Communications in Statistics—Theory and Methods, 40: 515–531, 2011

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920903411192

Estimating Variance Components and Random Effects Using the Box-Cox Transformation in the Linear Mixed Model

MATTHEW J. GURKA[1] AND LLOYD J. EDWARDS[2]

[1]Department of Community Medicine, School of Medicine, West Virginia University, Morgantown, West Virginia, USA
[2]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, USA

The linear mixed model assumes normality of its two sources of randomness: the random effects and the residual error. Recent research demonstrated that a simple transformation of the response targets normality of both sources simultaneously. However, estimating the transformation can lead to biased estimates of the variance components. Here, we provide guidance regarding this potential bias and propose a correction for it when such bias is substantial. This correction allows for accurate estimation of the random effects when using a transformation to achieve normality. The utility of this approach is demonstrated in a study of sleep-wake behavior in preterm infants.

Keywords Linear mixed model; Longitudinal data; Normality; Random effects; Transformation.

Mathematics Subject Classification Primary 62J05; Secondary 62P10.

1. Introduction

1.1. *Background and Motivation*

Seminal articles by Harville (1977) and Laird and Ware (1982) helped popularize the use of linear mixed models. Many theoretical and applied papers followed (e.g., Kenward and Roger, 1997), and several texts now include discussions of linear and nonlinear mixed models (e.g., Vonesh and Chinchilli, 1997). The linear mixed model has consequently become a standard tool when analyzing longitudinal data with continuous outcomes. This model has two sources of randomness: the random effects and the residual error. In the context of longitudinal

Received September 26, 2008; Accepted October 13, 2009
Address correspondence to Matthew J. Gurka, Department of Community Medicine, School of Medicine, West Virginia University, P.O. Box 9190, Morgantown, WV 26506-9190, USA; E-mail: mgurka@hsc.wvu.edu

data, the former represents among-subject variation, and the latter represents the within-unit variation. The random effects and the residual error are assumed to be normally distributed and independent of one another.

Previous work has focused on the consequences of fitting a linear mixed model when the random effects do not follow a Gaussian distribution. It has been shown that non normality of the random effects does not impact estimation of the fixed effects (Butler and Louis, 1992; Verbeke and Lesaffre, 1997). However, Verbeke and Lesaffre (1996) demonstrated that such a violation does in fact lead to poor inference about the random effects. Thus, alternative forms of the linear mixed model have been proposed that relax the assumption of normality of the random effects. Zhang and Davidian (2001) assumed a semi-nonparametric density of the random effects that includes the Gaussian distribution as a special case. Verbeke and Lesaffre (1996) allowed for a mixture of normal distributions for the random effects. Arellano-Valle (2005) assumed a skew-normal distribution of the random effects and the residual error.

Transformation of the response in the univariate linear model setting is a common tool when the assumption of normality of the error term is in question. The work of Box and Cox (1964), among others, led to the development of methods for transforming the response in linear models with independent and identically distributed (i.i.d) error terms. In the univariate setting, making inferences about the model parameters in the presence of a transformation has generated a great deal of debate (Hinkley and Runger, 1984). Some argue that treating the transformation parameter as non stochastic when making inferences on the other parameters is valid. However, many contend that one must take into account the estimation of the unknown transformation parameter in this case (Bickel and Doksum, 1981). Although debated from a methodological standpoint, the application of transformations in univariate models is commonplace in order to utilize a parametric model that best meets the assumptions made on its error term.

Transformations are also used in many fields as a simple way to achieve normality when fitting models of longitudinal data. One such example is recent research in accurately measuring food intake from multiple 24-h food recalls (Dodd et al., 2006; Tooze et al., 2006). Gurka et al. (2006, 2007) provided a general methodological overview of the extension of the Box-Cox transformation to the linear mixed model. Gurka et al. (2006) demonstrated that given the standard independence assumption of the two sources of randomness in the linear mixed model, a single transformation parameter would simultaneously target normality of both the random effects and the residual error. The Box-Cox method is thus one straightforward way to achieve normality of the random effects in the linear mixed model, allowing for accurate inference when the variance components and/or the random effects are of primary interest.

Gurka et al. (2006) showed that estimating the transformation parameter leads to biased estimates of the variance components of the mixed model. This bias was found to increase for increasing values of the transformation parameter, a phenomenon that was similarly observed in previous research on transformations for the univariate linear model (Spitzer, 1978). Correction for the bias observed in the variance component estimates could be very useful for some problems. In this article, we examine the implications of using a transformation approach when the variance components or the subsequent random effects of the mixed model are of interest. In doing so, we provide practical guidance for the data analyst regarding when one should be concerned about the incurred bias of these parameters when estimating the transformation. We then propose a correction for this bias in the variance component estimates when deemed necessary. This correction will allow for accurate estimation and inference about the random effects in the linear mixed model when a transformation is utilized to achieve normality.

1.2. *Sleeping and Waking States of Preterm Infants and Subsequent Development*

Whenever focus lies on the random effects of the model itself, the analyst must be concerned with the Gaussian distribution assumption. One study, involving the analysis of sleep-wake behavior measures in preterm infants, demonstrates an area of application where a simple transformation to achieve normality is appealing, and interest lies in the random effect estimates of the fitted mixed model. The sleeping and waking states of preterm infants have been widely studied because these states are strongly associated with developmental outcomes later in life (Parmelee et al., 1967; Dreyfus, 1975; Korner et al., 1988; Holditch-Davis, 1990; Whitney and Thoman, 1993; Holditch-Davis and Edwards, 1998; Gertner et al., 2002). Identifying subjects with extreme values of these states is thus extremely important from both a research and a clinical perspective. Holditch-Davis and Edwards (1998) modeled various measures of sleep-wake states on 37 high-risk preterm infants and another group of 34 infants over time via linear mixed models. They initially were interested in accurately modeling the changes in these states as the infants aged, as well as assess for a difference between the two cohorts. For most of the sleep-wake measures, no substantial difference was observed between the two groups. The results suggested that developmental patterns of sleep-wake states are fairly stable in the preterm period, leading the authors to conclude that substantial deviations in individual patterns (i.e., random effect estimates) may be useful in identifying infants with neurological problems.

