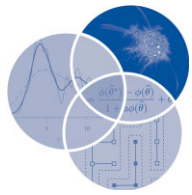
Advanced Review



Fixed and random effects models

Matthew J. Gurka[1]*, George A. Kelley[1] and Lloyd J. Edwards[2]

Traditional linear regression at the level taught in most introductory statistics courses involves the use of 'fixed effects' as predictors of a particular outcome. This treatment of the independent variable is often sufficient. However, as research questions have become more sophisticated, coupled with the rapid advancement in computational abilities, the use of random effects in statistical modeling has become more commonplace. Treating predictors in a model as a random effect allows for more general conclusions—a great example being the treatment of the studies that comprise a meta-analysis as random rather than fixed. In addition, utilization of random effects allows for more accurate representation of data that arise from complicated study designs, such as multilevel and longitudinal studies, which in turn allows for more accurate inference on the fixed effects that tend to be of primary interest. It is important to note the distinctions between fixed and random effects in the most general of settings, while also knowing the benefits and risks to their simultaneous use in specific yet common situations.

How to cite this article:
WIREs Comput Stat 2012, 4:181–190. doi: 10.1002/wics.201

Keywords: mixed models; multilevel studies; meta-analysis; longitudinal data

INTRODUCTION

The use of statistical models is pervasive. In almost every general area of research, models are used to explain relationships among variables or provide tools for prediction. There are numerous classes of models with countless options within each class. In general, however, a model has one (or more) dependent variables, or outcomes, and one (or more) independent variables, or predictors. No matter the type of model, the choice must be made (implicitly or explicitly) whether to treat the predictors as fixed or random effects.

Defining fixed and random effects, and comparing and contrasting the two, has been a focus of much discussion over the decades. There is no consensus mechanism for distinguishing between the two, and often interpretation depends on the context in which they are being used. In general, often the classification of a variable as a fixed or a random effect is driven by the motivation for that variable in the analysis. Variables where the analyst is interested in making statistical comparisons between its levels are typically viewed as fixed effects. As an example, consider a study comparing a new exercise regimen aimed at reducing falls among the elderly, and nursing home residents are randomized to either the new regimen ('intervention') or a standard exercise program ('control'). Primary interest lies in the comparison of fall rates between the residents in each of the two exercise programs—thus, the binary variable 'program' (intervention or control) should be classified as a fixed effect. Indeed, in most settings of this sort, where a simple χ^2 test for a binary outcome or a t-test for a continuous outcome is used to compare two levels of a variable, the 'treatment' variable is implicitly assumed to be a fixed effect. That is, the levels of the fixed effect are the only levels of interest in this study, and there is no interest in generalizing to other levels (in fact, often there are no other levels of the fixed effect).

Random effects, on the other hand, typically are assigned to predictors whose observed levels in the

*Correspondence to: mgurka@hsc.wvu.edu

[1]Department of Community Medicine, School of Medicine, West Virginia University, Morgantown, WV, USA

[2]Department of Biostatistics, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

study in question are a sample drawn from a larger population of levels. Designation of such a variable as a random effect would only be appropriate if interest lied in generalizing to the population of possible levels of that variable. In the nursing home example, if ten nursing homes were chosen to be a part of this study, and interest lied in generalizing to residents of all nursing homes (rather than conclusions aimed at only these ten nursing homes), 'nursing home' should be considered a random effect in any analysis. This topic, particularly in the scope of centers, schools, or 'clusters' of individuals in general, will be revisited in more detail.

To summarize, if one is interested in comparing means between levels of a variable, that variable should be considered a fixed effect in any sort of analysis. Comparisons of means among levels of random effects, on the other hand, are generally not of interest. Rather, interest generally lies in accounting for variance across levels of random effects. Such a distinction does not depend on a continuous outcome; the above example described a binary outcome of interest. Indeed, it could be generalized to state that interest lies in first moment comparisons of fixed effect variable levels, while the second moment is of interest when referring to random effects.

To outline the rest of the article, a general introduction to the 'mixed model' will be discussed first. Such a model allows for the inclusion of fixed and random effects and is becoming a standard tool for the practicing data analyst. The model's accommodation of both types of effects makes it a natural foundation for a discussion on the distinction of fixed and random effects. Next, the study design that utilizes random effects most often will be discussed: the longitudinal study. The discussion will be grounded using a real longitudinal study. Generalizing to multilevel studies with correlated data will follow. Another area of application where the distinction between fixed effects and random effects is important is meta-analysis; this topic will also receive an overview. Other types of analysis that allow for random effects will then be discussed. Finally, conclusions will be made and general areas of future research will be highlighted.

USING BOTH FIXED AND RANDOM EFFECTS: THE 'MIXED' MODEL

While models can include exclusively fixed or random effects, most often models will include both. To frame the discussion moving forward, the concept of the 'mixed model' is introduced here—which allows for both classifications of variables. While introducing the concept of mixed models, it is most straightforward to discuss with reference to linear models.

