior distribution is analytically intractable for multilevel models. Thanks to Markov Chain Monte Carlo (MCMC) methods, this distribution can be stochastically and efficiently estimated based on generated samples. A useful reference in this context is Congdon (2005).

In the mid 1990s, the WinBUGS software (Spiegelhalter et al., 1996a,b,c) was developed for Bayesian analysis with MCMC chains constructed using the Gibbs sampler. The beginnings of this program date back to 1989, when its predecessor BUGS (Bayesian inference Using Gibbs Sampling) was introduced. Currently, the development of the program is focused on the OpenBUGS project, drawing on the open source concept. In addition, users of the R software can call WinBUGS and OpenBUGS from R, and Gelman and Hill (2007) provide a good reference on this practice.

WinBUGS, available at `http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml`, can be used for a wide range of multilevel models. Bayesian analysis using the MCMC approach was also implemented for fitting GLMMs in the MLwiN software (Browne, 2009) and in the **MCMCglmm** (Hadfield, 2010) R package. Other software, for example MPlus (Muthén and Muthén, 1998–2010) and JAGS (Plummer, 2003), can also be used for Bayesian analysis of multilevel models. More recently, alternatives to the MCMC approach based on analytic approximations have been developed in Rue et al. (2009) and implemented in the R package **inla**.

26.2.3 *Recap*

Interestingly, many different groups have worked on developing software for multilevel models. Apart from general-purpose statistical software packages, such as SAS, SPSS, BMDP, Stata, and R, stand-alone software has been extensively developed. This software, dedicated specifically to multilevel models, was geared toward various audiences and became popular within specific disciplines. Examples include the HLM software's popularity in education and social science research, and aML's use by econometricians. NONMEM is a leading software tool in the pharmaceutical community for analysis of data from population pharmacokinetics/pharmacodynamics (PK/PD) studies. ASReml is generally dedicated to analytical problems in agriculture and genetics. There are also examples of applications of the **lme4** package in R to ecological and linguistic research.

Due to dynamic progress, stand-alone software dedicated to fitting multilevel models (e.g., MLWin and HLM) had a fairly complicated history, including changing names, alternating between research or production phases, and some of them being commercialized. We note similarities in developmental paths, however. Initially, this software was developed for models with two levels for a linear case. The next step was adding levels to a model's hierarchy, so multilevel linear models could be addressed. In the mid 1990s, with the advent of the Windows operating system, these software packages added graphing capabilities and some of them added "Win" to the software names.

In general, software for fitting LMMs came first, followed by software for fitting GLMMs, and possibly other classes of models. Recent advances have also resulted in procedures capable of fitting both LMMs and GLMMs in general-purpose statistical computing packages like SAS, SPSS, Stata, and R.

Software procedures for fitting NLMMs were developed in parallel, often as separate products. Interfaces between the different software were developed to gain flexibility and take advantage of other software environments.

Marginal likelihood approximations for GLMMs and NLMMs using deterministic and stochastic methods are critical components of computational algorithms and can be performed in a variety of ways. Unfortunately, in the case of many software products information about the approximation method being used is not readily available.

The Bayesian approach to model fitting and inference was limited in early development stages by computer speed, but this is not currently an issue. This is partially due to increased computing speed, but also due to the advent and widespread use of MCMC methods. More recently, analytical approximations of the posterior marginal distribution have been proposed. Consequently, the Bayesian approach provides a very attractive, unified framework, which allows analysts to systematically address a variety of multilevel and more advanced models.

Our presentation of historical developments is largely based on the work of Kreft et al. (1994), de Leeuw and Kreft (2001), Pillai et al. (2005), and de Leeuw and Meijer (2008).

At present, there are various venues of support for users thanks to internet resources. Examples of relevant web pages include `http://www.bristol.ac.uk/cmm/` and `http://statcomp.ats.ucla.edu/mlm/` at Bristol University and UCLA, respectively. Discussion lists, such as `r-sig-mixed-models@r-project.org` for R users, provide a vivid forum for discussing issues related to multilevel modeling.

In our personal view, the most influential software, besides NONMEM, developed over the years, includes HLM, MLwiN, SAS/PROC MIXED, and SAS/PROC NLMIXED. The `gllamm` command in Stata has proven widely useful, due to its versatility and ability to address models other than LMMs and GLMMs (see Section 26.5). These developments will certainly continue in the future, and the next section addresses the different software procedures that are currently available.

26.3 AVAILABLE SOFTWARE

In this section, we present a broad range of currently available software that can be used to fit many examples of LMMs, GLMMs, and NLMMs that frequently arise in a variety of applied fields. For the sake of simplicity, we concentrate on parametric models with normally distributed random effects. Extensions of this basic modeling framework and a discussion of related software procedures are deferred to Section 26.5.

Also for the sake of simplicity, we tentatively divide the available software into three groups: general-purpose software, stand-alone software dedicated for multilevel models, and software designed for a wider class of models encompassing mixed-effects models as a special case.

26.3.1 General-Purpose Software

In Table 26.1, we present examples of functions and modules available for fitting mixed-effects models in general-purpose statistical software. Typically, as reflected in the columns' labels, the available software contains functions/modules designed for a specific class of models.

Major general statistical software packages include SAS, SPSS, Stata, and R. Relevant references in this context are Littell et al. (2006), SAS Institute Inc. (2008), SPSS Inc. (2010), Rasbash (2004), West (2009), StataCorp. (2009), R Development Core Team (2010), and Pinheiro and Bates (2000). GENSTAT (Payne et al., 2009), SYSTAT, and STATISTICA are other general-purpose statistical software packages not included in the table that allow users to fit similar classes of models.

We note that the R/**nlme** package with the `lme` and `nlme` functions was available in S-plus (Pinheiro and Bates, 1995b) before being implemented in R. Also, the R/**lme4** package is under ongoing development. We also note that Stata has developed two specialized commands for selected GLMMs, namely `xtmelogit` and `xtmepoisson`.

It is also worth noting that although modules/functions are often limited to a particular class of models, their strength comes from being immersed in a rich software/package environment. Several examples of LMM syntax for HLM software and appropriate functions listed in Table 26.1 are given in West et al. (2007).