Subsequent research by Holditch-Davis and Edwards (2005) then used random effect estimates from the original sleep-wake models to predict developmental outcomes such as IQ at three years of age in this group of infants. The authors used random intercept and slope estimates from the models of various sleep-wake states as predictors for many of the outcomes measured three years later. For simplicity in demonstration of the proposed methods, we focused on models including only a random intercept. To summarize the analysis here, the general model for various sleep wake state measures is:

$$y_{ij} = \beta_0 + \beta_1 \mathrm{AGE}_{ij} + b_{0i} + e_{ij}, \tag{1}$$

where y_{ij} is the jth observation of the sleep-wake measure of interest on the ith preterm infant, $j = 1, \ldots, n_i$. The model-based estimates of b_{0i} from each infant could then be used as a predictor of various developmental outcomes of the child at three years of age. Table 1 lists only three of many sleep-wake states with brief descriptions that will be the focus of further analysis here, as well as the number of subjects and observations for the three variables. The variable associated with each of the states is a percentage of time in that particular state during a four-hour observation of the infant.

Table 1
Sleep-wake outcomes of interest and their descriptions

Sleep-wake states	Description	Number of subjects (observations) with data
Quiet Sleep	The infant's eyes are closed, and breathing is relatively regular. Motor activity is limited to occasional startles, sighs, or other brief movements.	69 (257)
Quiet Waking	The infant's eyes are open, or opening and is usually low and respiration even, but the infant may be active for brief periods when clearly alert and drowsy.	66 (257)
Active Waking	The infant's eyes are usually open, but unfocused. The infant may be fussing or crying. Motor activity is high and variable.	69 (287)

*The variable associated with each of the states is a percentage of time in that state during a four-hour observation of the infant.

As is standard in the linear mixed model, the random effects and within-subject error term are assumed to independently follow normal distributions, $b_{0i} \sim N(0, \sigma_b^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$, respectively. Previous sleep-wake studies assumed normality of both sources of randomness. Here, we evaluate the utility of a transformation in model (1) with respect to meeting the normality assumption. Rather than choosing a particular transformation a priori, we wish to estimate an appropriate transformation. If a transformation helps in achieving normality of the error terms in (1), improved estimates of b_{0i} will follow, which in turn will improve the analysis that uses the estimated individual deviation from "normal" sleep-wake behavior to predict development in the children three years later.

This example highlights an important possible area of application of this transformation method for the linear mixed model. It is true that other methods exist when normality of the random effects in particular is questioned. However, such methods may not be readily accessible and understandable to the analyst and his/her collaborators. The idea of transforming the data to better meet the assumptions of the analysis is a relatively standard and common procedure due to its lucidity. Interpretability of results is one major and understandable criticism of the transformation approach since the final model of the outcome is on a different scale. The sleep-wake research exemplifies an area of research where the final scale of the model does not matter; what is valuable is identifying the subjects (preterm infants) who deviate from normal sleep-wake behavior. It is vital to obtain accurate estimates of these deviations, no matter what the final scale of the model. To do this, normality of these random deviations is important, and we show the transformation approach to be extremely helpful in analyzing these data.

2. The Box-Cox Transformation for the Linear Mixed Model

To generalize, the linear mixed model with a parametrically transformed response for $i \in \{1, \dots, m\}$, m the number of independent sampling units (subjects), is given by

$$\boldsymbol{y}_i^{(\lambda)} = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i. \tag{2}$$

Here, $\boldsymbol{y}_i$ is an $n_i \times 1$ vector of observations on subject i, $\boldsymbol{X}_i$ is an $n_i \times p$ known, constant design matrix for the ith subject with rank p, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown, constant population parameters. Also, $\boldsymbol{Z}_i$ is an $n_i \times q$ known, constant design matrix for the ith subject with rank q corresponding to $\boldsymbol{b}_i$, a $q \times 1$ vector of unknown, random individual-specific parameters, and $\boldsymbol{e}_i$ is an $n_i \times 1$ vector of random within-subject, or residual, error terms. Additionally, let $\boldsymbol{\epsilon}_i = \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$ be the "total" error term of model (2).

When utilizing linear mixed models, in general, we assume that the following distributional assumptions are reasonably valid: $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ independently follow normal distributions with mean vectors $\mathbf{0}$ and covariance matrices $\boldsymbol{D}$ and $\boldsymbol{R}_i$, respectively. The covariance matrices $\boldsymbol{D}$ and $\boldsymbol{R}_i$ are characterized by unique parameters contained in the $k \times 1$ vector $\boldsymbol{\theta}$. Thus, the total variance for the response vector in (2) is $\mathscr{V}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{R}_i$, where $\mathscr{V}(\cdot)$ is the variance-covariance operator. Since the model pertains to a single outcome measured repeatedly, $\boldsymbol{y}_i$ is transformed using the form proposed by Box and Cox (1964) with a single transformation parameter λ, for all subjects $i \in \{1, \dots, m\}$ and measurement occasions $j \in \{1, \dots, n_i\}$, $y_{ij} > 0$:

$$y_{ij}^{(\lambda)} = \begin{cases} (y_{ij}^{\lambda} - 1)/\lambda & \lambda \neq 0 \\ \log(y_{ij}) & \lambda = 0 \end{cases}. \tag{3}$$

The standard linear mixed model with response y_{ij} is obtained when $\lambda = 1$. Using the transformation (3) requires $y_{ij} > 0$ for all i and j. Box and Cox (1964) also studied a shifted power transformation to avoid the restriction. Since the current interest is in the random effects, a residual maximum likelihood approach (REML) is employed because REML provides a less biased estimator of $\boldsymbol{\theta}$ than ML for finite values of m.