The Generalized Linear Model (Fixed Effects/Population-Averaged Model)

To begin, the generalized linear model, in the form discussed by McCullagh and Nelder,[1] also referred to as a population-averaged, has the following form:

$$E(\mathbf{y}_i) = f(\mathbf{X}_i\boldsymbol{\beta}). \quad (1)$$

Here, we are fitting a model to data from N independent sampling units (ISU). In model (1), $\boldsymbol{y}_i$ is the $(p \times 1)$ vector of responses from each of the N ISU's ($i = 1, 2, ...N$), $\boldsymbol{X}_i$ is the $(p \times q)$ design matrix of known variables for ISU i, and $\boldsymbol{\beta}$ is a $(q \times 1)$ vector of fixed, unknown parameters. If $p = 1$, the model represents a simple generalized linear model for one observation per ISU (i.e., $E(y_i) = f(\boldsymbol{x}_i'\boldsymbol{\beta})$). Model (1) allows for repeated observations on each ISU. Here, f is a function of the 'systematic component' of the model. The inverse of this function, say g, is typically called the 'link' function. So, $g\{E(\boldsymbol{y}_i)\} = \boldsymbol{X}_i\boldsymbol{\beta}$. There are many common link functions, each usually corresponding to an assumed distribution of $\boldsymbol{y}_i$. The simplest function is $g\{E(\boldsymbol{y}_i)\} = E(\boldsymbol{y}_i)$, the identity link, where $\boldsymbol{y}_i$ is assumed to be normally distributed. This simple case is the general linear model, a specific case of the generalized linear model. While this model is a type of generalized linear model, it has special properties that have made it much more widely studied. For logistic regression, the link function is called the logit link, $g(y_i) = \log\{y_i/(1 - y_i)\}$, where y_i is assumed to follow a Bernoulli distribution. Logistic regression is popular in many epidemiological and other biomedical studies where the outcome has two options, e.g., disease or no disease, and interest lies in estimating the odds of developing the disease. For Poisson regression, the link function is the log link, $g(y_i) = \log(y_i)$, where y_i is assumed to follow a Poisson distribution. Poisson regression is often used to model count or rate data.

The generalized linear model (1) contains only fixed effects—the parameters that comprise $\boldsymbol{\beta}$. In most settings such a model is sufficient. Inference about $\boldsymbol{\beta}$ will allow one to make conclusions regarding mean differences between treatments, demographic characteristics, and most other predictors with respect to the outcome of interest. The one limitation to such inference is that since the levels of the predictors are assumed to be fixed and not random, no generalizations can be made to levels of predictors that are not observed in the data.

Even in the presence of correlated data—such as a longitudinal study where repeated measurements are observed on each individual—model (1) can be appropriate, provided the analyst accommodates the lack of independence among observations in the model. For the general linear model using the identity link, $\boldsymbol{y}_i$ is assumed to exhibit a Gaussian distribution. Specifically, $\boldsymbol{y}_i \sim N(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$, with $\boldsymbol{\Sigma}_i$ denoting the covariance matrix that can take many forms depending on the nature of the dependence of the data. In the case of independent observations, $\boldsymbol{\Sigma}_i = \sigma^2\boldsymbol{I}_p$. The model with a general covariance structure to accommodate correlated data is often referred to as a 'marginal multivariate model'[2] or 'population-averaged model'.[3] There is no need to consider random effects in such a model to account for a lack of independence.

The Generalized Linear Mixed Model (Model with Both Fixed and Random Effects)

The generalized linear mixed model, in the form introduced by McCulloch and Searle,[4] simply adds random effects to the generalized linear model (1):

$$E\left(\boldsymbol{y}_i|\boldsymbol{b}_i\right) = f\left(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i\right), \tag{2}$$

where $\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$ and f is a function of the fixed and random effects of the model. The so-called general linear mixed model simply has an identity link and assumes a Gaussian distribution of $\boldsymbol{y}_i$, conditional on $\boldsymbol{b}_i$. In this setting the analyst can include variables where the assumption is that the levels of those predictors were drawn randomly from a larger population, and inferences on the entire population are of interest. However, the utility of the random effects in the mixed model setting is primarily in its ability to effectively model non-independent data. When multiple observations are collected on each person/family/hospital, independence of observations, at least taken from the same independent sampling unit, can no longer be assumed. The generalized linear mixed model (2), then, with its additional source of variation represented by the random effects ($\boldsymbol{b}_i$) can accommodate such data, which will lead to potentially accurate inference about the fixed effects. In addition, the random effects themselves (and their estimates) can provide meaningful information.

THE MIXED MODEL IN LONGITUDINAL STUDIES

One common application for random effects in a mixed model setting with Gaussian random deviations is when data are collected repeatedly, usually over time, on an individual sampling unit. The common form developed by Laird and Ware[5] for longitudinal data analysis involving a continuous outcome is as follows:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i\,. \tag{3}$$

Here, $i \in \{1, ..., N\}$, where N is the number of independent sampling units (subjects), $\boldsymbol{y}_i$ is an $p \times 1$ vector of observations on the *ith* subject; $\boldsymbol{X}_i$ is an $p \times q$ known, constant design matrix for the *ith* subject with rank q; $\boldsymbol{\beta}$ is a $q \times 1$ vector of unknown, constant population parameters; $\boldsymbol{Z}_i$ is an $p \times k$ known, constant design matrix for the *ith* subject with rank k corresponding to $\boldsymbol{b}_i$, a $k \times 1$ vector of unknown, random individual-specific parameters (the 'random effects'); and $\boldsymbol{e}_i$ is an $p \times 1$ vector of random 'within-subject,' or 'pure,' error terms.

