



## The American Statistician

Publication details, including instructions for authors and subscription information:  
<http://amstat.tandfonline.com/loi/utas20>

### An Overview of Current Software Procedures for Fitting Linear Mixed Models

Brady T. West<sup>a</sup> & Andrzej T. Galecki<sup>a</sup>

<sup>a</sup> Brady T. West is Assistant Research Professor, Survey Methodology Program, Survey Research Center, Institute for Social Research, and Assistant Research Scientist, Center for Statistical Consultation and Research, University of Michigan, Ann Arbor, MI 48109. Andrzej T. Galecki is Research Professor, Institute of Gerontology, Medical School, and Research Scientist, Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109. The authors wish to gratefully acknowledge the helpful comments from the Editor, an Associate Editor, and two anonymous referees, along with the detailed fact-checking provided by representatives from SAS, SPSS, Stata, SYSTAT, Mplus, and MLwiN. Any inaccuracies with regard to software procedures from the other packages are not intentional and may be inevitable given dynamic progress in this area; nevertheless, inaccuracies should be attributed to the authors. We also gratefully acknowledge support from the Claude Pepper Center Grants AG08808 and AG024824 from the National Institute of Aging.

Published online: 24 Jan 2012.

To cite this article: Brady T. West & Andrzej T. Galecki (2011) An Overview of Current Software Procedures for Fitting Linear Mixed Models, *The American Statistician*, 65:4, 274-282, DOI: [10.1198/tas.2011.11077](https://doi.org/10.1198/tas.2011.11077)

To link to this article: <http://dx.doi.org/10.1198/tas.2011.11077>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://amstat.tandfonline.com/page/terms-and-conditions>

# An Overview of Current Software Procedures for Fitting Linear Mixed Models

Brady T. WEST and Andrzej T. GALECKI

At present, there are many software procedures available that enable statisticians to fit linear mixed models (LMMs) to continuous dependent variables in clustered or longitudinal datasets. LMMs are flexible tools for analyzing relationships among variables in these types of datasets, in that a variety of covariance structures can be used depending on the subject matter under study. The explicit random effects in LMMs allow analysts to make inferences about the variability between clusters or subjects in larger hypothetical populations, and examine cluster- or subject-level variables that explain portions of this variability. These models can also be used to analyze longitudinal or clustered datasets with data that are missing at random (MAR), and can accommodate time-varying covariates in longitudinal datasets. Although the software procedures currently available have many features in common, more specific analytic aspects of fitting LMMs (e.g., crossed random effects, appropriate hypothesis testing for variance components, diagnostics, incorporating sampling weights) may only be available in selected software procedures. With this article, we aim to perform a comprehensive and up-to-date comparison of the current capabilities of software procedures for fitting LMMs, and provide statisticians with a guide for selecting a software procedure appropriate for their analytic goals.

**KEY WORDS:** Covariance structures; Longitudinal data analysis; Models for clustered data; Statistical software.

## 1. INTRODUCTION

Linear mixed models (LMMs) are flexible analytic tools for modeling correlated continuous data. Correlations among values on continuous dependent variables may arise from repeated measurements collected on study subjects, or from subjects being clustered in a way that would introduce similar values on the

measures of interest. These types of dependent variables arise in many different contexts from many different fields, where study designs may not result in balanced sets of continuous measures (either due to missing data or unbalanced designs). Due to their ability to accommodate these important features of continuous data collected in many studies, these models have become extremely popular tools among data analysts in a variety of fields.

Because of this increased popularity, recent years have seen a proliferation of software procedures capable of fitting these types of models. West, Welch, and Galecki (2006) provided a brief history of some of these developments. In the past year alone, new versions of software procedures for fitting LMMs have been released in the SAS, SPSS, R, S-Plus, Mplus, Stata, MLwiN, and HLM software packages, incorporating faster algorithms for fitting the models and options for accommodating some of the many components that these models may have (e.g., subgroup-specific variance-covariance structures for errors). With this article, we aim to provide readers with a current overview of the software procedures that are available for fitting these models, including comparisons of their features and their abilities to accommodate the many analytic aspects that accompany such a broad class of models.

We do not aim to provide a detailed overview of the theory or many possible applications of LMMs in this article. Readers interested in the theory underlying these models can turn to comprehensive texts by Demidenko (2004); Diggle et al. (2002); Fitzmaurice et al. (2004); Goldstein (2010); Jiang (2010); McCulloch, Searle, and Neuhaus (2008); Searle, Casella, and McCulloch (1992); or Verbeke and Molenberghs (2000). Readers more interested in applications of these models in practice using statistical software can turn to Brown and Prescott (2006); Faraway (2005); Gelman and Hill (2006); Hox (2010); Littell et al. (2006); Pinheiro and Bates (2009); Rabe-Hesketh and Skrondal (2008); Raudenbush and Bryk (2002); Singer and Willett (2003); Twisk (2006); Verbeke and Molenberghs (1997); West (2009); or West, Welch, and Galecki (2006). We do assume that readers have basic familiarity with essential LMM concepts, including alternative model specification, distributional assumptions, alternative estimation techniques, conceptual differences between fixed and random effects, and alternative covariance structures.