We wished to develop methods that are accessible to the widest possible audience and to take advantage of existing computational procedures to estimate λ. An estimate of λ cannot be obtained from transformation (3) using standard linear mixed model software because transformation (3) leads to a likelihood slightly different than the one from the standard model. If a goal is to use existing mixed model estimation procedures and software, one should replace (3) with a scaled transformation (Box and Cox, 1964):

$$w_{ij}^{(\lambda)} = \begin{cases} (y_{ij}^{\lambda} - 1)/(\lambda\tilde{y}^{\lambda-1}) & \lambda \neq 0 \\ \tilde{y} \cdot \log(y_{ij}) & \lambda = 0 \end{cases}, \tag{4}$$

where $\tilde{y} = \left(\prod_{i=1}^{m} \prod_{j=1}^{n_i} y_{ij}\right)^{\frac{1}{n}}$, the geometric mean of the response values. The scaled transformation leads to the model

$$\boldsymbol{w}_i^{(\lambda)} = \boldsymbol{X}_i\boldsymbol{\beta}_* + \boldsymbol{Z}_i\boldsymbol{b}_{*i} + \boldsymbol{e}_{*i}, \tag{5}$$

with the same assumptions as model (2). Conditioned on the geometric mean, the Jacobian of (4) is equal to one. Consequently, the residual log-likelihood in terms of the original observations $\boldsymbol{y}$, ignoring the constant terms, is that of a standard linear mixed model, where $\mathscr{V}(\boldsymbol{b}_{*i}) = \boldsymbol{D}_*$, $\mathscr{V}(\boldsymbol{e}_{*i}) = \boldsymbol{R}_{*i}$, $\boldsymbol{\theta}_*$ constitutes the set of variance parameters, and $\mathscr{V}(\boldsymbol{w}_i^{(\lambda)}) = \boldsymbol{\Sigma}_{*i} = \boldsymbol{Z}_i\boldsymbol{D}_*\boldsymbol{Z}_i' + \boldsymbol{R}_{*i}$. By using the scaled transformation (4), we can employ existing linear mixed model estimation procedures (such as SAS PROC MIXED) to obtain the REML estimates of the model (5) parameters for a given λ. We do this by fitting the model (5) for a wide range of values for λ and finding the value, $\hat{\lambda}$, that maximizes the corresponding residual likelihood. With respect to λ, maximizing the residual likelihood of (5) using the scaled transformation, $\boldsymbol{w}_i^{(\lambda)}$, is equivalent to maximizing the residual likelihood of (2) using $\boldsymbol{y}_i^{(\lambda)}$ (Spitzer, 1982). Since we assume that model (2) and its associated parameters and random effects are of practical interest, we need to convert the model parameter estimates that maximize the residual likelihood associated with (5) back to those based on (2). Given $w_{ij}^{(\lambda)} = y_{ij}^{(\lambda)}/\tilde{y}^{\lambda-1}$, we can state:

$$\begin{aligned} \boldsymbol{\beta} &= \tilde{y}^{\lambda-1} \cdot \boldsymbol{\beta}_* \\ \boldsymbol{b}_i &= \tilde{y}^{\lambda-1} \cdot \boldsymbol{b}_{*i} \\ \boldsymbol{e}_i &= \tilde{y}^{\lambda-1} \cdot \boldsymbol{e}_{*i}, \end{aligned} \tag{6}$$

where $\tilde{y}$ is considered fixed.

The discussion of transformations and their use in practice naturally leads to the issue of interpretability. Interpretability becomes even more problematic if one were to use the model of the scaled transformation (5) due to the inclusion of the geometric mean of the response data. In the case of the sleep-wake study, this is not an issue due to the desire in simply obtaining accurate random effect estimates to use in a subsequent analysis. However, even in this study, interest also lied in the fixed effects of the model as well (i.e., whether or not these states varied over time). In this case, estimation and inference of the parameters of (2) are favored, taking advantage of the fact that $\hat{\lambda}$ obtained from the scaled model (5) holds for the original transformation model (2). To summarize, we take advantage of a scaled transformation (4) in order to use existing software to obtain the appropriate value of the transformation, but use a more interpretable version of the Box-Cox transformation to obtain parameter estimates of the mixed model (2). SAS macros are available from the first author that implement this strategy and allow the user to avoid this process.

3. Estimation and Inference: Complications and Solutions

Gurka et al. (2006) proved that parametrically transforming the response will target both the normality of the random effects, $\boldsymbol{b}_i$, and the within-unit error term, $\boldsymbol{e}_i$, and they provided likelihood-based methods for estimating the parameters of the model (2). The authors specifically provide an approach for accurately estimating and inferring about the fixed effects of the model when a transformation is used, treating $\boldsymbol{\theta}$ as fixed. Here, we present methods for accurate estimation of $\boldsymbol{\theta}$, which will naturally lead to improved estimation of the random effects. As seen in Laird and Ware (1982), "empirical Bayes" (EB) estimates of the random effects (from the transformed model (2)):

$$\hat{\boldsymbol{b}}_i = \widehat{\boldsymbol{D}}\boldsymbol{Z}_i'\widehat{\boldsymbol{\Sigma}}_i(\boldsymbol{y}_i^{(\lambda)} - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) \tag{7}$$

are often used for diagnostic purposes Verbeke and Molenberghs (2000). The EB estimates (7) are based on the estimates of the model (2), and thus the importance of accurate estimates of the model parameters is evident. Simulations in Gurka et al. (2006) demonstrated that problems arise when estimating the parameters of model (2) and their variances in the presence of an unknown transformation parameter. We take the position that one should attempt to adjust for the estimation of λ when estimating the variances of the other model parameters. Lipsitz et al. (2000) also made this assessment and proposed a jackknife method to estimate the variances of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\lambda})$ given their dependence. This approach is sound for the most general case. Nevertheless, we present a likelihood-based method that is a computationally less intensive alternative and valid for large sample sizes. This alternative is potentially easier to implement, particularly given the availability of an associated SAS macro from the first author.