Table 26.1 General-Purpose Statistical Software Packages Containing Modules/ Functions Designed To Fit Various Classes of Mixed-Effects Models. Software version/approximate year of the corresponding first release are given in parentheses

Software/	Class of mixed-effects models		
Package	LMM	GLMM	NLMM
SAS	`PROC MIXED` (6.07/1992)	`PROC GLIMMIX` (9.1/2006)	`PROC NLMIXED` (8.0/1999)
SPSS	`MIXED` (11.0/2001)	`GENLINMIXED` (19.0/2010)	
Stata	`xtmixed` (9/2005)	`gllamm` (8/2003)	
R/**nlme**	`lme` (1999)		`nlme` (1999)
R/**lme4**	`lmer` (2005)	`glmer` (2006)	`nlmer` (2007)

26.3.2 *Software Dedicated to Multilevel Models*

In Table 26.2, we include examples of stand-alone software dedicated to fitting multilevel models, or at least selected classes of these models.

In the upper part of Table 26.2, we include stand-alone software applicable to a fairly wide range of multilevel models. HLM and MLwiN are major software products in this group. Both are commercially distributed by Scientific Software International (SSI) and the Centre for Multilevel Modeling based at the University of Bristol, respectively. We note that MLwiN functionality is substantially enhanced by the presence of macro capability, allowing users to write their own functions. EGRET stands for Epidemiological GRaphics Estimation Testing, originally developed at the University of Washington. This software product is currently available from Cytel.

The Mixed-Up Suite was originally developed at the University of Illinois at Chicago. It is also referred to as MIXFOO, and contains the MIXOR, MIXREG, MIXNO, MIXPREG functions written in FORTRAN for ordinal, normally distributed, nominal, and count response data, respectively. More recently, the aforementioned functions were combined into a single application SuperMix (Hedeker et al., 2008) distributed by SSI.

The name aML stands for applied Multilevel Modeling, and this is powerful statistical software for multilevel and multi-process models. In general, software outlined in the upper part of this table tends to address LMMs and GLMMs, but does not cover NLMMs.

In the lower part of Table 26.2, we include software applicable to a fairly narrow range of multilevel models. MLA (Busing et al., 2005) is research software designed to fit two-level models. An interesting feature of MLA, distinguishing it from many other software packages, is the implementation of bootstrap and jackknife techniques for variance estimation.

ASReml is specialized software designed for fitting linear mixed-effects models using REML. It was developed jointly by statisticians from the Biometrics Program of the NSW Department of Primary Industries in Australia and the Biomathematics Unit of Rothamsted Research in Britain. Initially it was stand-alone software; later, in 1997, its core was linked into Genstat (Release 4.1). At present it is distributed by VSN International Ltd (VSNi). SAS/PROC HPMIXED is a new procedure developed by SAS Institute, Inc. As we have already mentioned, both ASReml and SAS/PROC HPMIXED use techniques based on sparse matrices and are designed for efficiently fitting LMMs to large and complex clustered or longitudinal data sets.

Table 26.2 Software Dedicated to Fitting Multilevel Models

Software	First release	Reference(s)
Software for a wide range of models		
HLM	1988	Bryk et al. (1988)
		Raudenbush et al. (2004a)
MLwiN	1997	Rasbash et al. (2004)
EGRET	1990	Mauritsen (1990)
Mixed-Up Suite	1996	Hedeker and Gibbons (1996a,b)
		Hedeker (1998, 1999)
Supermix		Hedeker et al. (2008)
aML	2000	Lillard and Panis (2003)
Software for selected classes of models		
MLA	1994	Busing et al. (2005)
ASReml	1992	Butler et al. (2007)
		Gilmour et al. (2009)
SAS/HPMIXED	2010	SAS Institute Inc. (2010)
NONMEM	1980	Beal and Sheiner (2004)
Monolix	2008	Lavielle (2008)

NONMEM is a leading software package for fitting NLMMs in research communities conducting analysis of data from population PK/PD studies. Starting with Version 7 of NONMEM, ICON Developments Solutions has exclusive rights to this software. Monolix is a fairly new software package also designed for NLMMs, employing similar advanced estimation techniques to those recently implemented in NONMEM.

In stand-alone software dedicated to multilevel models, we anticipate more specialized options compared to the available software in general-purpose statistical packages. Compared to general-purpose software, the environment, e.g., graphics capabilities, may be less developed. We note that HLM, aML, and MLA are true multilevel software, in the sense that they allow for the specification of regression models at multiple levels, making them an attractive choice for these types of analyses.

26.3.3 Software for More General Models

In this section, we present software that can be used for fitting more general models than mixed-effects models. This software can also be used to fit mixed-effects models as a special case of the general class. Examples of the software are given in Table 26.3.

Interestingly, the `gllammm` command in Stata, included in Table 26.1, is an example of such a software procedure. More specifically, it allows the user to fit a more general class of models referred to as Generalized Linear Latent And Mixed Models (GLLAMM), described in Rabe-Hesketh et al. (2001) and Skrondal and Rabe-Hesketh (2004), with GLMMs obviously included as a special case.

The Mplus and LatentGold software packages were originally designed primarily for latent class models. More recent releases have also enabled analysts to fit GLMMs. Mplus and LatentGold are commercial products distributed by Muthén & Muthén and Statistical Innovations, respectively.

LISREL is designed for structural equation modeling, and is a commercial product distributed by Scientific Software International (SSI). Procedures for fitting GLMMs were implemented in Release 8.8 (2006).

The Automatic Differentiation Model Builder (ADMB) (Fournier, 1993–2011) is a very influential software package used in ecology and the field of fisheries.

Table 26.3 Software for Fitting a More General Class of Models

Software	First release	Reference(s)
MPlus	1998	Muthén and Muthén (1998–2010)
LatentGold	2000	Vermunt and Magidson (2008)
LISREL	1974	Jöreskog and Sörbom (2006)
ADMB	2004	Fournier (1993–2011)
WinBUGS	1996	Spiegelhalter et al. (1996a,b,c)

The distinguishing and attractive feature of ADMB is that the automatic differentiation (also known as algorithmic differentiation) capability is implemented. More specifically, the code to evaluate derivatives used for likelihood optimization is generated automatically by the computer program. ADMB-RE (Skaug and Fournier, 2004–2011), first introduced in 2004, is a software package that adds mixed-effects modeling capabilities to ADMB. Consequently, ADMB-RE enables users to include random effects in a wide range of non-linear models tackled by ADMB. Both ADMB and ADMB-RE are accurate, flexible, and computationally efficient programs.

WinBUGS is an MCMC-based package for Bayesian analysis developed at the Medical Research Council Biostatistics Research Unit in Cambridge, and can be used for a wide range of models, including multilevel models. As mentioned in Section 26.2.2, recent developments have also enabled users of R to call BUGS routines from R.