The following distributional assumptions are usually held: $\boldsymbol{b}_i$ is normally distributed with mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{D}$, and $\boldsymbol{b}_i$ is independent of $\boldsymbol{b}_j$, $i \neq j$. Also, $\boldsymbol{e}_i$ is distributed normally with mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{R}_i$, independent of $\boldsymbol{b}_i$. The covariance matrices $\boldsymbol{D}$ and $\boldsymbol{R}_i$ are typically assumed to be characterized by unique parameters contained in the $k \times 1$ vector $\boldsymbol{\theta}$. The total variance for the response vector in (3) is $\text{var}\left(\boldsymbol{y}_i\right) = \boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \boldsymbol{Z}_i\boldsymbol{D}(\boldsymbol{\theta})\boldsymbol{Z}_i' + \boldsymbol{R}_i(\boldsymbol{\theta})$.

The random effects within model (3) are most often viewed as subject-specific level deviations from the population estimates, or fixed effects ($\widehat{\boldsymbol{\beta}}$). Not only does the mixed model (3) allow an analyst to make inferences about the population via the fixed effects, but it also accommodates estimation and inference about subject-specific level deviation from the population estimates of typical interest ($\widehat{\boldsymbol{b}}_i$). The motivation behind the analysis or scientific question of interest will drive the interpretation of the estimates resulting from fitting the model (3) to the data. It is common to only be interested in inference about the fixed effects parameters, $\boldsymbol{\beta}$, and possibly the 'variance components', the variance parameters of $\boldsymbol{\theta}$. In this setting, model (3) with $\boldsymbol{\epsilon}_i = \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$, i.e., $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$, is often referred to as the marginal model[6] or the population-averaged model.[7] The random effects in this context, while not of direct interest, provide the analyst with a way to intuitively model the variation of the data.

In this context, restrictions are often placed on the structure and the number of parameters of both covariance matrices, $\boldsymbol{D}$ and $\boldsymbol{R}_i$. A 'conditionally independent' model is typically assumed; i.e., $\boldsymbol{R}_i = \sigma^2\boldsymbol{I}_{n_i}$. The structure of $\boldsymbol{D}$ is often dictated by the

number of random effects included in the model. For example, in the context of longitudinal data, if one included only a random intercept, then one only needs to estimate the variance of this random intercept term. However, if one also includes a random slope as well, then one must decide whether or not to allow the two random effects to covary. Most software can accommodate many different specified parametric models of both covariance matrices of the mixed model. For more detailed information, see Verbeke and Molenberghs.[6] Harville[8] provided a review of the maximum likelihood (ML) approach to estimation of the parameters in the linear mixed model. The ML estimates of $\boldsymbol{\theta}$ are biased downward since the loss of degrees of freedom resulting from the estimation of the fixed effects is not taken into account. Restricted, or residual, maximum likelihood (REML) estimation acknowledges this loss of degrees of freedom and hence leads to less biased estimates.[9]

Mixed models can be viewed as a result of a two-stage analysis. For example, for a linear mixed effects model (3), in the first stage, an appropriate function of predictors approximates each observed longitudinal profile allowing subject-specific regression coefficients:

$$\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\beta}_i + \boldsymbol{e}_i$$

where $\boldsymbol{\beta}_i$ is a $k \times 1$ vector of unknown subject-specific regression coefficients. In the second stage, another model explains the observed variability between the subjects:

$$\boldsymbol{\beta}_i = \boldsymbol{K}_i\boldsymbol{\beta} + \boldsymbol{b}_i$$

where $\boldsymbol{K}_i$ is a $k \times q$ matrix of known covariates. Note that $\boldsymbol{X}_i = \boldsymbol{Z}_i\boldsymbol{K}_i$ when we combine the two-stage models. When most variability among the measurements reflects between-subject variability, the two-stage approach helps to construct an appropriate mixed model. In turn, when variability depends on predictors that change within-subject, such as time, a valid mixed effects model requires an appropriate covariance structure for the residual in addition to the covariance structure modeled through the random effects. Under the two-stage model formulation, the predictors used in $\boldsymbol{X}_i$ (fixed effects) are often referred to as level 2 variables and the predictors used in $\boldsymbol{Z}_i$ (random effects) are referred to as level 1 variables.

Fixed Effects Inference

Quite often the primary goal in fitting a model with both fixed and random effects is inference on the fixed effects. Scientific questions are often based on means; fixed effects inference allows one to address such questions. Random effects are commonly included in these cases to accommodate the study design, such as the case with longitudinal studies. Random effects themselves, or the variance parameters based on the distributional assumptions of these random effects, may be of secondary or no interest other than to ensure accurate inference about the fixed effects. However, inclusion (or not) of random effects in a mixed model setting does impact the accuracy of fixed effect inference, and has been studied extensively.