We initially compute the observed information matrix for $\{\boldsymbol{\beta}_*, \boldsymbol{\theta}_*, \lambda\}$ from model (5); see the Appendix for details. Here, we are ultimately interested in $\hat{\boldsymbol{\theta}}$ in model (2), but as seen in simulations below, naive estimates of $\hat{\boldsymbol{\theta}}$ are biased and the magnitude of this bias increases for increasing values of λ. This bias was also observed for the variance estimate in the univariate linear model (Spitzer, 1978). To estimate this bias here and correct for it, we use a Taylor series expansion about $\hat{\boldsymbol{\theta}}$ utilizing (6) (see the Appendix for details). The corrected estimators that result then provide for improved estimation and inference about $\boldsymbol{b}_i$. Unfortunately, due to the complex nature of the corrected variance component estimators, estimating the variance of these corrected estimators using likelihood-based methods is extremely difficult. In the relatively rare instances when one wishes to perform inference on the corrected variance components, we suggest the use of a bootstrap or jackknife method. It is important to highlight that in those instances when the chosen transformation is small in magnitude (e.g., the log transformation), the bias of the resulting variance component estimates is negligible. Therefore, we recommend using the uncorrected variance component estimators and their readily available standard error estimates. A SAS macro that performs these calculations is available from the first author.

4. Simulation Results

We performed a small simulation study to demonstrate the impact of estimating λ on the variance component parameter estimates, and to evaluate the effectiveness of the adjusted estimators proposed in Sec. 3 and described in detail in the Appendix. For each $\lambda \in \{0, 0.5, 1\}$, we generated 10,000 Monte Carlo data sets, each with 65 subjects and 5 observations per subject at 5 time points equally spaced between 30 and 38 weeks. For the fixed effects, we used $\boldsymbol{\beta} = (\beta_0, \beta_1)'$, containing an intercept, β_0, and slope, β_1. For the sake of simplicity, only a random intercept was included in the model, $b_i \sim \mathcal{N}(0, 0.20)$, and $e_{ij} \sim \mathcal{N}(0, 1.55)$. We set the intercept parameter $\beta_0 = 5$ to assure $y_{ij}^{(\lambda)} > 0$. The following model was then generated:

$$y_{ij}^{(\lambda)} = 5 + (0.25)t_{ij} + b_i + e_{ij}. \tag{8}$$

The above model was used to parallel the analysis of the active waking outcome, but for a variety of transformation values to demonstrate the impact of the transformation on the estimation bias. Estimates and their corresponding standard errors for λ, β_1, σ_b^2, and σ_e^2 using the methods described in this article are displayed in Table 2. We wished to replicate the "naive" approach in finding an estimate of λ and then fitting model (8), transforming the response by $\hat{\lambda}$. This "naive" approach treats λ as fixed and known when estimating σ_b^2 and σ_e^2; we report the average values of the 10,000 bias-unadjusted estimates ($\hat{\sigma}_{b0}^2$ and $\hat{\sigma}_{e0}^2$, respectively). Additionally, the average values of our bias-adjusted estimators are displayed ($\hat{\sigma}_{b1}^2$ and $\hat{\sigma}_{e1}^2$, respectively).

Estimation of λ using the described methods is again proven reliable. As discussed earlier, an increasing bias of $\hat{\sigma}_{b0}^2$ and $\hat{\sigma}_{e0}^2$ is observed for increasing values of λ. This is in part due to the increasing variability for larger values of λ, which leads to greater likelihood of error in the estimates of the variance components. It is also of note that this bias exists in moderately sized settings; presumably this bias increases for smaller sample sizes. Guidance for practical solutions arise from this observation. If one estimates the transformation, and the value is close to 0, (e.g., the log-transformation), one does not need to worry

Table 2
Simulation results for 10,000 Monte Carlo replications for the transformed mixed model (8): means and standard errors of parameter estimates. Column values of λ are considered the true values from which the data sets were generated.

	$\lambda = 0$		$\lambda = 0.5$		$\lambda = 1$	
Estimates	Mean	SE	Mean	SE	Mean	SE
$\hat{\lambda}$	−0.001	0.029	0.492	0.111	0.988	0.192
$\beta_1 = 0.25$						
$\hat{\beta}_{1,0}$	0.253	0.051	0.256	0.082	0.261	0.101
$\hat{\beta}_{1,1}$	0.249	0.051	0.244	0.078	0.243	0.094
$\sigma_b^2 = 0.20$						
$\hat{\sigma}_{b0}^2$	0.211	0.128	0.228	0.191	0.248	0.241
$\hat{\sigma}_{b1}^2$	0.198	0.121	0.185	0.157	0.181	0.178
$\sigma_e^2 = 1.55$						
$\hat{\sigma}_{e0}^2$	1.628	0.604	1.772	1.141	1.912	1.545
$\hat{\sigma}_{e1}^2$	1.526	0.567	1.439	0.934	1.386	1.131

SE = Monte Carlo standard error of the model parameter estimates.
$\hat{\beta}_{1,0}$, $\hat{\beta}_{1,1}$ = Estimated "naive" and "adjusted" fixed effect age estimate.
$\hat{\sigma}_{b0}^2$, $\hat{\sigma}_{e0}^2$ = Estimated "naive" variance component parameter estimates.
$\hat{\sigma}_{b1}^2$, $\hat{\sigma}_{e1}^2$ = Estimated "adjusted" variance component parameter estimates.

about resulting bias of the variance component estimates. As values of λ increase (e.g., the square-root transformation), the analyst should be concerned about bias in the variance component estimates.

