<u>Topics for these notes</u>:

- *Notation, model assumptions and comments*
- *LMM's versus GLM's*
- *Some simpler LMMs discussed in 6612*
- *Important distributions in the LMM*
- *General parameter estimation in the LMM; ML and REML*

<u>Associated reading</u>: The LMM course notes (early chapters).

Notation, model assumptions and comments

- Considering longitudinal data collected on subjects, there are 3 basic ways that linear mixed models can be expressed:
  - subject-time level
  - subject level
  - complete data level

- The mixed model at the subject-time level is useful when you have defined the particular experiment and variables. For example, most of the models written in the notes up to this point are explicitly defined mixed models expressed at the subject-time level, with response Y_{ij} (i denoting subject, j denoting time).

- We can write a more general mixed model in terms of subject data:

$$\underset{r_i\times 1}{\mathbf{Y}_i} = \underset{r_i\times p}{\mathbf{X}_i}\,\underset{p\times 1}{\boldsymbol{\beta}} + \underset{r_i\times q}{\mathbf{Z}_i}\,\underset{q\times 1}{\mathbf{b}_i} + \underset{r_i\times 1}{\boldsymbol{\varepsilon}_i}, \qquad \text{for subjects } i=1,\ldots,n.$$

  - ○ $\mathbf{Y}_i$ are the $r_i\times 1$ responses for subject i
  - ○ $\mathbf{X}_i$ is the matrix of known covariates associated with fixed effects
  - ○ $\boldsymbol{\beta}$ are the $p\times 1$ fixed effects
  - ○ $\mathbf{Z}_i$ is the matrix of known covariates associated with the random effects
  - ○ $\boldsymbol{\varepsilon}_i$ is the residual error vector.

- We index **X** and **Z** by subject even when they may be the same across subjects, in order to identify the size of the matrices. We will keep **X** and **Z** without indices to denote the full-data versions of these matrices, which will be defined shortly.

- For the model above, we assume $\underset{q\times 1}{\mathbf{b}_i} \sim iid\,\mathrm{N}\left[\underset{q\times 1}{\mathbf{0}}, \underset{q\times q}{\mathbf{G}_i}\right]$ and $\underset{r_i\times 1}{\boldsymbol{\varepsilon}_i} \sim iid\,\mathrm{N}\left[\underset{r_i\times 1}{\mathbf{0}}, \underset{r_i\times r_i}{\mathbf{R}_i}\right]$, and that these random vectors are independent.

- In addition, subjects themselves are assumed to be independent of each other. However, in cases where subjects are not independent, we can work this into the model by defining appropriate cluster units, which will be discussed later.

- Generally speaking, $\mathbf{G}_i$ will be used to account for variability between subjects and $\mathbf{R}_i$ will be used to account for covariances between repeated measures within subjects. However, it will also be demonstrated that there are many ways to model correlated data that combine $\mathbf{G}_i$ and $\mathbf{R}_i$.

- The subject models can be combined into one 'complete-data' model by essentially stacking the *n* subject-specific models:

$$\underbrace{\begin{pmatrix} \underset{r_1\times 1}{\mathbf{Y}_1} \\ \underset{r_2\times 1}{\mathbf{Y}_2} \\ \vdots \\ \underset{r_n\times 1}{\mathbf{Y}_n} \end{pmatrix}}_{r_{tot}\times 1} = \underbrace{\begin{pmatrix} \underset{r_1\times p}{\mathbf{X}_1} \\ \underset{r_2\times p}{\mathbf{X}_2} \\ \vdots \\ \underset{r_n\times p}{\mathbf{X}_n} \end{pmatrix}}_{r_{tot}\times p} \underset{p\times 1}{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} \underset{r_1\times q}{\mathbf{Z}_1} & \underset{r_1\times q}{\mathbf{0}} & \cdots & \underset{r_1\times q}{\mathbf{0}} \\ \underset{r_2\times q}{\mathbf{0}} & \underset{r_2\times q}{\mathbf{Z}_2} & & \underset{r_2\times q}{\mathbf{0}} \\ \vdots & & \ddots & \\ \underset{r_n\times q}{\mathbf{0}} & \underset{r_n\times q}{\mathbf{0}} & & \underset{r_n\times q}{\mathbf{Z}_n} \end{pmatrix}}_{r_{tot}\times q_{tot}} \underbrace{\begin{pmatrix} \underset{q\times 1}{\mathbf{b}_1} \\ \underset{q\times 1}{\mathbf{b}_2} \\ \vdots \\ \underset{q\times 1}{\mathbf{b}_n} \end{pmatrix}}_{q_{tot}\times 1} + \underbrace{\begin{pmatrix} \underset{r_1\times 1}{\boldsymbol{\varepsilon}_1} \\ \underset{r_2\times 1}{\boldsymbol{\varepsilon}_2} \\ \vdots \\ \underset{r_n\times 1}{\boldsymbol{\varepsilon}_n} \end{pmatrix}}_{r_{tot}\times 1},$$