26.3.4 Recap

There are many software products currently available allowing analysts to fit multilevel models. For the sake of simplicity, we have classified the available software into three groups, with examples presented in Tables 26.1, 26.2, and 26.3.

The most popular software procedures available today for LMMs are SAS/PROC MIXED, the `lme` function in the R/**nlme** package, the `lmer` function in the R/ **lme4** package, HLM, the `xtmixed` command in STATA, and the MIXED procedure in SPSS. An extensive overview of the options currently available in software designed to fit LMMs can be found in West and Gałecki (2011). For GLMMs, the leading software procedures are SAS/PROC GLIMMIX, HLM, MlwiN, and the `gllamm` command in Stata. For NLMMs, the most important software procedures are SAS/PROC NLMIXED and NONMEM.

Over the last decade, we have observed an emerging trend in the development of software, initiated by the `gllamm` command in Stata, employing a unified framework for different but related classes of models, including multilevel, latent class, and/or structural equation models. We note, for example, that Mplus, LatentGold, LISREL, and ADMB, presented in Table 26.3 and originally designed for a different class of models, i.e., latent class or SEMs, have, over time, added the capability to fit multilevel models with random effects.

Examples of non-commercial products include R, the MIXED-Up suite, ADMB, aML, MLA, Monolix, and WinBUGS. It is also worth noting that in this group of software, R, ADMB, and OpenBUGS are developed as open source projects.

We also note that auxiliary software has been developed to address specific issues associated with mixed-effects modeling, such as sample size calculations (see Section 26.5).

26.4 IMPLEMENTATION DETAILS

In this section, we review some aspects of the implementation of procedures for fitting

LMMs, GLMMs, and NLMMs. Some of these aspects were already mentioned in Section 26.2 in a broader context. Here, we present them in more detail. We will stress issues/tasks common to LMMs, GLMMs, and NLMMs, and point out some differences.

26.4.1 Input Information

We consider two aspects of input information, namely data organization and model specification.

Data organization

The leading approach used by the majority of available software procedures is to use data in a "long" format, with multiple records per subject or cluster. Another very natural method, implemented in the HLM software, is to have data organized essentially in the form of a relational database, with separate tables of data for variables measured at different levels of the data hierarchy (e.g., for subjects at Level 2 and individual measurements on subjects at Level 1). In contrast to the aforementioned data formats, the "wide" format, used for example in the 5V BMDP software (one of the very first software tools for fitting LMMs), poses substantial difficulties in specifying models for longitudinal data with irregular time observations and is essentially no longer supported by today's software for fitting mixed-effects models.

Model specification

As stated in Chapter 2 of this volume, multilevel models can be specified in a variety of ways. In most software, syntax typically follows a general matrix formulation. For example, PROC MIXED employs a single-equation specification of mixed-effects models. In contrast, the HLM, MLwiN, MLA, and aML software packages implement model specification based on "level notation", which naturally corresponds to the hierarchy of multilevel models.

Current software procedures for fitting multilevel models offer a variety of ways for users to specify different components of hierarchical models. Some software procedures (e.g., HLM, MLwiN) offer users a graphical interface for specifying models in a multilevel manner, while others rely on syntax (e.g., SAS/PROC MIXED) or possibly selection of choices from a menu (e.g., SPSS MIXED) for specifications more in the spirit of a "single equation". In general, in a given software package, the specification of design matrices for fixed and random effects in LMMs is similar to specifications used for classical linear models. A desirable feature that future enhancements of these software procedures might include is to have well-integrated polynomials and regression splines, so that specification of the design matrices may be enhanced.

Crossed (as opposed to nested) random effects are increasingly implemented in many software packages. Examples include SAS/PROC MIXED, SPSS MIXED, HLM, and R/`lmer`. Goldstein (1999, 2010) describes a trick for setting up the crossed random-effects model for a hierarchical software package such as MLwiN or `gllamm` in Stata.

Software procedures have different syntax or different options allowing users to specify alternative covariance structures in these models, which is especially important for errors in the context of longitudinal continuous data. Although the choices of variance-covariance matrices may be impressive (e.g., SAS/PROC MIXED), it may also be desirable to have an option to allow for user-defined structures, as is done in the R/**nlme** package. Readers can refer to West et al. (2007) for more details about specifying alternative covariance structures in a variety of software packages capable of fitting LMMs.

In the context of GLMMs, it may be desirable to allow for user-defined link functions, and this feature of GLMMs has not yet been widely implemented. For NLMMs, an important feature is to allow for the variance function to depend on the mean value of the dependent variable. This feature is available, for example, in the NONMEM software, but it is not available in SAS/PROC NLMIXED.

Caution needs to be exercised when using some software. For example, in aML, a potential source of confusion for some users is that the labeling of hierarchical levels is the opposite of that used in most of the multilevel literature and software.

26.4.2 Estimation

In most of the software for multilevel models, estimation is primarily based on optimization of the likelihood function, but other methods such as iterative generalized least squares (IGLS) are also available.

In the *linear* case, likelihood-based software typically relies on optimization of a marginal likelihood for covariance parameters with fixed effects being profiled out. Optimization of the likelihood in its general form is complicated by underlying non-linear constraints, assuring that covariance matrices remain positive-definite. Some remedies, such as transformation of the variance-covariance parameters (resulting in an unconstrained optimization) have been proposed (Pinheiro and Bates, 1996). The most common computational approach for ML optimization is based on the N–R algorithm. In some cases, e.g., HLM or the `xtmixed` command in Stata, the EM algorithm is used to refine initial values. Other methods such as Ridge-stabilized N–R are also often available.

Most current software procedures provide users with a choice of REML or ML estimation, depending on the test of interest. Typically, restricted maximum likelihood (REML) is used as a default, due to its property of producing unbiased estimates of the covariance parameters. Software procedures may vary in terms of methods used to derive starting values for estimates of covariance parameters, and some software procedures (e.g., SAS/PROC MIXED) allow the user to specify starting values for the covariance parameter estimates. Most current software procedures, in addition to having default methods for optimizing likelihood functions, use various combinations of these methods to arrive at the final estimates. Current computing power generally makes differences between these methods trivial for models with nested random effects. Models with crossed random effects require more advanced methods that are becoming more widely implemented in LMM software procedures (e.g., R/**lmer**).

As mentioned in Section 26.2.1 in a *non-linear case*, i.e. GLMMs and NLMMs, optimization is typically preceded by approximation of the marginal likelihood. In this section, we elaborate on this issue in more detail in the context of implementation in a variety of software.