With this article, we aim to identify capabilities of currently available software procedures for fitting LMMs that are both shared in common and also less commonly implemented across the procedures. We do not aim to declare certain software procedures as “better” than others, but only to present up-to-date, user-friendly comparisons of the procedures based on features that we feel are important for fitting these models. Section 2 compares currently available software procedures in terms of these features, focusing on various analytic aspects that will be

Brady T. West is Assistant Research Professor, Survey Methodology Program, Survey Research Center, Institute for Social Research, and Assistant Research Scientist, Center for Statistical Consultation and Research, University of Michigan, Ann Arbor, MI 48109 (E-mail: [bwest@umich.edu](mailto:bwest@umich.edu)). Andrzej T. Galecki is Research Professor, Institute of Gerontology, Medical School, and Research Scientist, Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109 (E-mail: [agalecki@umich.edu](mailto:agalecki@umich.edu)). The authors wish to gratefully acknowledge the helpful comments from the Editor, an Associate Editor, and two anonymous referees, along with the detailed fact-checking provided by representatives from SAS, SPSS, Stata, SYSTAT, Mplus, and MLwiN. Any inaccuracies with regard to software procedures from the other packages are not intentional and may be inevitable given dynamic progress in this area; nevertheless, inaccuracies should be attributed to the authors. We also gratefully acknowledge support from the Claude Pepper Center Grants AG08808 and AG024824 from the National Institute of Aging.

of interest to statisticians working with these models. Section 3 summarizes our comparisons with a concluding discussion.

## 2. COMPARISONS OF SOFTWARE PROCEDURES

In this section, we introduce important analytic aspects of fitting LMMs that statisticians will need to consider carefully when selecting a software procedure for their analysis. In general, modern software procedures currently available for fitting LMMs share many common analytic capabilities, and we discuss these briefly below. Our focus in this review is on key analytic aspects of fitting LMMs that are not yet widely implemented across the available software procedures, and we present comparisons of the available software<sup>1</sup> in terms of current implementation of these aspects in Tables 1 through 5. More details on these different aspects can be found in West, Welch, and Galecki (2006).

*Aspect 1: Data Organization.* Most current software procedures (e.g., PROC MIXED in SAS/STAT) require a clustered or longitudinal dataset to be in a “long” or “vertical” format, with multiple rows per cluster (or subject in a longitudinal study). Nearly all of the statistical software packages mentioned in this review will accept as input raw text files in a delimited format (where rows represent cases, columns represent variables, and data for different variables are separated by a character, such as a comma or a blank space), or data files in a format native to a specific package (e.g., .dta files in Stata). However, input data files in a “wide” format (which is common in longitudinal studies, where different variables contain repeated measures) will need to be restructured to the vertical format required by procedures that fit LMMs. Most of the “general-purpose” packages (e.g., SPSS, Stata, SYSTAT) will provide users with data management tools allowing for easy restructuring of “wide” or “horizontal” datasets into the necessary vertical structure for an LMM analysis (e.g., the VARSTOCASES command in SPSS, the reshape command in Stata, or the WRAP command in SYSTAT). Users of software procedures specifically dedicated to LMM analyses (e.g., HLM) will need to perform this data management in another package prior to analysis. Some software procedures (e.g., HLM) require the user to input multiple datasets in the “long” form, with each dataset corresponding to a specific level of a hierarchical study design (e.g., a student-level dataset with both a school ID and student ID, and a school-level dataset).

*Aspect 2: Model Specification.* Current software procedures for fitting LMMs offer a variety of ways for users to specify the different components in a model of interest. Some software procedures (e.g., HLM, MLwiN) offer users a graphical interface

for specifying models in a multilevel (or hierarchical) 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” for the desired model. In general, in a given software package, the specification of design matrices for fixed and random effects is similar to specifications used for classical linear models. However, different software procedures have different syntax or different options allowing users to specify alternative covariance structures in these models, and we aim to provide some clarity on these differences below. Table 1 compares currently available software procedures in terms of different model specification aspects.

Statisticians working with subject matter where heterogeneous error covariance structures are expected between different groups of analytic units will be attracted to procedures that are capable of allowing heterogeneity in the error covariance matrix across different groups (e.g., PROC MIXED in SAS, GENLINMIXED in SPSS Statistics, xtmixed in Stata, glamm in Stata, hglm() in R, etc.). Table 1 shows that most software procedures are capable of fitting models with crossed random effects, but important implementation holes still exist with regard to models allowing for different variance-covariance structures for different groups in a given dataset, which is a very important aspect of fitting LMMs. We also see in Table 1 that the vast majority of the available software procedures require users to specify models using syntax, where the syntax defines a “single equation” form of the model. Only the HLM and MLwiN packages offer users the convenience of specifying models graphically in a multilevel structure, using point-and-click interfaces. Users of syntax-driven procedures need to take extra care in making sure that the model of interest has been correctly specified.