Approximate Wald and F statistics are often used to make inferences regarding $\boldsymbol{\beta}$. However, Wald tests can underestimate the true variability in the estimated fixed effects because they do not take into account the variability incurred by estimating $\boldsymbol{\theta}$.[10] Various approximate F tests have been proposed, with major focus on the degrees of freedom for these tests. These degrees of freedom are dependent on the sample size—and in the case of hierarchical data (longitudinal data, clustered data), the effective sample size is the number of independent sampling units (e.g., individuals measured over time, schools with measurements on multiple students). In the analysis of longitudinal data, if the number of subjects is sufficiently large, the degrees of freedom from the various approximations will be similar, leading to similar inference.[6] Unfortunately, the approximate F statistic is known to result in inflated Type I errors and poor power approximations when there are a small number of independent sampling units.[11] Kenward and Roger[12] presented a scaled Wald statistic with an approximate F distribution for testing fixed effects with REML estimation that performs well, even in small samples. However, even this inference technique is not ideal, as documented performance of the Kenward–Roger F statistic for some small sample cases has revealed inflated type I error rates with various covariance model selection techniques.[13] Alternatives have been proposed with no inflation of the type I error rate, but are low in power.[14]

Random Effects Inference

Even though focus typically lies on the fixed effects, it is important to effectively model the variation of the data via the covariance parameters in such a model. Inclusion of random effects and the covariance parameters associated with them allow for an intuitive manner to model the variation observed in the data. Making valid conclusions about the variability of the data is important information in itself, but it also leads to proper inference about the fixed effects as well. The topic of inference about the 'variance components' has been researched extensively.[15]

When one is interested in the random effects themselves in the mixed model, then one needs to make inferences from the hierarchical model perspective. It is most convenient to estimate the random effects using Bayesian techniques, resulting in the following form of the estimates of $\boldsymbol{b}_i$, assuming $\boldsymbol{\theta}$ is known:

$$\widehat{\boldsymbol{b}}_i = \boldsymbol{D}\boldsymbol{Z}_i'\boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right). \quad (4)$$

In most cases $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are not known beforehand. In these settings, $\widehat{\boldsymbol{b}}_i$ in (4) is known as the 'empirical Bayes' estimate of $\boldsymbol{b}_i$, and is based on estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$. Similar to fixed effects inference, inference on the random effects should be based on an approximate *F*-test with an appropriate estimate of the denominator degrees of freedom.[6]

Model Assumptions

The standard mixed model assumes the random effects exhibit a Gaussian distribution with a particular covariance structure. In many specific forms of the generalized linear mixed model, such as the model of continuous outcomes (3); the random effects are assumed to be independent of additional 'residual' error, $\boldsymbol{e}_i$, which also is assumed to follow a Gaussian distribution with a given covariance structure. Estimation and inference about the fixed and random effects is based on these assumptions.

What should the practicing analyst do in the presence of data that do not allow for such assumptions to be valid? In the case of the linear mixed model, which assumes normality of the random effects and the residual error, others have shown that violations of the normality assumption do not greatly impact inference on the fixed effects, but do affect estimation of the random effects.[16] Research has been done on both simple fixes as well as extensions that allow for different distributions. Transformation of the outcome to achieve normality has been utilized extensively for fixed effects models; it has shown to be effective for models with both fixed and random effects as well.[17] Others have extended the traditional linear mixed model that allow more general distributions of the random effects.[18,19] Of course the emergence of the generalized linear mixed model extends the simultaneous modeling of fixed and random effects to more general distributional settings.

Model Selection and Goodness-of-Fit

Selecting the appropriate fixed effects to include in a model is often based on science—either prior knowledge of the background research, or the aims of the study at hand. Standard inferential techniques, most often involving hypothesis tests discussed above, goodness-of-fit measures, and assessing distributional assumptions are used to decide on the final model of fixed effects. Modeling strategies used in more straightforward settings where random effects are also not included, such as backward selection, can be done in these settings as well.[20]

The notion of 'selecting the best model' becomes unique in the mixed model setting when attempting to determine the appropriate random effects to include in the model. One can make inferences using hypothesis tests of variance components.[15] However, candidate models are often not nested, and hence standard inferential techniques are not valid. Information theoretic criteria, such as the Akaike Information Criterion (AIC)[21] and the Bayesian Information Criterion (BIC),[22] are common tools in mixed model selection due to their relative validity in comparing non-nested models. These tools, developed in more general settings, have not been thoroughly studied in the context of mixed models, and hence various forms have been proposed and used with no consensus.[23]

Methods for assessing goodness-of-fit are an active area of research in mixed models. In practice, the coefficient of multiple determination (R^2) is often used to measure the overall goodness of fit of an ordinary multiple linear regression model. The larger the R^2, the better the model fits the data. Edwards et al.[24] developed a model R^2 statistic for the linear mixed model based on an appropriate *F* statistic. The R^2 statistic measures multivariate association between the repeatedly measured outcome and the fixed effects and assesses the goodness of fit of the fixed effects model (the means model).