Due to its simplicity, a basic version of the Laplace approximation is implemented in many software packages. More advanced versions of this approximation introduced in Raudenbush et al. (2000) have been implemented in the HLM software.

Approximations based on non-adaptive Gaussian quadrature have been implemented, for example, in MIXOR, MIXNO, aML, and EGRET.

Quasi-likelihood methods are implemented by iteratively executing two steps, namely creating pseudo-data and fitting a working LMM. An example of such an approach is PQL, as implemented in the `glmmPQL` function from the **MASS** package (Venables and Ripley, 2002) in R.

AGQ is currently implemented in several popular software packages that are aimed at fitting GLMMs, e.g., SAS/PROC GLIMMIX, SAS/PROC NLMIXED, and the `gllamm` command in Stata. In software packages implementing both methods, one can choose between non-adaptive and adaptive Gaussian quadrature. Interestingly, SAS/PROC NLMIXED allows the user to modify the number of quadrature points during the iteration process, so that better approximation can be accomplished. Additionally, SAS/PROC NLMIXED allows for the importance sampling method, an example of stochastic approximation, to be used.

Another example of software implementing importance sampling is ADMB.

The Mplus software can provide users with both frequentist and Bayesian inference. The numerical integration of the marginal likelihood can be carried out with or without adaptive quadrature. Bayesian analyses conducted using Mplus employ MCMC methods.

Assuming that the method of marginal likelihood approximation is selected, the next step is to optimize the likelihood. The most commonly used techniques are the EM, N–R, and Fisher scoring algorithms. We briefly describe the EM algorithm only. At first glance, it appears that the EM algorithm manages to bypass the likelihood approximation step and allows the user to go directly to finding optimum values of parameter estimates. Unfortunately, at the E-step, the EM algorithm also requires the calculation of an intractable integral, and therefore the E-step needs to employ a deterministic or, alternatively, a stochastic approximation. An example of the latter approach, called stochastic approximation EM (SAEM), was introduced in Delyon et al. (1999) and implemented in the Monolix software. More specifically, this method replaces the analytically intractable integral in the E-step of the EM algorithm with one iteration of a stochastic approximation procedure.

A different estimation technique known as IGLS, presented in Goldstein (1986) and Davidian and Giltinan (1995), is referred to in Chapter 3 and is used less often in software. The IGLS method is implemented, for example, in the `lme` function from R/**nlme**, and can prove useful for fitting models where the variance depends on the mean.

26.4.3 Tools for Inference and Model Selection

In this section, we present selected tools for inference and model selection, focusing primarily on hypothesis tests. We conclude with a short note on information criteria, and more recent developments.

Although tests of hypotheses for both fixed effects and covariance parameters are often performed using approximate Wald-type tests, we describe them separately due to important differences.

We first discuss hypothesis tests for *fixed effects*. In the LMM case, software procedures will typically automatically compute F- or t-test statistics, approximate degrees of freedom for the underlying null distributions, and corresponding p-values. Software procedures will vary in terms of the approximations used (or available for use) for the degrees of freedom, and some software procedures (e.g., the `lmer` function in R) will not propose any degrees of freedom approximations and p-values will not be reported. Denominator numbers of degrees of freedom for these tests are currently a topic of vigorous debate.

GLMM estimation is also based on maximum likelihood principles, and therefore Wald-type tests for fixed effects can also be performed. A word of caution is needed when quasi-likelihood approximations of the likelihood are used, because the output is based on the results of fitting a working LMM to pseudo-data, not to observed data. We encounter this type of situation, for example, when fitting the models using the `glmmPQL` function from the **MASS** package in R.

Since analytical forms of null distributions for the test statistics are merely approximations in most cases when fitting mixed-effects models, it is often desirable to derive them empirically using bootstrapping, jackknife procedures (for example in MLA), or by applying Bayesian methods (i.e., generating draws from posterior distributions based on a fitted model, for example in the **lme4** R package).

Hypothesis tests for *covariance* parameters can also be performed using Wald tests. In the case where the null hypotheses for covariance parameters do not define the parameters to be on the boundary of a parameter space (e.g., a null hypothesis that the covariance between two random errors is equal to zero), asymptotic likelihood ratio tests for the parameters can be performed in a standard way by referring likelihood ratio test statistics to a χ^2 distribution with an appropriate number of degrees of freedom (typically the difference in the number of parameters between a null model and an alternative model). Difficulties arise, for example, when testing whether variance components are equal to 0 (which places the variance under the null hypothesis on the boundary of its parameter space) and consequently classical testing procedures are no longer valid. Some software procedures will present the asymptotic Wald test (i.e., the estimated variance component divided by its estimated standard error) for covariance parameters automatically; others might report confidence intervals for covariance parameters (see Bottai and Orsini, 2004, for a method implemented in Stata). The HLM software uses a χ^2 test for variance components explained by Bryk and Raudenbush (2002, pp. 63–4). Selected procedures for fitting LMMs, e.g., SAS/PROC GLIMMIX and the `xtmixed` command in Stata, have currently implemented asymptotic likelihood ratio tests for variance components based on mixtures of chi-square distributions (Gutierrez et al., 2001).

The exact null distribution of the likelihood ratio test statistic for a variance component under more general conditions (including small samples) has been defined (Crainiceanu and Ruppert, 2004), and the R/**RLRsim** package (Scheipl, 2010) implements likelihood ratio tests based on simulations from this distribution. Appropriate null distributions of likelihood ratio test statistics for multiple covariance parameters have not been derived to date; classical likelihood ratio tests comparing nested models with multiple variance components constrained to be 0 in the reduced model should be considered conservative. The `xtmixed` command in Stata, for example, makes explicit note of this when users fit models with multiple random effects. MCMC methods based on draws from posterior distributions defined by a given model are generally considered more appropriate for making inferences about covariance parameters; however, readily available software procedures for this type of inference (e.g., the `mcmcsamp` generic function from the **lme4** package in R) are not as common.

Akaike's (AIC) and Bayesian (BIC) information criteria are often included as a part of the default output after fitting models. Also, the **lmmlasso** package (Schelldorfer, 2011) in R allows for model selection using LASSO penalties. Methodology related to information criteria and LASSO penalties is discussed in Chapter 7 of this volume.