*Aspect 3: Estimation Methods and Likelihood Optimization.* Most current software procedures capable of fitting LMMs provide users with a choice of residual maximum likelihood (REML) or maximum likelihood (ML) estimation, depending on the hypothesis test of interest. 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 also have default methods for optimizing likelihood functions, and various combinations of these methods may be used 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 not yet widely programmed into LMM software procedures. Table 2 compares the currently available software procedures in terms of these estimation aspects.

Occasionally, analysts wish to fit LMMs to survey datasets collected from samples with complex designs (see Heeringa et al. 2010, chap. 12). Complex sample designs are generally characterized by division of the population into strata, multistage selection of clusters of individuals from within the strata, and unequal probabilities of selection for both clusters and the ultimate individuals sampled. These unequal probabilities of se-

<sup>1</sup> We were unable to fully review the stand-alone ASREML software (Version 3; Gilmour et al. 2009) and the more general-purpose GENSTAT software (Version 14) in this article, due to cost restrictions. Both procedures use the same computational engine for fitting mixed-effects models. GENSTAT is a more general-purpose statistical computing package that enables users to fit mixed-effects models (and perform other analyses) while also managing data. ASREML is used strictly for fitting mixed-effects models to large datasets using REML and efficient computational methods. Visit <http://www.vsni.co.uk/software/asreml/> for more details.

Table 1. Comparisons of model specification aspects

Software procedure (version)	Graphical multilevel specification	Crossed random effects	Multiple outcomes	Choices for covariance matrix of random effects ( <i>D</i> )	Group-specific <i>D</i> matrices	Group-specific covariance matrices for model errors ( <i>R</i> )
HLM (7)	✓	✓	✓			
MIXREG* (1.2)						
MLwiN (2.22)	✓	✓	✓		✓	✓
Mplus (6.1)		**	✓	✓	✓	✓ <sup>1</sup>
R (2.12.2): lme()				✓		✓
R (2.12.2): lmer()		✓		✓		
R (2.12.2): hglm()		2		✓		✓
SAS (9.2): PROC MIXED		✓	3	✓	✓	✓
SAS (9.2): PROC GLIMMIX		✓	✓	✓	✓	✓
SAS (9.2): PROC HP MIXED		✓			✓	
SPSS (20): MIXED / GENLIN MIXED		✓	✓	✓	✓	✓
Stata (12): gllamm		4	✓			✓
Stata (12): xtmixed		✓		✓	5	✓
Statistica (10)		✓				
SYSTAT (13)		✓	✓	✓		✓
WinBUGS (1.4.3)		✓			✓	✓

<sup>1</sup> Mplus provides group-specific covariance matrices by multiple group analysis via the KNOWNCLASS option and TYPE=MIXTURE. When all classes are known, this is identical to multiple group analysis. Heterogeneity in Level-1 variances can also be accommodated using multilevel regression mixture models; see Muthén and Asparouhov (2009a, 2009b) for additional details.

<sup>2</sup> The HGLMfit() function in the R package HGLMMM can be used to fit models with crossed random effects using the same estimation methods as hglm(). See Ronnegard et al. (2011) for more details on the hglm package in R, which offers the flexibility of fitting generalized linear mixed models with different families of distributions for the random effects (aside from Gaussian).

<sup>3</sup> See Galecki (1994) for computational details.

<sup>4</sup> In general, software like the gllamm procedure that has been designed for fitting models with nested random effects only can be “tricked” into fitting models with crossed random effects: see <http://www.stata.com/statalist/archive/2004-04/msg00762.html> for details. We advise analysts to use procedures like xtmixed that have been programmed to handle crossed random effects appropriately so as to minimize possible programming errors.

<sup>5</sup> From Stata: The -xtmixed- command can accommodate group-specific *D* (random-effects) matrices through repetitive independent-panel specifications. As a simpler example, consider a two-level random-coefficient model for outcome variable -y- with a random coefficient for -z-. Suppose we want the random-effects covariance structure, the covariance matrix of random intercept and random coefficient for -z-, to be different between various groups, say boys and girls. This can be done as follows:

```
. xtmixed y z <...> || id: boy boyXz, nocons cov(un) || id: girl girlXz, nocons cov(un)
```

where the -id- variable is the second-level identifier; -boy- and -girl- are dummy variables for boys and girls, respectively; and -boyXz- and -girlXz- contain gender-specific interactions with covariate -z-.

\* Hedeker and Gibbons (1996).

\*\* Crossed random effects will be available in Version 7 of Mplus.

lection generally lead to the construction of sampling weights for individuals, which ensure unbiased estimation of descriptive parameters when incorporated into an analysis. These weights might be further adjusted for survey nonresponse and calibrated to known population totals. Traditionally, analysts might consider a *design-based* approach to incorporating these complex sampling features when estimating regression models (Heeringa et al. 2010). More recently, statisticians have started to explore *model-based* approaches to analyzing these data, using LMMs to incorporate fixed effects of sampling strata and random effects of sampled clusters.