In this case, our proposed adjustment taking into account the estimation of λ works well in removing most of this bias, as observed in Table 2. In fact, the adjustment over-corrects for this bias, but this should not be seen as surprising. The correction is relatively simple, using estimates of the variances and covariances of the parameters in subtracting out the quantity used to estimate the bias via a Taylor series expansion. However, the improvement in accuracy of the variance component parameter estimates is striking, even for small values of λ. The unadjusted estimate of σ_b^2 in the case of the log-transformation has roughly a 5% bias, while the adjusted estimate has only a 1% bias. In the case of the null transformation ($\lambda = 1$), the unadjusted estimate of σ_b^2 has roughly an 24% bias, compared to the 9.5% bias in the opposite direction of its adjusted counterpart. It is also worth noting that the adjusted estimates are more precise, as exhibited by the smaller Monte Carlo standard errors relative to their unadjusted counterparts.

5. Sleep-Wake Behavior Analysis

We explored the data from the sleep-wake study using our proposed methods, not only to demonstrate the utility of our methods, but also to examine the necessity of a transformation for the model of interest (1). The transformed linear mixed model for the data is a generalization of (1):

$$y_{ij}^{(\lambda)} = \beta_0 + \beta_1 \mathrm{AGE}_{ij} + b_{0i} + e_{ij}, \tag{9}$$

for the three outcomes listed in Table 1. The proposed methods are used to estimate λ (using the scaled transformed model) and to obtain accurate estimates of σ_b^2 and σ_e^2. These estimators are then used to compute the EB estimates of b_{0i}.

For all three outcomes, Table 3 displays the estimate of λ, as well as the unadjusted and adjusted estimates of σ_b^2 and σ_e^2. Additionally, estimates of the variance components when $\lambda = 1$ (no transformation) are also reported. Thus, we have three estimates of each of the variance components: (1) from the model of the untransformed response; (2) from the model of the transformed response not adjusting for the estimation of λ; and (3) adjusting for the estimation of λ. A transformation appears beneficial for all three sleep-wake state outcomes, with values of λ ranging from 0.1–0.7. In practice, if interest lied in the fixed effects, or the actual values of the random effects, one may need to worry about the interpretability of such a precise estimate of the transformation. In this case, it may be helpful to explore the use of the log-transformation when $\lambda = 0.11$, or the square-root transformation when $\lambda = 0.34$, or no transformation at all when $\lambda = 0.71$. However, in this particular application, the resulting random effect estimates are used as predictors in a second analysis to help identify and characterize infants who are at high-risk of developmental problems at three years of age. Hence, interpretability of the final scale used for each outcome is not as vital.

The results in Table 3 are consistent with the simulation results in Table 2, in that the bias of the variance component estimates when not taking into account the estimation of λ is fairly minimal for small values of λ, but increases as λ gets closer to 1. For active waking ($\hat{\lambda} = 0.11$), the adjusted variance component estimates

Table 3
Parameter estimates from models of sleep-wake outcomes

	Model of sleep-wake state measures		
Parameter	Quiet sleep	Quiet waking	Active waking
$\hat{\lambda}$	0.71	0.34	0.11
Age fixed effect:			
$\hat{\beta}_1(\lambda = 1)$	0.83	0.56	0.71
$\hat{\beta}_{1,0}(\hat{\lambda})$	0.36	0.15	0.23
$\hat{\beta}_{1,1}(\hat{\lambda})$	0.34	0.15	0.23
$\hat{\sigma}^2_b(\lambda = 1)$	7.59	5.06	1.02
$\hat{\sigma}^2_{b0}(\hat{\lambda})$	1.64	0.50	0.19
$\hat{\sigma}^2_{b1}(\hat{\lambda})$	1.42	0.50	0.18
$\hat{\sigma}^2_e(\lambda = 1)$	62.47	34.28	14.97
$\hat{\sigma}^2_{e0}(\hat{\lambda})$	11.24	2.81	1.56
$\hat{\sigma}^2_{e1}(\hat{\lambda})$	9.75	2.78	1.55
Subject ID's of "Low" Outliers*			
$\lambda = 1$	25,66,68,**79**,83,86	22,29,**33**,36,81,**88**	29,32,33,36,**55**,87
$\hat{\lambda}$, $\hat{\sigma}^2_{b0}$, $\hat{\sigma}^2_{e0}$	**2**,25,66,68,83,86	22,29,**31**,36,**61**,81	**12**,29,32,33,36,87
$\hat{\lambda}$, $\hat{\sigma}^2_{b1}$, $\hat{\sigma}^2_{e1}$	**2**,25,66,68,83,86	22,29,**31**,36,**61**,81	**12**,29,32,33,36,87

$\hat{\sigma}^2_b(\lambda = 1)$, $\hat{\sigma}^2_e(\lambda = 1)$ = Variance component (VC) parameter estimates from the untransformed model.

$\hat{\sigma}^2_{b0}$, $\hat{\sigma}^2_{e0}$ = "Naive" VC estimates from the transformed model.

$\hat{\sigma}^2_{b1}$, $\hat{\sigma}^2_{e1}$ = "Adjusted" VC estimates from transformed model.

$\hat{\beta}_1(\lambda = 1)$, $\hat{\beta}_{1,0}(\hat{\lambda})$, $\hat{\beta}_{1,1}(\hat{\lambda})$ = Age estimates: untransformed model, and transformed model (naive and adjusted).

*As determined by the EB estimates of the random intercepts, calculated from each of the three sets of variance component estimators above (those with values less than the 10th percentile). ID's in bold indicate subjects not identified by both the untransformed and transformed model (in this case, using the adjusted variance component estimates did not influence the identification of outlying subjects).

are nearly identical to their "naive" counterparts, while for quiet sleep ($\hat{\lambda} = 0.71$), the adjusted estimates differ more from the naive estimates. We again suggest taking into account the magnitude of λ in deciding whether or not there is a need to "adjust" for bias incurred by estimating and using this transformation.