26.4.4 *Model Diagnostics*

Assessment of model diagnostics is more complicated in the case of multilevel models as opposed to a "one-level" model, given that there are more underlying assumptions. Most multilevel software can readily compute empirical Bayesian (EB) estimates of random effects and model-based residuals for assessing simple model diagnostics. Work by Schabenberger (2004) has allowed SAS users to examine influence diagnostics when using SAS/PROC MIXED and SAS/PROC GLIMMIX, but diagnostic procedures still tend to be fairly limited in most software procedures. See the case studies in West et al. (2007) for examples of simpler model diagnostics for LMMs that can be implemented in most current software packages, including plots of residuals and plots of EB estimates. The ability of a given software procedure to generate relevant diagnostic statistics for LMMs is an important consideration for model assessment, and future software developments are certainly needed in this area. See Chapter 24 for further discussion.

26.4.5 *Recap*

At present, desirable features for multilevel modeling software include hierarchical specifications of the model, supported by graphical representation (e.g., HLM). Efficient estimation algorithms, including accurate approximations of the marginal likelihood in a nonlinear case, need to be implemented. Availability of various structured covariance matrices at different levels of the model hierarchy (both for the random effects and errors), dealing with crossed random effects, finding starting values for variance-covariance parameters, assuring that variance-covariance matrices remain positive-definite during the iteration process, imposing user-defined constraints on variance parameters, allowing for multiple levels of model hierarchy, and utilities for testing for the need for random effects are all features that we feel should be included in "ideal" software for fitting mixed-effects models. Methods for empirically deriving the null distributions of test statistics also need to be more widely implemented in the available software. Finally, graphical facilities for visualizing model fit and diagnostics also tend to be lacking in much of the available software, and we hope that future developments will simplify generation of these types of graphics, especially in the spirit of Cleveland (1993). Several aspects of the multilevel model implementation in various software are compared by Ben Bolker at `http://glmm.wikidot.com/pkg-comparison`.

26.5 MORE ADVANCED MODELS

So far, this chapter has focused on fairly standard examples of multilevel models, namely LMMs, GLMMs, and NLMMs, with random effects being normally distributed. Here, we briefly discuss extensions of these models, going in a variety of directions. When relaxing some of the assumptions underlying these models, the models become even more flexible, and stand-alone and specialized software procedures are more likely to address these advanced issues.

Accounting for complex sample design features

Data sets having hierarchical structures are often produced from large complex samples, where several units of analysis are sampled from randomly sampled clusters of units, and the clusters of units may be stratified in some way. In addition, sampled units may not have equal probabilities of selection in complex multi-stage sample designs, leading to the use of weights to offset the unequal probabilities of selection (and also possibly account for nonresponse and calibration of estimates to known population totals). Pfeffermann et al. (1998), Asparouhov and Muthén (2006), and Rabe-Hesketh and Skrondal (2006) have developed theory for estimating multilevel models in a way that incorporates the weights, and Rabe-Hesketh and Skrondal (2006), Carle (2009), and Heeringa et al. (2010) have presented applications of these approaches using currently available software procedures. See Gelman (2007) for a summary of the issues surrounding the use of weights when fitting multilevel and mixed-effects models. In particular, the Mplus, HLM, and MLwiN software packages allow users to address complex sample design features when fitting a wide class of models. More research in this area is very likely in the near future.

Different types of the dependent variables

Dependent variables measured on an ordinal scale or with censored observations lead to models that are not examples of GLMMs. Examples of software implementing models for these types of dependent variables include MLWin and aML. Also, specific functions, such as the MIXOR and MIXSUR functions incorporated in the Mixed-Up suite or in the Supermix software, can be used.

In the context of NLMMs, such as those discussed in Chapter 14, an important role is being played by models expressed as a solution of ordinary differential equations (ODE). These can be fitted, for example, using the NONMEM software. In the general-purpose statistical package SAS, the NLMEM macro (Gałecki, 1998), which builds on the SAS macro NLINMIX, is a possible choice.

Similarly, the R/**nlmeODE** package (Tornøe et al., 2004; Tornøe, 2010), which builds on the `nlme` function in R/**nlme** package, can be used.

Joint/multi-process modeling

Multi-process or joint modeling refers to the capability of jointly fitting models to more than one dependent variable, measured over time or for clustered data.

Initial approaches to the problem of specifying a joint density of outcomes of the same type go back to Gałecki (1994) and were implemented for LMMs in SAS/PROC MIXED. Examples of joint modeling for outcomes of different types, for example continuous and discrete data, using standard software (e.g., SAS/PROC GLIMMIX and SAS/NLMIXED) are given in Molenberghs and Verbeke (2005). See Chapter 21 for further discussion of data of this type. For joint modeling of longitudinal and time-to-event data the **JM** (Rizopoulos, 2010) package in R can be used. The aML software appears to be especially well-suited for joint (or multi-process) modeling of correlated outcomes of *any* type.

Distributions for random effects

In some cases, the assumption of a normal distribution for the random effects may be too restrictive. Relaxing this assumption can be accomplished in a variety of ways.

In a nonparametric approach, the researcher may leave this (possibly continuous) distribution entirely unspecified. In such a case, the distribution function is estimated with a step function having a finite number of steps (Laird, 1978). The method of finding such a step function is called nonparametric maximum likelihood estimation (NPMLE) or fully semi-parametric estimation, and is implemented in the `gllamm` command in Stata.

Other choices of distributions for the random effects are possible. For example, the random effects might follow a mixture of two or more normal distributions, as presented in Verbeke and Lesaffre (1996). Both Mplus and LatentGold are well suited to fit these types of model.

Another way to relax assumptions about random effects is introduced in Hierarchical Generalized Linear Models, with generalized linear models specified at every level of the model hierarchy. Estimation based on the aforementioned h-likelihood method (see Section 26.2.1) is implemented in the HG system of facilities, called from GENSTAT, and in the R/**hglm** package (Rönnegård et al., 2010).

Another group of extensions aims to build on existing classes of models, and incorporate random effects into them. We discuss examples of these types of models below.

Generalized Additive Mixed-effects Models (GAMM)

This class of models builds on Generalized Additive Models (GAM) and allows flexible functional dependence of an outcome variable on covariates by using nonparametric regression. This class of models can be fitted for example using the R/**amer** package, which extends the range of models fitted by R/**lme4**. Implementation is based on the observation that the penalized least squares problem for smooth terms can be reformulated as a likelihood for mixed-effects models, with the smoothing parameter becoming a variance component (see Chapter 18).

Semi-parametric nonlinear mixed-effects models (SNMM) have also been proposed by Ke and Wang (2001). The function `snm` in the R/**assist** package (Wang and Ke, 2011) can be used for fitting models from this class. Similar models are considered in Chapter 17.

Random effects incorporated in Structural Equation Models (SEM)

These models build on SEMs by incorporating random effects (possibly to account for clustering of observations) and are implemented in programs such as LISREL and Mplus.