The primary difficulty with the development of model-based approaches to analyzing these data has been choosing appropri-

ate methods for incorporating the sampling weights (see Gelman 2007 for a summary of the issues). 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 survey weights, and Rabe-Hesketh and Skrondal (2006), Carle (2009), and Heeringa et al. (2010, chap. 12) have presented applications using current software procedures, but this continues to be an active area of statistical research. Software procedures capable of fitting LMMs are at various stages of implementing the approaches that have been proposed in the literature thus far for incorporating complex design features, and analysts need to consider this when fitting LMMs to complex sample survey data. Ana-



Table 2. Comparisons of estimation and likelihood optimization aspects

Software procedure (version)	Alternative algorithms to initiate parameter estimation	Correct handling of survey weights	Constraints on variance- covariance parameters	Sparse matrices <sup>1</sup> for computational speed
HLM (7)		✓	✓	✓
MIXREG (1.2)	✓		✓	
MLwiN (2.22)		✓	✓	✓
Mplus (6.1)	✓	✓	✓	
R (2.12.2): lme()	✓		2	
R (2.12.2): lmer()			✓	✓
R (2.12.2): hglm()				
SAS (9.2): PROC MIXED	✓		✓	
SAS (9.2): PROC GLIMMIX	✓		3	
SAS (9.2): PROC HP MIXED			✓	✓
SPSS (20): MIXED / GENLIN MIXED	✓		✓	✓
Stata (12): gllamm	✓	✓	✓	
Stata (12): xtmixed	✓	✓	✓	
Statistica (10)	✓		✓	
SYSTAT (13)	✓		✓	
WinBUGS (1.4.3)		4	✓	

<sup>1</sup> See Section 5 in the paper <http://cran.r-project.org/web/packages/lme4/vignettes/Theory.pdf> for a primer.

<sup>2</sup> Log-Cholesky decomposition to assure variance-covariance matrices remain positive definite.

<sup>3</sup> Modeling the variance with structures based on the Cholesky root, as well as constructing and solving the mixed model equation using the Cholesky root of the covariance matrix of random effects.

<sup>4</sup> In purely model-based methods, as implemented via the MCMC-based Gibbs Sampler in the WinBUGS software, inclusion of auxiliary variables used to define survey weights in LMMs is generally recommended; see Gelman (2007) for a discussion of the issues with this approach.

lysts interested in fitting LMMs to data collected from complex sample surveys will be attracted to procedures that are capable of correctly incorporating the survey weights into the estimation procedures (HLM, MLwiN, Mplus, xtmixed, and gllamm), consistent with the present literature in this area.

Presently, only selected software procedures are relying on estimation methods using sparse matrices, including ASREML, HLM, MLwiN, the lmer() function in R (Bates et al. 2011), GENLIN MIXED in SPSS Statistics, and PROC HP MIXED in SAS. These methods greatly improve computational speed, especially for larger datasets. Slightly more procedures are using alternative methods to invoke parameter estimation, which can result in more stability of parameter estimates (see West et al. 2006, sec. 2.5). Given modern computing power, these alternative aspects will only truly impact models being fitted to very large datasets. Finally, most software procedures have built-in constraints on covariance parameters during estimation routines, preventing negative estimates of variance components. Upon encountering negative variance estimates, some estimation routines working under specific constraints will simply display estimates of 0 for the variance components (e.g., PROC MIXED in SAS; this constraint can be relaxed by using the NOBOUND option on the PROC MIXED line). The final estimates of variance components displayed by the software procedures (and corresponding “notes” or warning messages generated by executing the procedures) may vary depending on the constraints being used.

*Aspect 4: Hypothesis Testing.* Most software procedures will automatically compute  $F$ - and  $t$ -test statistics, approximate degrees of freedom for the test statistics, and corresponding  $p$ -values for fixed effects (to test null hypotheses that the fixed effects are equal to zero). Procedures will vary in terms of the approximations used (or available for use) for the degrees of freedom. For instance, Satterthwaite and Kenward-Roger DF approximations (Kenward and Roger 1997) are computed by PROC MIXED and PROC GLIMMIX in SAS. Some software procedures (e.g., the mcmcsmpl() function in the contributed R package lme4) enable the user to apply Bayesian methods (i.e., generating draws from posterior distributions based on a fitted model) post-estimation for making inferences. Bayesian approaches to making inferences for fixed effects avoid the complication of selecting a degrees of freedom approximation for the test statistics.

When 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 errors is equal to zero), performing asymptotic likelihood ratio tests for the parameters is straightforward using modern statistical software. Computational difficulties arise when testing whether variance components are equal to 0 (which places the variance on the boundary of its parameter space). Some software procedures will automatically present some form of an asymptotic Wald test (e.g., the estimated variance component divided by its estimated standard error) for covariance parameters, which has

Table 3. Comparisons of hypothesis testing aspects

Software procedure (version)	Alternative methods for computing approximate DF	Bayesian inference methods	Appropriate hypothesis tests for variance components
HLM (7)			✓
MIXREG (1.2)			
MLwiN (2.22)		✓	✓
Mplus (6.1)		✓	
R (2.12.2): lme()			2
R (2.12.2): lmer()		1	2
R (2.12.2): hglm()			
SAS (9.2): PROC MIXED	✓	✓	
SAS (9.2): PROC GLIMMIX	✓		3
SAS (9.2): PROC HP MIXED			
SPSS (20): MIXED/GENLINMIXED	✓		
Stata (12): gllamm			
Stata (12): xtmixed			4
Statistica (10)			
SYSTAT (13)			
WinBUGS (1.4.3)		✓	

<sup>1</sup> With `mcmcscamp()` in the `lme4` package, or `pvals.func()` in the R package.