What are the consequences of model misspecification, particularly misspecification of the random effects model on inference of the fixed effects? Jacqmin-Gadda et al.[25] demonstrated an inflated type I error rate with covariance misspecification in simulations in the general linear mixed model with Gaussian errors. Gurka et al.[26] prove that under-specification of the covariance inflates type I error of fixed effects inference, no matter the sample size. This is particularly troublesome in the context of longitudinal data. Some authors have considered covariance misspecification for tests about the fixed effects in mixed models. Liang and Zeger[27] proposed the 'sandwich' estimator for $\boldsymbol{\mathcal{V}}(\widehat{\boldsymbol{\beta}})$ that aims for inference about the fixed effects that is robust to misspecified covariance. However, use of this sandwich estimator is not ideal in all settings, such as studies with certain types of missingness[6] or in small sample settings.[26]

The Study of Early Child Care and Youth Development

The Study of Early Child Care and Youth Development (SECCYD) was a longitudinal design that followed children from 10 U.S. sites from birth through age 15.[28] Families were recruited at hospital visits following the birth of the child in 1991. A total of 1364 families enrolled from the 8986 eligible births. By phase IV of the study (2005–2008; age 14–15), 1056 families remained enrolled. Numerous outcomes were measured repeatedly throughout the study, namely a variety of standard childhood psychometric measures. Gurka et al.[28] used mixed models to compare children classified as being born 'late-preterm' (34–36 weeks gestational age) to those children born full term (37–40 weeks gestational age) with respect to numerous measures over the span of childhood. Here, a condensed and simplified analysis is presented to demonstrate the utility of a model with both fixed and random effects. Namely, does the average longitudinal profile of externalizing behavior problems for children born late-preterm differ from the profile for those born full term?

This question can be answered by making inferences about fixed effects. The repeated measurements on each child allow for powerful conclusions to be made, but analyzing such data must be done judiciously. The mixed model, with the inclusion of random effects that accommodate these repeated measures, provides one possible strategy. It has advantages over alternative repeated measures models in the fact it allows for incomplete, unbalanced data, and data that are missing (at random).[6]

A linear mixed model was fit to the outcome, the Child Behavior Check List (CBCL) externalizing behavior score.[28] It is a widely used, parent-completed checklist that identifies various types of behavior and emotional problems that occur in children aged 4 and above. The resulting score is age-standardized (mean = 50, SD = 10), with higher scores indicate more externalizing behavior problems. The model below was fit on 1166 children with measures at 4.5, 5, 6, and 8 years of age:

$$y_{ij} = \alpha + \beta_1\,(\mathrm{MALE})_i + \beta_2\,(\mathrm{WHITE})_i + \beta_3\,(\mathrm{MAGE})_i + \beta_4\,(\mathrm{EDU})_i + \beta_5\,(\mathrm{PRE}_i) + \beta_6 t_{ij} + \beta_7\,(\mathrm{PRE}_i)\,t_{ij} + b_{1i} + b_{2i}t_{ij} + e_{ij}\,. \quad (5)$$

Here, y_{ij} is the CBCL score for child i at year of age t_{ij} ($t_{ij} = 4.5, 5, 6, 8$). The following covariates were included in the model as increments to the intercept term: male gender (MALE), white race (WHITE), maternal age (MAGE), and maternal education in years (EDU). PRE_i is an indicator variable for late preterm children. Thus, the above model is fitting a line representing CBCL scores over this period of time, separate for each of the two term groups, while adjusting for possible confounding variables—in other words, separate intercepts and separate slopes for the two groups. Inference on β_5 allows one to conclude whether late-preterm children had higher overall scores than full-term children; inference on β_7 leads to conclusions on whether the linear trajectory over time differs between the two groups. Collectively, model (5) provides mean CBCL trajectories for the two groups via these fixed effects. This model, though, also includes a random intercept (b_{1i}) and random slope (b_{2i}) for each individual i. The variance parameters for each of these two terms, σ^2_{b1} and σ^2_{b2}, respectively, effectively model the among-subject variation with respect to the intercept and slope. Inferences can be made to determine whether not one needs to include these terms, and the two random effects can be allowed to covary, as they are here. The random effect estimates on individual i represent the deviations of each subject's estimated intercept and slope from the population intercept and slope, depending on whether the child was preterm or not. In this instance, we assumed a constant within-unit variation between the two groups.

TABLE 1 | Preterm Infant Study: Mixed Model Results

Model Variable	Parameter Estimate	Standard Error	p-value*
Fixed effects			
Intercept	68.148	1.698	<0.001
Male gender	−1.082	0.507	0.033
White race	0.349	0.645	0.589
Maternal education (years)	−0.430	0.121	<0.001
Maternal age at birth (years)	−0.193	0.055	<0.001
Preterm	−3.375	2.485	0.175
Months	−0.098	0.006	<0.001
Preterm × Months	0.029	0.030	0.339
Random effects variance components			
Intercept variance	68.050	4.138	<0.001
Slope variance	0.011	0.002	<0.001
Covariance (Intercept, Slope)	0.289	0.068	<0.001
Correlation (Intercept, Slope)	0.328		

*For fixed effects: based on *F*-test using Kenward-Roger approximation for the denominator degrees of freedom; for variance components: based on Wald tests.