or more succinctly, $\underset{r_{tot}\times 1}{\mathbf{Y}} = \underset{r_{tot}\times p}{\mathbf{X}} \underset{p\times 1}{\boldsymbol{\beta}} + \underset{r_{tot}\times q_{tot}}{\mathbf{Z}} \underset{q_{tot}\times 1}{\mathbf{b}} + \underset{r_{tot}\times 1}{\boldsymbol{\varepsilon}}$,

where $\begin{pmatrix} \underset{q_{tot}\times 1}{\mathbf{b}} \\ \underset{r_{tot}\times 1}{\boldsymbol{\varepsilon}} \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \underset{q_{tot}\times 1}{\mathbf{0}} \\ \underset{r_{tot}\times 1}{\mathbf{0}} \end{pmatrix}, \begin{pmatrix} \underset{q_{tot}\times q_{tot}}{\mathbf{G}} & \underset{q_{tot}\times r_{tot}}{\mathbf{0}} \\ \underset{r_{tot}\times q_{tot}}{\mathbf{0}} & \underset{r_{tot}\times r_{tot}}{\mathbf{R}} \end{pmatrix}\right]$, $q_{tot} = nq$, $r_{tot} = \Sigma\, r_i$,

$\underset{q_{tot}\times q_{tot}}{\mathbf{G}} = \operatorname*{diag}_{i=1}^{n}\left\{\underset{q\times q}{\mathbf{G}_i}\right\}$ and $\underset{r_{tot}\times r_{tot}}{\mathbf{R}} = \operatorname*{diag}_{i=1}^{n}\left\{\underset{r_i\times r_i}{\mathbf{R}_i}\right\}$.

- Note that $\mathbf{R}_i$ will often differ between subjects due to different numbers of repeated measures (although the underlying parameters are usually the same).

- Even when $\mathbf{R}_i$ or $\mathbf{G}_i$ are the same across subjects (this is usually the case for $\mathbf{G}_i$), we keep the subscript i since $\mathbf{R}$ and $\mathbf{G}$ are used for complete data form. [When $\mathbf{R}_i$ does differ between subjects due to missing data, we will later discuss how we can keep dimensions of $\mathbf{R}_i$ the same across subjects and just partition the matrix into 'observed' and 'missing' pieces.]

- When $\mathbf{G}_i$ is the same across subjects, note that $\underset{q_{tot}\times q_{tot}}{\mathbf{G}} = \underset{n\times n}{\mathbf{I}} \otimes \underset{q\times q}{\mathbf{G}_i}$, where '$\otimes$' denotes the Kronecker product. Generally, for an m×n matrix **A** and p×q matrix **B**, the Kronecker product is defined as

$$\mathbf{A}\otimes\mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & & a_{2n}\mathbf{B} \\ \vdots & & \ddots & \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & & a_{mn}\mathbf{B} \end{pmatrix}$$

- The normal distribution assumption of the random effects is common. There have been methodological developments to account for non-normal random effects by considering mixtures of normals (which can yield quite a variety of distributions).