It is ultimately of interest to obtain accurate random intercept estimates in this example; namely, we wish to identify those who deviate the most from the average profile over time for each of these sleep-wake states. To do this, outlying subjects were identified as those whose random intercepts were less than the 10th percentile. Interestingly, the same outlying subjects were not identified in both the untransformed and the transformed models (Table 3). We also compute the EB estimates (7) of the random intercepts using the adjusted estimates of σ^2_b and σ^2_e from the transformed model, as well as similar biased-corrected estimates of $\hat{\boldsymbol{\beta}}$ proposed by Gurka et al. (2006). The EB estimates adjusting for the estimation of λ led to the identification of the same individuals as when using the estimates

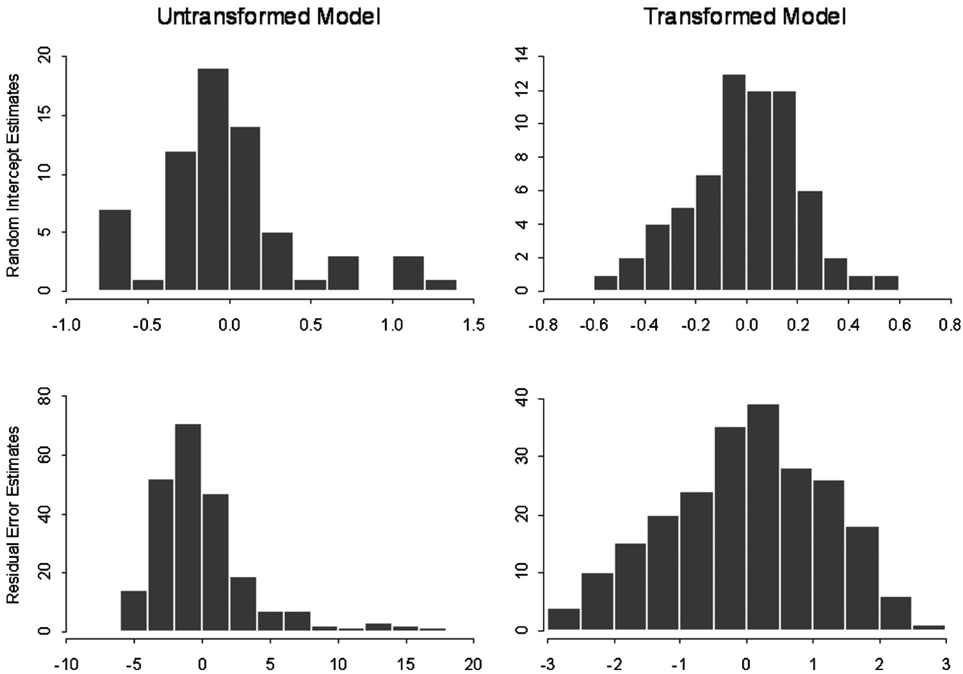

Figure 1. Frequency histogram of random intercept and residual error estimates of the untransformed and transformed ($\lambda = 0.11$) model of active waking.

calculated when ignoring the estimation of λ, although this is not surprising for two of the outcomes where the adjusted and naive estimates of the variance components are nearly identical. We anticipate that one cannot generalize this conclusion beyond this specific application.

The utility of the transformation approach in the linear mixed model setting is demonstrated when comparing the identified outliers to when no transformation is used in Table 3. For all three outcomes, subjects identified as having low values when fitting a simple linear mixed model with no transformation are different than the subjects singled out after implementing the appropriate transformation. To further show the advantages of the transformation approach, we display the frequency histograms of the resulting residuals, $\hat{b}_{0i}$, and $\hat{e}_{ij}$, from the untransformed model of active waking as well as the transformed model ($\lambda = 0.11$) (Fig. 1). Random intercept estimates computed from the uncorrected variance component estimates are displayed here, due to their near equivalence with their "corrected" counterparts (since λ is close to 0). The comparison of the histograms for the untransformed models to those of the transformed models indicate that the assumption of normality for both error terms is improved when transforming the response.

6. Discussion

Many areas of application employ transformations, most often in the context of cross-sectional data. The increasing popularity of longitudinal data analyses requires a thorough examination of all of the available tools to improve upon the assumptions of the models. One such model that is very popular in practice is the linear mixed model. Previous research has indicated that when interest lies in the random effects themselves of the mixed model, the assumption of normality of these random effects is important. We extended examination of transformation of the response of the linear mixed model, an approach shown to be effective in targeting normality of the random effects as well as the residual error. In doing so, we have demonstrated the possible bias in the estimates of the variance components of the mixed model (and hence the random effect estimates) when not taking into account the estimation of the transformation parameter. This bias is primarily a concern for larger values of the transformation parameter. Bias is not an issue for smaller values of the transformation (e.g., the log transformation). As pointed out by a reviewer, the effect of estimating the transformation is also negligible when the mean of the model is very large and the variance of the model is very small. However, due to the focus of the present paper and space considerations, we leave this to future research. This work provides guidance regarding how to implement a transformation to the linear mixed model when interest lies in the variance components and/or the random effects. We then propose a correction to take into account the resulting bias when that is deemed a concern. Hence, we provide a straightforward approach when interest lies in the random effects and their normality is in doubt.