<sup>2</sup> With `exactLRT()` in the `RLRsim` package.

<sup>3</sup> A mixture of central chi-square distributions is used to compute the  $p$ -values for the likelihood ratio test for several recognized special cases.

<sup>4</sup> For tests of single variance components.

several shortcomings (see Berkhof and Snijders 2001). Other procedures might report confidence intervals for covariance parameters (see Bottai and Orsini 2004 for a Stata procedure). The HLM software uses a chi-square test for variance components explained by Raudenbush and Bryk (2002, p. 63–64). From a frequentist perspective, the current literature advocates the reference of likelihood ratio test statistics for variance components to distributions defined (asymptotically) by mixtures of chi-square distributions [see Verbeke and Molenberghs (2000) and Zhang and Lin (2008) for nice discussions of this issue].

The exact null distribution of the likelihood ratio test statistic for a *single* variance component under more general conditions (including small samples) has been defined (Crainiceanu and Ruppert 2004), and Fabian Scheipl has written software in R implementing exact likelihood ratio tests based on simulations from this distribution (the `exactLRT()` function in the `RLRsim` package<sup>2</sup>). 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. Bayesian 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 `pvals.func()` function from the `languageR` package in R) are not as common. Table 3 compares the currently available software procedures in terms of hypothesis testing capabilities.

Table 3 indicates that appropriate inferential methods for the variance components associated with random effects have (to date) not been widely implemented. While post-estimation likelihood ratio tests for variance components are fairly easy to program post-estimation (see West et al. 2006), the software procedures currently available rarely automate this kind of test. SAS PROC GLIMMIX recognizes several special cases and uses a mixture of central chi-square distributions to compute  $p$ -values for likelihood ratio tests. The `xtmixed` command in the Stata software currently automates this test correctly for null hypotheses that *single* variance components are equal to zero; when multiple random effects are included in a model, Stata uses a classical likelihood ratio test, and reminds users that results of the test should be considered conservative.

Bayesian methods for making inferences about variance components are also becoming more widely advocated, but these methods have yet to be widely implemented outside of specialized post-estimation routines available in selected software packages [e.g., the `mcmcscamp()` function in the R software, or available Markov chain Monte Carlo (MCMC) procedures for inference in MLwiN]. We also note that the HLM software implements a relatively unique but fairly straightforward

<sup>2</sup><http://cran.r-project.org/web/packages/RLRsim/RLRsim.pdf>

Table 4. Comparisons in terms of model diagnostic aspects

Software procedure (version)	Influence statistics	Diagnostic plots
HLM (7)		✓
MIXREG (1.2)		
MLwiN (2.22)	✓	✓
Mplus (6.1)	✓	✓
R (2.12.2): lme()		
R (2.12.2): lmer()		
R (2.12.2): hglm()	✓	✓
SAS (9.2): PROC MIXED	✓	✓
SAS (9.2): PROC GLIMMIX		✓
SAS (9.2): PROC HP MIXED		
SPSS (20): MIXED / GENLIN MIXED		
Stata (12): gllamm	✓	
Stata (12): xtmixed		
Statistica (10)		
SYSTAT (13)		✓
WinBUGS(1.4.3)		

automated test of null hypotheses for single variance components, which only relies on the ability to compute ordinary least squares (OLS) estimates of the random coefficients associated with each cluster or subject (in a longitudinal study). For more information on this chi-square test, see Raudenbush and Bryk (2002, p. 63–64).

*Aspect 5: Model Diagnostics.* Assessment of model diagnostics is more complicated in the case of the LMM as opposed to a simple linear regression model, given that there are more underlying assumptions (see Fung et al. 2002; Zewotir and Galpin 2005). Most LMM software can readily compute empirical best linear unbiased predictors (EBLUPs) for random effects and model-based residuals for assessing simple model diagnostics. Recent work by Schabenberger (2004) has allowed SAS users to examine influence diagnostics when using PROC MIXED, but diagnostic procedures still tend to be fairly limited in most software procedures. See the case studies in West et al. (2006) for examples of simpler model diagnostics that can be implemented in most current software packages, including plots of residuals and plots of EBLUPs. The ability of a given software procedure to generate important diagnostic statistics for LMMs is an important consideration for model assessment. Table 4 compares the currently available software procedures in terms of model diagnostic capabilities.

Table 4 indicates that tools for model diagnostics are also currently lacking in the available LMM software. Thorough assessment of model diagnostics for LMMs is essential for making sure that inferences are not being heavily influenced by unusual observations or unusual clusters (or subjects, in the case of a longitudinal study), and there is clearly room for software development in this area. See West et al. (2006, chap. 3) for exam-

ples of the diagnostic capabilities within SAS PROC MIXED.