- In fitting a linear mixed model with SAS, PROC MIXED, the RANDOM statement is used to specify **Z** and **G**, while the REPEATED statement is used to specify **R**. When a REPEATED statement is not included, the model will use $\mathbf{R}_i = \mathbf{I}_{r_i}\sigma_\varepsilon^2$ (the independent structure).

- In modeling a random intercept term by subject, we discussed how the following approaches were essentially equivalent (however, see course notes for differences in computation between these approaches):
  - random intercept / subject=id;
  - random id;

- You can add the option 'g' after the slash in the RANDOM statement to get the form and fit for what SAS calls '**G**'.

- The notation for mixed models varies from text to text. We will use $\boldsymbol{\beta}$ to denote the set of regression coefficients. We can denote the set of all covariance parameters in the covariance matrix of $\mathbf{Y}_i$ (Var($\mathbf{Y}_i$)=$\mathbf{V}_i$) as $\boldsymbol{\alpha}$. Collectively, $\boldsymbol{\theta}$=($\boldsymbol{\alpha}$, $\boldsymbol{\beta}$) is the set of all parameters in a particular mixed model. The variance function $\mathbf{V}_i$ is often written as $\mathbf{V}_i(\boldsymbol{\alpha})$ to indicate that all parameters in the matrix involve $\boldsymbol{\alpha}$.

- We will typically use $\mathbf{b}_i$ for random effects. When we have only a random intercept, we might call it b_i. If there is a random intercept and slope, we can use $\mathbf{b}_i =(b_{0i},b_{1i})^t$ (first element for intercept, 2nd for slope).

- We use $\mathbf{R}_i$ for the 'within-subject' covariance matrix and $\mathbf{G}_i$ for the covariance matrix of random effects that expresses 'between-subject' variability. But note that both of these matrices impact $Var(\mathbf{Y}_i)$, which is the covariance matrix for the responses [you could also call it $Cov(\mathbf{Y}_i)$].

A simple way to account for correlation in an LMM is to add a 'random intercept':

- The basic model is $$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 x_{1ij} + ... + \beta_{p-1} x_{p-1,ij} + b_i + \varepsilon_{ij} \\ &= \mathbf{X}_{ij}\boldsymbol{\beta} + b_i + \varepsilon_{ij}\ , \end{aligned}$$

  where *Y* is the outcome, $\mathbf{X}_{ij} = (x_{1\mathrm{ij}},\ldots,x_{p-1,\mathrm{ij}})$ is a row vector of predictors, both for subject *i* at time *j*, and where $\varepsilon_{ij} \sim N(0,\sigma_\varepsilon^2)$ and $b_i \sim N(0,\sigma_b^2)$. These random terms are assumed to be independent of each other.

- The main element that distinguishes this from a general linear model is the addition of the random term b_i. We also use subject and time indices here, with the addition of the repeated measures.

- What is $Var(\mathbf{Y}_i)$?

A special case of the LMM: a GLM!

- A linear mixed model with no random effects and a simple error covariance structure ($\mathbf{R}=\sigma^2\mathbf{I}$) is a general linear model.

- For cross-sectional data without correlation, we can fit using SAS PROC GLM or the lm function in R. But using LMM software will also work! (Like PROC MIXED.)

- For GLM $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$, the least squares estimator of $\boldsymbol{\beta}$ is $(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$. To show, apply calculus to matrix quantities (see lecture notes). This is also the MLE of $\boldsymbol{\beta}$. The MLE of σ^2 is biased!

- Examples of GLM data
  - Myostatin data
  - Mouse and smoke data

LMM's versus GLM's

- For many simpler models, GLM and LMM modeling approaches will be the same (since GLM's are special cases of LMM's).

- Treatment of predictors in LMMs and GLMs are generally the same. Writing ESTIMATE and CONTRAST statements are similar for GLMs and LMMs.

- Inference in the GLM versus LMM
  - Both commonly us ML (or REML) to estimate parameters; Wald (*t*) and LRT (*F*) tests are used for testing hypotheses.
  - In the GLM, *F*-tests come via the ANOVA table (algebraic).
  - In the LMM, we identify quantities that have approximate t or F distributions, then calibrate the test by estimating the appropriate degrees of freedom.