Even though the proposed transformation approach is promising in regards to addressing the normality assumptions of the mixed model, we demonstrate that one must use caution when applying a transformation to the mixed model. The simulation results indicate that one should take into account the estimation of λ when focusing on the other parameters of the model when the estimate of λ is relatively large in magnitude. In this article, we provided straightforward large-sample methods for estimating the variance components, which then can be used to estimate and infer on the random effects. Of course our method does not take into account the problem of shrinkage that one must keep in mind when using random effect estimates to identify outlying subjects. Although other approaches exist that allow for more flexible distributions of the random effects, our proposed methods provide a simple and convenient alternative that may be practical in collaborative research where transformations are commonplace. Our example of the study of sleep-wake behavior, in which random effects estimates from a mixed model of preterm infants were then used to predict outcomes later in life, highlights the benefit of this transformation approach to achieve normality when interpretability of the scale of the final model is not of utmost importance.

This article serves as one component of a general exploration of the idea of transformations in the linear mixed model for longitudinal data, with a valuable application to an important area of biomedical research. It must be noted that the simulation study was not extensive. But, in the interest of a parallel study with the motivating example, we aimed at providing simulation results that were applicable to the type of study exemplified by the sleep-wake example. The results are promising, but more extensive enumerations could be performed. Many further areas of research include examination of influential data points and robust estimation. As alluded to earlier, in many cases the scale of the final model is important, and in these instances researchers may not want to infer about random effect estimates using a transformation. "Transforming back" to the original scale has been studied in the univariate linear model setting (Duan, 1983), and may be worth exploring further for the linear mixed model when random effect estimates in the original scale are of interest. One such area of research where this may prove useful is the study of health care cost and utilization, specifically the identification of outlying clinicians, hospitals or clinics.

As statistical models become more sophisticated, the utility of transformations is increasingly questioned by statisticians. Yet, a quick review of biomedical and other applied literature reveals the continued use of transformations to meet the assumptions of straightforward modeling approaches, such as the univariate linear model. It is therefore important to study the transformation technique when applied to models of correlated data such as the linear mixed model. In doing so, we discovered a real benefit to a simple transformation in meeting the multiple assumptions of normality of the mixed model. However, we also highlighted the pitfalls of transforming the data, namely the bias of the resulting parameter estimates from the model when the transformation parameter is large in magnitude. This bias adversely affects inference about the variance component parameters and consequently the random effect estimates. Our proposed adjustment allows for the valid use of the transformation approach by greatly reducing this bias. Consequently, we provided a straightforward alternative for the practicing statistician and his/her collaborators.

Appendix: Theory and Details from Sec. 3

The residual likelihood from model (5) is maximized to obtain the REML estimates of the parameters. Following Demidenko (2004, Ch. 2), it will be simpler to solve the derivatives using $\boldsymbol{D}_+ = \frac{1}{\sigma_*^2}\boldsymbol{D}_*$. Consequently, denote $\boldsymbol{V}_{*i} = \frac{1}{\sigma_*^2}\boldsymbol{\Sigma}_{*i} = \boldsymbol{Z}_i\boldsymbol{D}_+\boldsymbol{Z}_i' + \boldsymbol{I}_{n_i}$. The following version of the residual log-likelihood will then be utilized:

$$\begin{aligned} L_R(\boldsymbol{\beta}_*, \boldsymbol{\theta}_*) = {} & -\frac{1}{2}(N-p)\log(\sigma_*^2) - \frac{1}{2}\sum_{i=1}^{m}\log(|\boldsymbol{V}_{*i}|) - \frac{1}{2}\log\left|\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{V}_{*i}^{-1}\boldsymbol{X}_i\right| \\ & - \frac{1}{2\sigma_*^2}\sum_{i=1}^{m}\left(\boldsymbol{w}_i^{(\lambda)} - \boldsymbol{X}_i\boldsymbol{\beta}_*\right)'\boldsymbol{V}_{*i}^{-1}\left(\boldsymbol{w}_i^{(\lambda)} - \boldsymbol{X}_i\boldsymbol{\beta}_*\right). \end{aligned} \tag{10}$$

Here, we do not assume that $\boldsymbol{D}_+$ is symmetric, and let $\gamma_+ = vec(\boldsymbol{D}_+)$. Again, following Demidenko (2004, p. 85) and Lindstrom and Bates (1988), we solve for the negative Hessian matrix of the residual log-likelihood function (10) in terms of the $\{p+1+k^2+1\} \times 1$ vector $\boldsymbol{\eta}_* = (\boldsymbol{\beta}_*', \sigma_*^2, \gamma_+', \lambda)'$:

$$\boldsymbol{H}_* = \mathcal{I}(\boldsymbol{\eta}_*) = \begin{pmatrix} -\frac{\partial^2 l_R}{\partial\boldsymbol{\beta}_*\partial\boldsymbol{\beta}_*'} & -\frac{\partial^2 l_R}{\partial\boldsymbol{\beta}_*\partial\sigma_*^2} & -\frac{\partial^2 l_R}{\partial\boldsymbol{\beta}_*\partial\gamma_+'} & -\frac{\partial^2 l_R}{\partial\boldsymbol{\beta}_*\partial\lambda} \\ -\frac{\partial^2 l_R}{\partial\sigma_*^2\partial\boldsymbol{\beta}_*'} & -\frac{\partial^2 l_R}{\partial\sigma_*^2\partial\sigma_*^2} & -\frac{\partial^2 l_R}{\partial\sigma_*^2\partial\gamma_+'} & -\frac{\partial^2 l_R}{\partial\sigma_*^2\partial\lambda} \\ -\frac{\partial^2 l_R}{\partial\gamma_+\partial\boldsymbol{\beta}_*}' & -\frac{\partial^2 l_R}{\partial\gamma_+\partial\sigma_*^2} & -\frac{\partial^2 l_R}{\partial\gamma_+\partial\gamma_+'} & -\frac{\partial^2 l_R}{\partial\gamma_+\partial\lambda} \\ -\frac{\partial^2 l_R}{\partial\lambda\partial\boldsymbol{\beta}_*'} & -\frac{\partial^2 l_R}{\partial\lambda\partial\sigma_*^2} & -\frac{\partial^2 l_R}{\partial\lambda\partial\gamma_+'} & -\frac{\partial^2 l_R}{\partial\lambda^2} \end{pmatrix}. \tag{11}$$