Table 1 includes the estimates of the parameters in model (5) from SAS PROC MIXED, including the fixed effects as well as the covariance parameters

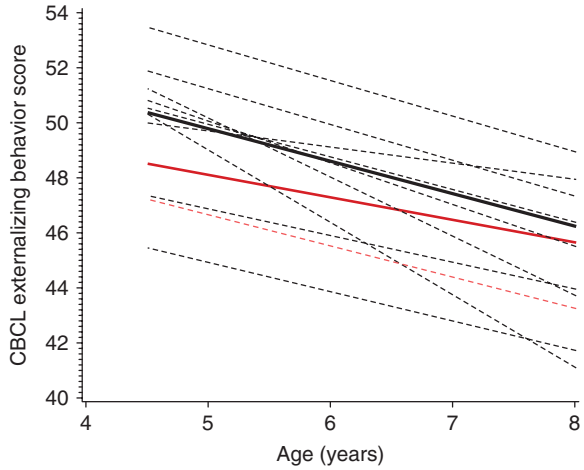


FIGURE 1 | Model-Predicted Mean and Individual Behavior Profiles Over Time. Solid bold lines represent model-predicted mean (fixed effects) profiles for white children whose mothers were 30 years old and had 14 years of education (black = full-term; red = late preterm). Dashed lines indicate individual trajectories as estimated from the random effects included in the model.

associated with the two random effects and the within-unit variance. Inference on the fixed effects leads to the conclusion that no mean differences were observed between the two groups of children. This particular model of the fixed effects allows for two different linear trajectories for the two term groups and the comparisons that naturally follow. Figure 1 displays the model-predicted mean CBCL scores for white males with mothers who were 30 years of age at the time of birth and had 14 years of education, that are based on the fixed effect estimates. Each individual has a corresponding estimate that represents their deviation from the mean trajectory (both intercept and slope). Figure 1 additionally contains a random sampling of individual profiles that are derived from the fixed effects and the random effect estimates (intercept and slope). The random effect estimates themselves (not shown), which represent the subject-specific deviations from the average trend over time for the groups, could be useful in helping to identify underlying individual factors that may influence unusual CBCL trajectories.

The variance parameter estimates represent the among-subject variation with respect to the intercept and slope, and allow for the conclusion that these individuals do vary significantly with respect to overall CBCL value as well as the slope of this value over time. These variance component estimates are also used to derive the random effect estimates, which allows one to conclude there is significant variation among the children with respect to CBCL trajectories over this time period.

THE MIXED MODEL FOR MULTILEVEL DATA

The preceding discussion of the mixed model focused on its use for longitudinal data, as it presents an extremely intuitive as well as powerful tool for modeling repeated measurements. The fixed effects in general represent population averages over time; the random effects symbolize the subject-specific deviations from those population averages and thus can be used to estimate individual trajectories over time. The ability to handle unbalanced and incomplete data of various forms makes it especially appealing, as missing data is almost a certainty in longitudinal studies.

The longitudinal study is one specific type of a general multilevel study, where observations are assumed to be correlated within clusters. In that specific case, the cluster is the individual person, and the observations are the repeated measurements. However, more and more studies involve cluster designs that are not longitudinal, but nonetheless have to account for lack of independence of the data. For example, nursing homes can be considered a cluster in the example discussed in the *Introduction*. It is very possible that residents within a nursing home are not independent of one another relative to residents from other nursing homes in the study. In this case, there are two sources of variation: between-cluster variation and within-cluster variation. Any model of the data should initially include a random effect for this cluster variable. Such a random effect is equivalent to a random intercept in the longitudinal data setting, and leads to exchangeable correlation among observations within a cluster, also referred to as compound symmetry. In other words, in model (3), $\boldsymbol{Z}_i = \mathbf{1}_{p_i}$, and $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e) = \sigma_e^2 \boldsymbol{I}_{p_i}$ (a conditional independence assumption for within-cluster residual). The responses have

$$\boldsymbol{\mathcal{V}}(\boldsymbol{y}_i) = \boldsymbol{\Sigma}_i = \mathbf{1}_{p_i}\sigma_d^2 \mathbf{1}'_{p_i} + \sigma_e^2 \boldsymbol{I}_{p_i}. \tag{6}$$

The intraclass correlation coefficient (ICC), $\rho = \sigma_d^2/(\sigma_d^2 + \sigma_e^2)$ is an important function of the variance components that relates the between-cluster variability to the total variability of the outcome.

Prior discussion related to fixed and random effects inference in a longitudinal setting applies in this more general setting as well. It is important to recognize the accurate fixed effect inference is contingent on the number of independent sampling units; often in clustered designs the number of independent sampling units is relatively small. Of course, the amount of correlation, quantified by the ICC (ρ), is usually much smaller than what one

observes in longitudinal studies. Hence, the need for 'large' numbers of subjects is of utmost importance in longitudinal designs; in clustered designs one typically does not need the equivalent number of independent sampling units.

Random effects model selection in this more general setting is usually not a concern. In contrast to longitudinal data, where the correlation among repeated measurements over time can be modeled countless ways, assuming an exchangeable correlation among observations from the same cluster (i.e., a random effect for the cluster variable) is pragmatically the only option. Of course, nested designs with multiple levels, such as classrooms within schools, would allow for more complicated covariance models. But, even in these settings one would include nested random effects for each level of nesting observed in the data.