*Aspect 6: Statistical Output.* LMM analyses can produce a great deal of output, and alternative software procedures vary in terms of the amount of output produced by default. Basic output shared in common by the available software procedures includes parameter estimates, estimates of asymptotic standard errors for the parameter estimates, simple test statistics, and information criteria based on the maximized likelihood function. However, many alternative tests and model diagnostics are possible, and software procedures vary in terms of this default output. An important consideration for statisticians is what output is produced by default, and what output needs to be specially requested. Table 5 compares the currently available software procedures in terms of selected aspects of statistical output.

Table 5 suggests that additional software developments enabling more advanced visualization of model fit (through automated plotting of estimated marginal means and measures of uncertainty associated with these estimates) and adjustments for post-hoc multiple comparisons of means between different levels of categorical fixed factors are needed in general. Computing marginal fitted values and their standard errors is fairly straightforward in most software procedures (e.g., using the /EMMEANS subcommand in SPSS MIXED), enabling users to make their own custom plots if automated procedures are not available.

The display of estimated marginal variance-covariance matrices for blocks of observations on the dependent variable in a clustered or longitudinal dataset can be a useful diagnostic tool (West et al. 2006, sec. 6.11), but this option is not widely available in current LMM software. Users can also compute the marginal variance-covariance matrix implied by a fitted model using any matrix computing language (see West et al. 2006,

Table 5. Comparisons in terms of selected statistical output

Software procedure (version)	Estimates of marginal variance-covariance matrices	Adjustments for multiple comparisons	Graphs of marginal fitted values
HLM (7)			✓
MIXREG (1.2)			
MLwiN (2.22)			✓
Mplus (6.1)	✓		✓
R (2.12.2): lme()	✓	1	
R (2.12.2): lmer()		1	
R (2.12.2): hglm()			
SAS (9.2): PROC MIXED	✓	✓	
SAS (9.2): PROC GLIMMIX	✓	✓	
SAS (9.2): PROC HP MIXED			
SPSS (20): MIXED / GENLIN MIXED		✓	
Stata (12): gllamm			
Stata (12): xtmixed	2	3	✓
Statistica (10)	✓		✓
SYSTAT (13)		✓	✓
WinBUGS (1.4.3)			

<sup>1</sup> Possible when using the multcomp package for lme or lmer objects.

<sup>2</sup> Estimates of marginal variance-covariance matrices are available automatically when using the unofficial user-written post-estimation command -xtmixed\_corr- after -xtmixed-. Although this is not an official part of Stata 12, the -xtmixed\_corr- command is written by one of StataCorp's developers and is available for free download from within Stata.

<sup>3</sup> When using the post-estimation commands pwcompare or contrast.

chap. 6), but automating this computation would be a nice feature to include in the available software.

### 3. DISCUSSION

The objective of this review article was to provide an overview of the more advanced features of current software procedures capable of fitting LMMs, and to compare their relative strengths when considering all of the analytic complexities that can arise when fitting these models. We did not aim to highlight any one procedure as being the “best” procedure, and we hope that the information in this article will provide practicing statisticians with a useful guide for selecting the best procedure for a given analysis. We have attempted to provide a concise summary of current similarities and differences between the many different software procedures capable of fitting LMMs in one article; the Centre for Multilevel Modeling at the University of Bristol provides a well-designed web site (<http://www.bristol.ac.uk/cmm/>) with detailed reviews of many different software packages including procedures for fitting LMMs.

In our comparisons, we emphasize the less common components that have *already been programmed* for the use of analysts and are essentially “automatic,” rather than less common components that would require additional programming on the part of the user. For example, generation of standard diagnostic plots for LMMs is fairly straightforward when using saved predicted

values and residuals from the various procedures. Not all procedures have programmed options for automatically generating diagnostic plots, and we highlight procedures with this capability in Table 4. Evidence of a missing feature for a particular software procedure is not entirely detrimental, as our focus here is on more specialized features that we believe to be important for analysts and reflect state-of-the-art developments in the literature. As indicated in Section 2, all of the analysis aspects of fitting LMMs discussed in this article can be programmed with enough effort; we focus on components that have already been implemented and are readily available for analysts.

We believe that the identified strengths of the currently available software procedures should be used by statisticians to select procedures that provide the best fit for a given line of applied work. Of course, “blank” cells in our comparison tables do not indicate that certain analytical procedures are impossible in a given software package; ultimately, all of these procedures can be programmed (with enough effort) in any statistical software environment. Although some of our “availability” indicators in Tables 1–5 may soon be incorrect given rapid changes in the development and capabilities of modern statistical software procedures, we feel that our review provides a fairly clear indication of where user-friendly developments are needed in the software procedures that are currently available for fitting LMMs. We will attempt to provide updates on availability of these features at the web site <http://www.umich.edu/~bwest/almmussp.html>.