The following terms are derived for later use:

$$\begin{aligned} \frac{\partial\boldsymbol{w}_i^{(\lambda)}}{\partial\lambda} &= \lambda^{-1}\tilde{y}^{1-\lambda}\boldsymbol{s}_i^{(\lambda)} - [\lambda^{-1} + \log(\tilde{y})]\boldsymbol{w}_i^{(\lambda)} \\ \frac{\partial^2\boldsymbol{w}_i^{(\lambda)}}{\partial\lambda^2} &= \lambda^{-1}\tilde{y}^{1-\lambda}\left[\boldsymbol{t}_i^{(\lambda)} - 2\{\lambda^{-1} + \log(\tilde{y})\}\boldsymbol{s}_i^{(\lambda)}\right] + \left[\{\lambda^{-1} + \log(\tilde{y})\}^2 + \lambda^{-2}\right]\boldsymbol{w}_i^{(\lambda)}, \end{aligned}$$

References

Arellano-Valle, R. B., Bolfarine, H., Lachos, V. H. (2005). Skew-normal linear mixed models. *J. Data Sci.* 3:415–438.

Bickel, P. J., Doksum, K. A. (1981). An analysis of transformations revisited. *J. Amer. Statist. Assoc.* 76:296–311.

Box, G. E., Cox, D. R. (1964). An analysis of transformations. *J. Roy. Statist. Soc. Ser. B* 26:211–252.

Butler, S. M., Louis, T. A. (1992). Random effects models with non-parametric priors. *Statist. Med.* 11:1981–2000.

Demidenko, E. (2004). *Mixed Models: Theory and Application*. New York: Wiley and Sons.

Dodd, K. W., Guenther, P. M., Freedman, L. S., Subar, A. F., Kipnis, V., Midthune, D., Tooze, J. A., Krebs-Smith, S. M. (2006). Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. *J. Amer. Dietetic Assoc.* 106:1640–1650.

Dreyfus, C. (1975). Neurophysiological studies in human prematures after 32 weeks of conceptional age. *Biol. Psych.* 10:485–496.

Duan, N. (1983). Smearing estimate: a nonparametric retransformation method. *J. Amer. Statist. Assoc.* 78:605–610.

Gertner, S., Greenbaum, C. W., Sadeh, A., Dolfin, Z., Sirota, L., Ben-Nun, Y. (2002). Sleep-wake patterns in preterm infants and 6 month's home environment: implications for early cognitive developement. *Early Hum. Develop.* 68:93–102.

Gurka, M. J., Edwards, L. J., Muller, K. E., Kupper, L. L. (2006). Extending the Box-Cox transformation to the linear mixed model. *J. Roy. Statist. Soc. Ser. A* 169:273–288.

Gurka, M. J., Edwards, L.J., Nylander-French, L. (2007). Testing transformations for the linear mixed model. *Computati. Statist. Data Anal.* 51:4297–4307.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* 72:320–338.

Hinkley, D. V., Runger, G. (1984). The analysis of transformed data (with discussion). *J. Amer. Statist. Assoc.* 79:302–320.

Holditch-Davis, D. (1990). The development of sleeping and waking states in high-risk preterm infants. *Infant Behav. Develop.* 13:513–531.

Holditch-Davis, D., Edwards, L. J. (1998). Modeling development of sleep-wake behaviors. II. Results of two cohorts of preterms. *Physiol. Behav.* 63:319–328.

Holditch-Davis, D., Edwards, L. J. (2005). Prediction of three-year developmental outcomes from sleep development over the preterm period. *Infant Behav. Develop.* 28:118–131.

Kenward, M. G., Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53:983–997.

Korner, A. F., Brown, B. W., Reade, E. P., Stevenson, D. K., Fernbach, S. A., Thom, V. A. (1988). State behavior of preterm infants as a function of development, individual and sex differences. *Infant Behav. Develop.* 11:111–124.

Laird, N. M., Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* 38:963–974.

Lindstrom, M. J., Bates, D. M. (1988). Newton–Rapson and EM algorithms for linear mixed-effects models for repeated-measures data. *J. Am. Stat. Assoc.* 83:1014–1022.

Lipsitz, S. R., Ibrahim, J., Molenberghs, G. (2000). Using a Box-Cox transformation in the analysis of longitudinal data with incomplete responses. *Appl. Statist.* 49:287–296.

Parmelee, A. H., Wenner, W. H., Akiyama, Y., Schultz, M., Stern, E. (1967). Sleep states in premature infants. *Dev. Med. Child Neurol.* 9:70–77.

Spitzer, J. J. (1978). A Monte Carlo investigation of the Box-Cox transformation in small samples. *J. Amer. Statist. Assoc.* 73:488–495.

Spitzer, J. J. (1982). A primer on Box-Cox estimation. *Rev. Econ. Statist.* 64:307–313.

Tooze, J. A., Midthune, D., Dodd, K. W., Freedman, L. S., Krebs-Smith, S. M., Subar, A. F., Guenther, P. M., Carroll, R. J., Kipnis, V. (2006). A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *J. Amer. Dietetic Assoc.* 106:1575–1587.

Verbeke, G., Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *J. Amer. Statist. Assoc.* 91:217–221.

Verbeke, G., Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computat. Statist. Data Anal.* 23:541–556.

Verbeke, G., Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.

Vonesh, E. F., Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker.

Whitney, M. P., Thoman, E. B. (1993). Early sleep patterns of premature infants are differentially related to later developmental disabilities. *J. Dev. Behav. Pediatr.* 14:71–80.

Zhang, D., Davidian M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics* 57:795–802.