Social relations analysis

Analysis of social relations, which takes into account correlations of observations in dyads or groups of subjects, requires specialized software. Examples include the **TripleR** package (Schönbrodt et al., 2012) and StOCNET (Huisman and Van Duijn, 2003). The **TripleR** package is used for Social Relation Model (SRM) analyses for single or multiple round-robin groups. On the other hand, StOCNET is a project that builds on a software system developed for the advanced analysis of social networks and has been written in collaboration between software engineers of Science Plus and other researchers. Data of this type are discussed in Chapter 33.

Sample size considerations

Specialized software for power calculations includes, but it is not limited to, Raudenbush et al. (2004b) and Browne et al. (2009). Sample size calculations are discussed in the context of study design considerations in Chapter 11 of this volume.

26.6 CONCLUSION

The recent trend discussed in this chapter of developing software encompassing a wide range of models, in a unified framework enabling both likelihood-based and Bayesian approaches, will be continued in the future. Likelihood-based approaches are valid for fitting these models under the assumption that missing data are missing at random (MAR), or equivalently that missingness depends on the observed data, but not on the unobserved data. Future developments for mixed-effects models might consider models for data sets with data exhibiting a not missing at random (NMAR) mechanism (Little and Rubin, 1987, 2002).

With more powerful hardware, certain options for fitting these models will become more attractive and will (hopefully) be widely implemented (e.g., empirically deriving null distributions of test statistics and sampling from these distributions). This is especially the case with resampling methods such as the bootstrap, multiple imputation, and MCMC.

Given dynamic developments and the rapidly changing software landscape, we have merely touched on features of the software procedures without going into too much detail. In many places we have provided our personal impressions of the currently available software. This is a wide-ranging topic encompassing many disciplines, and we have indicated areas where we think that additional developments are needed. We take responsibility for any inaccuracies or omissions in this chapter.

ACKNOWLEDGEMENTS

We gratefully acknowledge valuable comments provided by Ben Bolker from McMaster University and Sue Welham from Rothamsted Research. We are thankful to the editors and anonymous reviewers of this chapter for their patience and feedback. We thank Marta Gałecki for editing early drafts of this chapter. We acknowledge support from the Claude Pepper Center Grants AG08808 and AG024824 from the National Institutes of Aging.

REFERENCES

Alam, M., Ronnegard, L., and Shen, X. (2010) *hglm: Hierarchical Generalized Linear Models*. R package version 1.1.1.

Asparouhov, T. and Muthén, B. (2006) 'Multilevel modeling of complex survey data', in *Proceedings of the American Statistical Association*, Seattle, WA: American Statistical Association.

Bates, D., Maechler, M., and Bolker, B. (2011) *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-42.

Beal, S. and Sheiner, L. (1980) 'The NONMEM system', *American Statistician*, 34: 118–19.

Beal, S. and Sheiner, L. (1982) 'Estimating population kinetics', *Critical Reviews in Biomedical Engineering*, 8: 195–222.

Beal, S. and Sheiner, L. (2004) *NONMEM User's Guide*.

Bottai, M. and Orsini, N. (2004) 'XTVC: Stata module to compute confidence intervals for the variance component of random-intercept linear models', *Statistical Software Components*, Boston College Department of Economics.

Breslow, N. and Clayton, D. (1993) 'Approximate inference in generalized linear mixed models', *Journal of the American Statistical Association*, 88: 9–25.

Browne, W. (2009) *MCMC Estimation in MLwiN, v2.13*. Centre for Multilevel Modeling, University of Bristol.

Browne, W., Golalizadeh Lahi, M., and Parker, R. (2009) *A guide to sample size calculations for random effect models via simulation and the MLPowSim software package*. Centre for Multilevel Modeling, University of Bristol.

Bryk, A. and Raudenbush, S. (2002) *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage.

Bryk, A., Raudenbush, S., Seltzer, M., and Congdon, R. (1988) *An Introduction to HLM: Computer Program and Users Manual*. Chicago: University of Chicago Dept. of Education.

Busing, F., Meijer, E., and van der Leeden, R. (2005) *MLA: Software for MultiLevel Analysis of Data with Two Levels. User's Guide for Version 4.1*. Universiteit Leiden.

Butler, D., Cullis, B., Gilmour, A., and Gogel, B. (2007) *ASReml-R reference manual*. State of Queensland Department of Primary Industries and Fisheries.

Carle, A. (2009) 'Fitting multilevel models in complex survey data with design weights: Recommendations', *BMC Medical Research Methodology*, 9: 49.

Cleveland, W. (1993) *Visualizing Data*. Summit, NJ: Hobart Press.

Congdon, P. (2005) *Bayesian Statistical Modeling*, 2nd edition. New York: Wiley.

Crainiceanu, C. and Ruppert, D. (2004) 'Likelihood ratio tests in linear mixed models with one variance component', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66: 165–85.

Davidian, M. and Giltinan, D. (1995) *Nonlinear Models for Repeated Measurement Data*. Boca Raton, FL: Chapman and Hall/CRC.

de Leeuw, J. and Kreft, I. (2001) 'Software for Multilevel Analysis', in A.H. Leyland and H. Goldstein (eds), *Multilevel Modeling of Health Statistics*. New York: Wiley, 187–204.

de Leeuw, J. and Meijer, E. (2008) *Handbook of Multilevel Analysis*. New York: Springer Verlag.

Delyon, B., Lavielle, M., and Moulines, E. (1999) 'Convergence of a Stochastic Approximation Version of the EM Algorithm', *Annals of Statistics*, 27: 94–128.

Dempster, A., Laird, N., and Rubin, D. (1977) 'Maximum Likelihood from Incomplete Data Via the EM Algorithm', *Journal of the Royal Statistical Society. Series B (Methodological)*, 39: 1–38.

Fournier, D. (1993–2011) *An Introduction to AD Model Builder for Use in Nonlinear Modeling and Statistics Version 10.0*.

Gałecki, A. (1994) 'General Class of Covariance Structures for Two or More Repeated Factors in Longitudinal Data Analysis', *Communications in Statistics – Theory and Methods*, 23: 3105–19.

Gałecki, A. (1998) 'NLMEM: A New SAS/IML Macro for Hierarchical Nonlinear Models', *Computer Methods and Programs in Biomedicine*, 55: 207–16.

Gelman, A. (2007) 'Struggles With Survey Weighting and Regression Modeling', *Statistical Science*, 22: 153–64.