One analysis dimension that we have not focused on from the user point-of-view is whether these models can be speci-



fied using point-and-click graphical user interfaces (GUIs), or whether package-specific syntax is needed to fit the models. For example, fitting these models in SAS, R, glamm (in Stata), and Mplus requires users to write syntax for specifying the models and selecting desired options. Users of SPSS, xtmixed (in Stata), HLM, MLwiN, SYSTAT, Statistica, MIXREG, and GENSTAT may prefer the convenience of point-and-click interfaces for setting up the models, although the trade-offs between available features and ease of use should be carefully considered depending on the analysis problem.

Another important consideration in the selection of a software procedure is cost. Some of the software that has been reviewed in this article is available online and can be downloaded free of charge (R, WinBUGS, and MIXREG). We recommend that statisticians working on tight budgets first fit LMMs using whatever commercial software is available for a given project, and then apply more specialized procedures in freely available software (e.g., exactLRT() in R) for more specific aspects of the analysis, such as hypothesis testing. In general, we hope that this article will spur developers of statistical software to focus on continuous implementation of new published methodologies for LMM analysis, so that disparities between software procedures in terms of available capabilities will slowly cease to exist. Many user communities are forced to use a particular statistical software package (due to budgetary constraints, historical usage, etc.), and the need to use other statistical software for specific aspects of LMM fitting can be an inconvenience for inexperienced and experienced users alike.

Finally, this review article does not address software tools providing power analysis capabilities for longitudinal or clustered study designs, software procedures for fitting generalized linear mixed models (GLMMs) to nonnormal outcomes in longitudinal or clustered datasets, or nonlinear mixed models (NLMMs). We find the Optimal Design software (<http://sitemaker.umich.edu/group-based/optimal-design-software>) and the MLPowSim software (<http://www.bristol.ac.uk/cmm/learning/multilevel-models/samples.html#mlpowsim>; see Browne et al. 2009) to be flexible (and free) software tools for power analysis in clustered or longitudinal study designs. Gelman and Hill (2006) also provide an extremely useful reference on simulation-based methods for examining power in multilevel studies. Finally, we hope to publish a similar article addressing these topics for GLMMs and NLMMs in the near future. We note that other researchers have already begun similar comparisons of software for fitting GLMMs (see <http://glmm.wikidot.com/pkg-comparison>), and we hope to contribute further to this work in the near future.

[Received May 2011. Revised December 2011.]

## REFERENCES

- Asparouhov, T., and Muthén, B. (2006), "Multilevel Modeling of Complex Survey Data," *Proceedings of the Joint Statistical Meetings*, August 2006, ASA Section on Survey Research Methods, Alexandria, VA: ASA, pp. 2718–2726. 276
- Bates, D., Maechler, M., and Bolker, B. (2011), "lme4: Linear Mixed-Effects Models using S4 Classes," R package version 0.999375-42. 277
- Berkhof, J., and Snijders, T.A.B. (2001), "Variance Component Testing in Multilevel Models," *Journal of Educational and Behavioral Statistics*, 26, 133–152. 278
- Bottai, M., and Orsini, N. (2004), "A New Stata Command for Estimating Confidence Intervals for the Variance Components of Random-Effects Linear Models," presented at United Kingdom Stata Users' Group Meetings, Stata Users Group, 2004. 278
- Brown, H., and Prescott, R. (2006), *Applied Mixed Models in Medicine* (2nd Ed.), New York: Wiley. 274
- Browne, W.J., Lahi, M.G., and Parker, R.M.A. (2009), *A Guide to Sample Size Calculations for Random Effect Models via Simulation and the MLPowSim Software Package*, available online at <http://seis.bris.ac.uk/~fwj/b/esrc/MLPOWSIMmanual.pdf>. 281
- Carle, A.C. (2009), "Fitting Multilevel Models in Complex Survey Data with Design Weights: Recommendations," *BMC Medical Research Methodology*, available online at <http://www.biomedcentral.com/1471-2288/9/49>. 276
- Crainiceanu, C. M., and Ruppert, D. (2004), "Likelihood Ratio Tests in Linear Mixed Models With One Variance Component," *Journal of the Royal Statistical Society, Series B*, 66, 165–185. 278
- Demidenko, E. (2004), *Mixed Models: Theory and Applications*, London: Wiley-Interscience. 274
- Diggle, P., Heagerty, P., Liang, K., and Zeger, S. (2002), *Analysis of Longitudinal Data* (2nd Ed.), New York: Oxford University Press. 274
- Faraway, J.J. (2005), *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, New York: Chapman and Hall/CRC Press. 274
- Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2004), *Applied Longitudinal Data Analysis*, Hoboken, NJ: Wiley. 274
- Fung, W.-K., Zhu, Z.-Y., Wei, B.-C., and He, X. (2002), "Influence Diagnostics and Outlier Tests for Semiparametric Mixed Models," *Journal of the Royal Statistical Society, Series B*, 64, 565–579. 279
- Galecki, A.T. (1994), "General Class of Covariance Structures for Two or More Repeated Factors in Longitudinal Data Analysis," *Communications in Statistics: Theory and Methods*, 23, 3105. 276
- Gelman, A. (2007), "Struggles With Survey Weighting and Regression Modeling," *Statistical Science*, 22, 153–164. 276, 277
- Gelman, A., and Hill, J. (2006), *Data Analysis using Regression and Multilevel/Hierarchical Models*, New York: Cambridge University Press. 274, 281
- Gilmour, A.R., Gogel, B.J., Cullis, B.R., and Thompson, R. (2009), *ASReml User Guide, Release 3.0*, VSN International Ltd., UK. <http://www.vsn.co.uk>. 275
- Goldstein, H. (2010), *Multilevel Statistical Models* (4th Ed.), New York: Wiley. 274
- Hedeker, D., and Gibbons, R.D. (1996), "MIXREG: A Computer Program for Mixed-Effects Regression Analysis with Autocorrelated Errors," *Computer Methods and Programs in Biomedicine*, 49, 229–252. 276
- Heeringa, S.G., West, B.T., and Berglund, P.A. (2010), *Applied Survey Data Analysis*, Boca Raton, FL: Chapman and Hall/CRC Press. 275, 276
- Hox, J. (2010), *Multilevel Analysis: Techniques and Applications* (2nd ed.), New York: Routledge Academic. 274
- Jiang, J. (2010), *Linear and Generalized Linear Mixed Models and Their Applications*, New York: Springer. 274
- Kenward, M. G., and Roger, J.H. (1997), "Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood," *Biometrics*, 53, 983–997. 277
- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., and Schabenberger, O. (2006), *SAS for Mixed Models* (2nd ed.), Cary, NC: SAS Publishing. 274
- McCulloch, C.E., Searle, S.R., and Neuhaus, J.M. (2008), *Generalized, Linear, and Mixed Models* (2nd ed.), New York: Wiley. 274
- Muthén, B., and Asparouhov, T. (2009a), "Growth Mixture Modeling: Analysis with non-Gaussian Random Effects," in *Longitudinal Data Analysis*, eds. G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs, Boca Raton, FL: CRC Press, pp. 143–165. 276
- (2009b), "Multilevel Regression Mixture Analysis," *Journal of the Royal Statistical Society, Series A*, 172, 639–657. 276
- Pfeiffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H., and Rasbash, J. (1998), "Weighting for Unequal Selection Probabilities in Multilevel Models," *Journal of the Royal Statistical Society, Series B*, 60, 23–40. 276
- Pinheiro, J., and Bates, D. (2009), *Mixed-Effects Models in S and S-PLUS* (2nd