FIXED VERSUS RANDOM EFFECTS IN META-ANALYSIS

An area of study where the issue of fixed versus random effects has received great attention is meta-analysis research. Meta-analysis is an approach in which the findings for a similar outcome from different studies are combined quantitatively, and is most often accomplished by using aggregate data (AD) from each study. The statistical aspects of an AD meta-analysis encompass a two-stage approach. In the first stage, the summary statistics from each study are calculated. In the second stage, these summary statistics from each study are combined to yield an overall result, which is ultimately of interest.

Consideration of the general definitions of fixed and random effects is relevant to the desired approach taken when performing a meta-analysis. A fixed-effects meta-analysis assumes that all observed variation is due to within-study sampling error, and all studies are assumed to measure the same overall effect, i.e. a single true effect size, θ. In contrast, the intercept-only random-effects model can be written using notation similar to other contexts described earlier:

$$y_i = \theta + b_i + e_i.$$

Hence, the random-effects approach allows the effect size to randomly vary between studies, ($\theta_i = \theta + b_i$, where i indexes the study). Typically, $b_i \sim \mathcal{N}(0, \tau^2)$. The fixed-effects model is appropriate for an AD meta-analysis when all included studies are identical and the goal is to estimate a common effect size for the identified population. However, motivation to generalize beyond the precise population of included studies which are most likely not identical results in the need to use the random-effects model when performing a meta-analysis.

In addition to generating an overall effect size from the pooling of studies, one may also be interested in looking at the effect of potential covariates on the overall effect size, especially, but not exclusively, when significant between-study heterogeneity exists. Since the random versus fixed-effects model is usually the more appropriate model to use, we focus our discussion on the random-effects model in the form of regression, i.e., meta-regression, denoted as:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + b_i + e_i \; ; \; i = 1, 2, ..., N. \qquad (7)$$

Model (7) is a version of model (3), where each independent sampling unit (ISU) i has an outcome y_i (the 'effect' reported from the study), a vector of fixed effects $\boldsymbol{x}_i$ with associated parameters $\boldsymbol{\beta}$, and a random effect associated with the study (b_i) which represents between-study variation from the overall effect y_i, adjusting for covariates. The primary distinction between this 'mixed model' and others introduced thus far is the fact that each ISU has one observation associated with it, whereas longitudinal and clustered data have multiple observations per ISU. Nonetheless, the random effect plays an important role in modeling the variation between these independent sampling units.

Estimation of model (7) has been extensively researched, much of the time using methods similar to those used in longitudinal and other multilevel settings. However, many of the concepts introduced earlier do not translate in a straightforward fashion due to the fact the above mixed model is used in a setting with one observation per ISU. Many methods for estimating the parameters in (7) have been proposed, with the primary distinction among them being the manner in which the between-study variance τ^2 is estimated.[29] Currently, the most common approach is an extension of the noniterative method of moments approach originally proposed by Dersimonian and Laird,[30] but flaws in this estimation approach have been found. For example, Brockwell and Gordon found that even with large sample sizes, confidence intervals of odds ratios based on the DerSimonian and Laird method did not attain the preferred coverage.[31] While not commonly used, DerSimonian and Kacker have recently provide updated random-effects models that they believe are more valid than the original DerSimonian and Laird model.[32]

Alternative and more computationally intensive parametric random-effects models for estimating

between-study heterogeneity in an AD meta-analysis have been proposed. These include, but are not necessarily limited to, ML, REML, and profile likelihood.[29,33,34] One nonparametric random-effects model, the permutations model, has also been proposed for estimating the between-study variance in an aggregate data meta-analysis.[35] While not discussed further, the empirical Bayes model for estimating τ^2 has also been proposed for random-effects meta-analyses.[34] However, the results for such have been mixed.[36] Thompson and Sharp[29] advocate for the use of a REML over ML for the same reasons REML is advocated in longitudinal settings, a particularly relevant argument when the number of studies included are small, a common occurrence in meta-analysis.

RANDOM EFFECTS IN OTHER SETTINGS

The discussion up to this point has revolved around linear models with random effects. However, nonlinear models can include random effects as well. The 'nonlinear mixed model' follows the same general form as (2), but the function f is typically more complex than the standard functions used for the generalized linear mixed model. Nonlinear mixed models are often used in pharmacokinetics and various biological growth models where there is a known or suspected form of this function.

Random effects can also be included in models of time-to-event data as well. The mixed model approach in estimating time to a certain event has two main uses that parallel the above discussion, depending on the nature of the event to be modeled. In the case of time-to-event data where the event can only occur once (e.g., death), but subjects are clustered, a random effect of the cluster variable should be included when making inferences on the fixed effects of the traditional survival model. Mixed time-to-event models are also utilized when the event occurs repeatedly on the same individuals (e.g., falls in the elderly), and thus we have repeated durations on each individual that should be modeled accordingly. For a detailed discussion of what is often referred to as 'multilevel' survival data models, see Goldstein.[37]

CONCLUSIONS

The issue of classifying effects as fixed versus random has long been a point of discussion and debate. There are inherent characteristics of both, however, and understanding these characteristics, particularly considering the context of the analysis, can lead to informative and powerful conclusions regarding the relevant research questions. Additionally, the simultaneous use of fixed and random effects has distinct advantages, but also pitfalls of which the data analyst must be cognizant. Mixed models that contain both types of variables in longitudinal settings can allow for conclusions about time trends to be made at both the population level and the individual level. The use of these models in multilevel studies in general allow for conclusions about the means of the data as well as the variation of the data. Utilization of random effects in particular for meta-analysis greatly increases the scope of the conclusions made regarding the results. These are just a few major areas of study where the utilization of fixed and random effects, either separately or in tandem, can aid the researcher in making the most accurate and robust statements. Their use carries to multiple other areas of research as well. As the use of these 'mixed' models extend to other more general settings, with various classes of outcomes and study designs, the need for careful statistical research into these extensions and their proper use is necessary. The potential impact of these tools in answering important research questions is matched by the risk of misusing these models.