Gelman, A. and Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

George, A. and Liu, J. (1981) *Computer Solution of Large Sparse Positive Definite Systems*. Englewood Cliffs, NJ: Prentice-Hall.

Gilmour, A., Gogel, B., Cullis, B., and Thompson, R. (2009) *ASReml User Guide Release 3.0*.

Goldstein, H. (1986) 'Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares', *Biometrika*, 73: 43–56.

Goldstein, H. (1991) 'Nonlinear Multilevel Models, With an Application to Discrete Response Data', *Biometrika*, 78: 45–51.

Goldstein, H. (1999, 2010) *Multilevel Statistical Models*. New York: Wiley.

Goldstein, H. and Rasbash, J. (1996) 'Improved Approximations for Multilevel Models With Binary Responses', *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159: 505–13.

Gutierrez, R., Carter, S., and Drukker, D. (2001) 'On boundary-value likelihood-ratio tests', *Stata Technical Bulletin*, 10(60): 15–18.

Hadfield, J. (2010) 'MCMC Methods for Multi-Response Generalised Linear Mixed Models: The MCMCglmm R Package', *Journal of Statistical Software*, 33(2): 1–22.

Hedeker, D. (1998) *MIXPREG: A Computer Program for Mixed-Effects Poisson Regression*. Chicago: School of Public Health, University of Illinois at Chicago.

Hedeker, D. (1999) 'MIXNO: A Computer Program for Mixed-Effects Nominal Logistic Regression', *Journal of Statistical Software*, 4(5): 1–92.

Hedeker, D. and Gibbons, R. (1996a) 'MIXOR: A Computer Program for Mixed-Effects Ordinal Regression Analysis', *Computer Methods and Programs in Biomedicine*, 49: 157–76.

Hedeker, D. and Gibbons, R. (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, R., du Toit, M., and Cheng, Y. (2008) *SuperMix: Mixed Effects Models*. Scientific Software International.

Heeringa, S., West, B., and Berglund, P. (2010) *Applied Survey Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC.

Huisman, M. and Van Duijn, M. (2003) 'StOCNET: Software for the Statistical Analysis of Social Networks', *Connections*, 25: 7–26.

Jennrich, R. and Schluchter, M. (1986) 'Unbalanced Repeated-Measures Models With Structured Covariance Matrices', *Biometrics*, 42: 805–20.

Jöreskog, K. and Sörbom, D. (2006) *LISREL 8.8 for Windows*.

Ke, C. and Wang, Y. (2001) 'Semiparametric Nonlinear Mixed-Effects Models and Their Applications', *Journal of the American Statistical Association*, 96: 1272–98.

Kreft, I., de Leeuw, J., and van der Leeden, R. (1994) 'Review of Five Different Statistical Packages for Hierarchical Linear Regression: BMDP-5V, GENMOD, HLM, ML3 and VARCL', *American Statistician*, 48: 324–35.

Laird, N. (1978) 'Nonparametric Maximum Likelihood Estimation of a Mixing Distribution', *Journal of the American Statistical Association*, 73: 805–11.

Laird, N. and Ware, J. (1982) 'Random-Effects Models for Longitudinal Data', *Biometrics*, 38: 963–74.

Lavielle, M. (2008) *Monolix Users Guide V.2.4*.

Lee, Y. and Nelder, J. (1996) 'Hierarchical Generalized Linear Models', *Journal of the Royal Statistical Society. Series B (Methodological)*, 58: 619–78.

Lee, Y. and Nelder, J. (2001) 'Hierarchical Generalised Linear Models: A Synthesis of Generalised Linear Models, Random-Effect Models and Structured Dispersions', *Biometrika*, 88: 987–1006.

Lillard, L. and Panis, C. (2003) *aML Multilevel Multiprocess Statistical Software, version 2.0*. Los Angeles, CA: EconWare.

Lindstrom, M. and Bates, D. (1990) 'Nonlinear Mixed Effects Models for Repeated Measures Data', *Biometrics*, 46: 673–87.

Littell, R., Milliken, G., Stroup, W., and Wolfinger, R. (1996) *SAS® System for Mixed Models*. Cary, NC: SAS Institute.

Littell, R., Milliken, G., Stroup, W., Wolfinger, R., and Schabenberger, O. (2006) *SAS® for Mixed Models*, 2nd edition. Cary, NC: SAS Institute.

Little, R. and Rubin, D. (1987, 2002) *Statistical Analysis With Missing Data*. New York: Wiley.

Liu, Q. and Pierce, D. (1994) 'A Note on Gauss-Hermite Quadrature', *Biometrika*, 81: 624–9.

Longford, N. (1990) *VARCL. Software for Variance Component Analysis of Data With Nested Random Effects (Maximum Likelihood)*. Princeton, NJ: Educational Testing Service.

Mauritsen, R. (1990) *EGRET Reference Manual*. Seattle: Statistics and Epidemiology Research Corporation.

Molenberghs, G. and Verbeke, G. (2005) *Models for Discrete Longitudinal Data*. New York: Springer.

Muthén, L. and Muthén, B. (1998–2010) *Mplus: Statistical Analysis With Latent Variables: User's guide*, 6th edition.

Payne, R., Harding, S., Murray, D., Soutar, D., Baird, D., Glaser, A., Channing, I., Welham, S., Gilmour, A., Thompson, R., et al. (2009) *The Guide to GenStat Release 12, Part 2: Statistics*.

Pfeffermann, D., Skinner, C., Holmes, D., Goldstein, H., and Rasbash, J. (1998) 'Weighting for unequal selection probabilities in multilevel models', *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60: 23–40.

Pillai, G., Mentré, F., and Steimer, J. (2005) 'Non-Linear Mixed Effects Modeling – From Methodology and Software Development to Driving Implementation in Drug Development Science', *Journal of Pharmacokinetics and Pharmacodynamics*, 32: 161–83.

Pinheiro, J. and Bates, D. (1995a) 'Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model', *Journal of Computational and Graphical Statistics*, 4: 12–35.

Pinheiro, J. and Bates, D. (1995b) *lme and nlme: Mixed-Effects Methods and Classes for S and S-Plus*.

Pinheiro, J. and Bates, D. (1996) 'Unconstrained Parameterizations for Variance-Covariance Matrices', *Statistics and Computing*, 6: 289–96.

Pinheiro, J. and Bates, D. (2000) *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

Plummer, M. (2003) 'JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling', in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, 20–22.

Prosser, R., Rasbash, J., Goldstein, H., and University of London Institute of Education (1991) *ML3, Software for Three-level Analysis: User's Guide*.