- ed.), New York: Springer. 274
- Rabe-Hesketh, S., and Skrondal, A. (2006), "Multilevel Modeling of Complex Survey Data," *Journal of the Royal Statistical Society, Series A*, 169, 805–827. 276
- Rabe-Hesketh, S., and Skrondal, A. (2008), *Multilevel and Longitudinal Modeling using Stata* (2nd ed.), College Station, TX: Stata Press. 274
- Raudenbush, S.W., and Bryk, A.S. (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.) Newbury Park, CA: Sage. 274, 278, 279
- Ronnegard, L., Shen, X., and Alam, M. (2011), "The hglm Package," a Vignette downloaded from <http://cran.r-project.org/web/packages/hglm/vignettes/hglm.pdf>. 276
- Schabenberger, O. (2004), "Mixed Model Influence Diagnostics," in *Proceedings of the Twenty-Ninth Annual SAS Users Group International Conference*, Paper 189-29, Cary, NC: SAS Institute. 279
- Searle, S.R., Casella, G., and McCulloch, C.E. (1992), *Variance Components*, New York: Wiley. 274
- Singer, J.D., and Willett, J.B. (2003), *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, New York: Oxford University Press. 274
- Twisk, J.W.R. (2006), *Applied Multilevel Analysis: A Practical Guide for Medical Researchers*, Cambridge: Cambridge University Press. 274
- Verbeke, G., and Molenberghs, G. (1997), *Linear Mixed Models in Practice: A SAS-Oriented Approach*, New York: Springer. 274
- (2000), *Linear Mixed Models for Longitudinal Data*, Berlin: Springer-Verlag. 274, 278
- West, B.T. (2009), "Analyzing Longitudinal Data with the Linear Mixed Models Procedure in SPSS," *Evaluation and the Health Professions*, 32, 207–228. 274
- West, B.T., Welch, K.B., and Galecki A.T. (with Contributions from B.W. Gillepie) (2006), *Linear Mixed Models: A Practical Guide using Statistical Software*, Boca Raton, FL: Chapman and Hall/CRC Press. 274, 275, 277, 278, 279
- Zewotir, T., and Galpin, J.S. (2005), "Influence Diagnostics for Linear Mixed Models," *Journal of Data Science*, 3, 153–177. 279
- Zhang, D., and Lin, X. (2008), "Variance Component Testing in Generalized Linear Mixed Models for Longitudinal/Clustered Data and other Related Topics," Chapter 2 in *Random Effect and Latent Variable Model Selection*, ed. D. B. Dunson, Berlin: Springer Lecture Notes in Statistics, 192. 278