ACKNOWLEDGMENTS

Dr. Gurka's work on the SECCYD dataset was supported in part by NIH/NICHD grant R03-HD055298. Use of the SECCYD data was classified as exempt by the University of Virginia and West Virginia University Institutional Review Boards.

REFERENCES

1. McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd ed. New York: Chapman and Hall; 1989.
2. Molenberghs G, Verbeke G. *Models for Discrete Longitudinal Data*. New York: Springer; 2005.
3. Neuhaus JM, Kalbfleisch JD, Hauck WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *Int Stat Rev* 1991, 59:25–35.

4. McCulloch CE, Searle SR. *Generalized, Linear and Mixed Models*. New York: Wiley; 2001.
5. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982, 38:963–974.
6. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag; 2000.
7. Zucker DM, Lieberman O, Manor O. Improved small sample inference in the mixed linear model. *J R Stat Soc B* 2000, 62:827–838.
8. Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. *J Am Stat Assoc* 1977, 72:320–338.
9. Harville DA. Bayesian inference for variance components using only error contrasts. *Biometrika* 1974, 61:383–385.
10. Dempster AP, Rubin RB, Tsutakawa RK. Estimation in covariance components models. *J Am Stat Assoc* 1981, 76:341–353.
11. Catellier DJ, Muller KE. Tests for Gaussian repeated measures with missing data in small samples. *Stat Med* 2000, 19:1101–1114.
12. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997, 53:983–997.
13. Gomez EV, Schaalje GB, Fellingham GW. Performance of the Kenward-Roger method when the covariance structure is selected using AIC and BIC. *Commun Stat Simul Comput* 2005, 34:377–392.
14. Skene SS, Kenward MG. The analysis of very small samples of repeated measurements I: an adjusted sandwich estimator. *Stat Med* 2010, 29:2825–2837.
15. Verbeke G, Molenberghs G. The use of score tests for inference on variance components. *Biometrics* 2003, 59:254–262.
16. Verbeke G, Lesaffre E. The effect of misspecifying the random effects distribution in linear mixed models for longitudinal data. *Comput Stat Data Anal* 1997, 23:541–556.
17. Gurka MJ, Edwards LJ, Muller KE, Kupper LL. Extending the Box-Cox transformation to the linear mixed model. *J R Stat Soc A* 2006, 169:273–288.
18. Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random-effects population. *J Am Stat Assoc* 1996, 91:217–221.
19. Ho HJ, Lin TI. Robust linear mixed models using the skew t distribution with application to schizophrenia data. *Biom J* 2010, 52:449–469.
20. Cheng J, Edwards LJ, Maldonado-Molina MM, Komro KA, Muller KE. Real longitudinal data analysis for real people: building a good enough mixed model. *Stat Med* 2010, 29:504–520.
21. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control* 1974, AC-19: 716–723.
22. Schwarz G. Estimating the dimension of a model. *Ann Stat* 1978, 6:461–464.
23. Gurka MJ. Selecting the best linear mixed model under REML. *The American Statistician* 2006, 60:19–26.
24. Edwards LJ, Muller KE, Wolfinger RD, Qaqish BF, Schabenberger O. An R^2 statistic for fixed effects in the linear mixed model. *Stat Med* 2008, 27:6137–6157.
25. Jacqmin-Gadda H, Sibillot S, Proust C, et al. Robustness of the linear mixed model to misspecified error distribution. *Comput Stat Data Anal* 2007, 51:5142–5154.
26. Gurka MJ, Edwards LJ, Muller KE. Avoiding bias in mixed model inference for fixed effects. *Stat Med* (Epub ahead of print; 2011).
27. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986, 73:13–22.
28. Gurka MJ, LoCasale-Crouch J, Blackman JA. Long term cognition, achievement, socioemotional, and behavioral development of healthy late preterm infants. *Arch Pediatr Adolesc Med* 2010, 164:525–532.
29. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999, 18:2693–2708.
30. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986, 7:177–188.
31. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001, 20:825–840.
32. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials* 2007, 28:105–114.
33. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Stat Med* 1996, 15:619–629.
34. Morris CN. Parametric empirical Bayes inference—theory and applications. *J Am Stat Assoc* 1983, 78:47–55.
35. Follmann DA, Proschan MA. Valid inference in random effects meta-analysis. *Biometrics* 1999, 55:732–737.
36. Sidik K, Jonkman JN. A comparison of heterogeneity variance estimators in combining results of studies. *Stat Med* 2007, 26:1964–1981.
37. Goldstein H. *Multilevel Statistical Models*. 3rd ed. London: Arnold; 2003.