R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rabe-Hesketh, S., Pickles, A., and Skrondal, A. (2001) 'Gllamm: A Class of Models and a Stata Program', *Multilevel Modeling Newsletter*, 13(1): 17–23.

Rabe-Hesketh, S. and Skrondal, A. (2006) 'Multilevel Modeling of Complex Survey Data', *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 169: 805–27.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004) *GLLAMM manual*.

Rasbash, J. (2004) 'Statistical Software Review: Mixed Models in SPSS', *British Journal of Mathematical and Statistical Psychology*, 57: 189–90.

Rasbash, J., Prosser, R., and Goldstein, H. (1989) *ML2: Software for Two-Level Analysis, User's Guide*.

Rasbash, J., Steele, F., Browne, W., and Prosser, B. (2004) *A User's Guide to MLwiN, Version 2*.

Raudenbush, S., Bryk, A., Cheong, Y., and Congdon, R. (2004a) *HLM 6: Hierarchical Linear and Nonlinear Modeling*. Scientific Software International.

Raudenbush, S., Spybrook, J., Liu, X., and Congdon, R. (2004b) *Optimal Design for Longitudinal and Multilevel Research: Documentation for the Optimal Design Software*.

Raudenbush, S., Yang, M., and Yosef, M. (2000) 'Maximum Likelihood for Generalized Linear Models With Nested Random Effects Via High-Order, Multivariate Laplace Approximation', *Journal of Computational and Graphical Statistics*, 9: 141–57.

Rizopoulos, D. (2010) 'JM: An R Package for the Joint Modeling of Longitudinal and Time-to-Event Data', *Journal of Statistical Software*, 35(9), 1–33.

Rodriguez, G. and Goldman, N. (1995) 'An Assessment of Estimation Procedures for Multilevel Models With Binary Responses', *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 158: 73–89.

Rönnegård, L., Shen, X., and Alam, M. (2010) 'hglm: A Package for Fitting Hierarchical Generalized Linear Models', *R Journal*, 2(2): 20–28.

Rue, H., Martino, S., and Chopin, N. (2009) 'Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations', *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 71: 319–92.

SAS Institute Inc. (2008) *SAS/STAT® 9.2 Users Guide*.

SAS Institute Inc. (2010) *SAS/STAT® 9.22 Users Guide*.

Schabenberger, O. (2004) 'Mixed Model Influence Diagnostics. Paper 189–29', in *Proceedings of the Twenty-Ninth Annual SAS® Users Group International (SUGI) Conference*. Cary, NC: SAS Institute Inc.

Scheipl, F. (2010) *RLRsim: Exact (Restricted) Likelihood Ratio Tests for Mixed and Additive Models*. R package version 2.0–5.

Schelldorfer, J. (2011) *lmmlasso: Linear Mixed-Effects Models With Lasso*. R package version 0.1–2.

Schluchter, M. (1988) *BMDP 5V: Unbalanced Repeated Measures: Models with Structured Covariance Matrices*. BMDP Statistical Software.

Schönbrodt, F., Back, M., and Schmukle, S. (2012) 'TripleR: An R Package for Social Relations Analyses Based on Round-Robin Designs', *Behavior Research Methods*, 44: to appear.

Sheiner, L., Rosenberg, B., and Melmon, K. (1972) 'Modeling of Individual Pharmacokinetics for Computer-Aided Drug Dosage', *Computers and Biomedical Research*, 5: 441–59.

Skaug, H. and Fournier, D. (2004–2011) *Random Effects in AD Model Builder. ADMB-RE User Guide. Version 10.0*.

Skrondal, A. and Rabe-Hesketh, S. (2004) *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: CRC Press.

Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1996a) *BUGS 0.5: Bayesian Inference Using Gibbs Sampling Manual (Version ii)*.

Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1996b) *BUGS 0.5 Examples*, Volume 1.

Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1996c) *BUGS 0.5 Examples*, Volume 2.

SPSS Inc. (2010) *SPSS 19.0 Advanced Models*. Englewood Cliffs, NJ: Prentice Hall.

StataCorp (2009) *Stata Statistical Software: Release 11*. College Station, TX: Stata Corporation.

Stroud, A. and Secrest, D. (1966) *Gaussian Quadrature Formulas*. Upper Saddle River, NJ: Prentice-Hall.

Tierney, L. and Kadane, J. (1986) 'Accurate Approximations for Posterior Moments and Marginal Densities', *Journal of the American Statistical Association*, 81: 82–6.

Tornøe, C., Agersø, H., Jonsson, E., Madsen, H., and Nielsen, H. (2004) 'Non-Linear Mixed-Effects Pharmacokinetic/Pharmacodynamic Modeling in NLME Using Differential Equations', *Computer Methods and Programs in Biomedicine*, 76: 31–40.

Tornøe, C.W. (2010) *nlmeODE: Non-Linear Mixed-Effects Modeling in nlme Using Differential Equations*. R package version 12.2.2.

Venables, W.N. and Ripley, B.D. (2002) *Modern Applied Statistics with S*, 4th edition. New York: Springer.

Verbeke, G. and Lesaffre, E. (1996) 'A Linear Mixed-Effects Model With Heterogeneity in the Random-Effects Population', *Journal of the American Statistical Association*, 91: 217–21.

Vermunt, J. and Magidson, J. (2008) *LG-Syntax User's Guide: Manual for Latent GOLD 4.5 Syntax Module*.

Wang, Y. and Ke, C. (2011) *assist: A Suite of S-Plus Functions Implementing Smoothing Splines*. R package version 3.1.2.

West, B. (2009) 'Analyzing Longitudinal Data With the Linear Mixed Models Procedure in SPSS', *Evaluation and the Health Professions*, 32: 207–28.

West, B. and Gałecki, A. (2011) 'An Overview of Current Software Procedures for Fitting Linear Mixed Models', *The American Statistician*, 65: 274–82.

West, B., Welch, K., and Gałecki, A.T., with contributions from Gillespie B. (2007) *Linear Mixed Models: A Practical Guide Using Statistical Software*. Boca Raton, FL: Chapman and Hall/CRC.

Wolfinger, R. (1999) 'Fitting Nonlinear Mixed Models With the New NLMIXED Procedure', in *Proceedings of the 24th Annual SAS Users Group International Conference (SUGI 24)*, 1–10.

Wolfinger, R. and O'Connell, M. (1993) 'Generalized Linear Mixed Models: A Pseudo-Likelihood Approach', *Journal of Statistical Computation and Simulation*, 48: 233–43.