- There are several (D)DF estimation methods: containment, between-within, residual, Satterthwaite, Kenward-Roger. Depending on the model, some methods may produce the same (D)DF. One of the key issues is accurately estimating the true distribution of the test statistic under the null hypothesis.

- In SAS, there are default (D)DF estimation approaches depending on whether you have a RANDOM or REPEATED statement (or both). You can specify the method you want, or even just numerically specify the DF. Recall that both t and F distributions are indexed by (D)DF.

- For more details, see Verbeke and Molenberghs, Linear Mixed Models in Practice, Springer, 1997, Appendix A, and also the SAS Help Documentation.

Some simpler LMMs discussed in 6612

- Simple random intercept models (fixed effects plus one random intercept for subjects)
    - 1-sample
    - Multi-sample

- For the simple random intercept model, tests for predictors can also be conducted via the GLM, using 'repeated measures ANOVA'.

- Model with crossed random effects
    - Side question: when to make an effect fixed or random?
    - Judge data

- Please review the course notes.

Important distributions in the LMM

- The conditional distribution of **Y** given the random effects **b**

$$\mathbf{Y}\,|\,\mathbf{b} \sim \mathbf{N}\left[\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \mathbf{R}\right] = \mathbf{N}\left[\begin{pmatrix}\mathbf{X} & \mathbf{Z}\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta} \\ \mathbf{b}\end{pmatrix}, \mathbf{R}\right]$$

- The classical method to analyze longitudinal data, “RM ANOVA”, essentially makes inference using this conditional distribution, since the random effects are treated as fixed effects. Adjustments are then made to tests in order to make ‘correct’ inference for estimators that account for the clustered data. In some cases this approach may yield the same or similar results as fitting a linear mixed model, but generally is much more limited.

- The conditional distribution is used to conduct inference for random effects using standard LMM methods (i.e., the ‘Laird and Ware’ approach).

- The marginal distribution of **Y**
  - The joint distribution of **Y** and **b** is

$$\begin{pmatrix}\mathbf{Y}\\ \mathbf{b}\end{pmatrix} \sim N\left[\begin{pmatrix}\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{b}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{Z}\mathbf{G}\mathbf{Z}^t+\mathbf{R} & \mathbf{Z}\mathbf{G}\\ \mathbf{G}^t\mathbf{Z}^t & \mathbf{G}\end{pmatrix}\right].$$

  - The marginal distribution of **Y** can be obtained by integrating out the random effects **b** from the joint distribution to obtain $\mathbf{Y} \sim N\left[\mathbf{X}\boldsymbol{\beta}, \mathbf{V}=\mathbf{Z}\mathbf{G}\mathbf{Z}^t+\mathbf{R}\right]$.
  - The marginal distribution is used in the likelihood functions that are used to estimate parameters in the LMM.

- Modern mixed model methodology maximizes the likelihood associated with **Y**, or it equivalently minimizes

$$\ell = -2ln(L) = r_{tot}\ ln\left(2\pi\right) + ln\left|\mathbf{V}\right| + \left(\mathbf{y}-\mathbf{X\beta}\right)^{t}\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X\beta}\right)$$

  in order to make inferences about the regression coefficients **β** and covariance parameters **α**.

- The likelihood is also often expressed as a combination of subject-specific components:

$$L\left(\theta\right) = \prod_{i=1}^{n}\{\left(2\pi\right)^{-r_i/2}|\mathbf{V}_i\left(\boldsymbol{\alpha}\right)|^{-1/2}e^{-\left(\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}\right)^{t}\mathbf{V}_i^{-1}\left(\mathbf{Y}_i-\mathbf{X}\boldsymbol{\beta}\right)/2}\}$$

$$\Rightarrow \ell = -2ln(L) = \sum_{i=1}^{n} r_i ln\left(2\pi\right) + \sum_{i=1}^{n} ln\left|\mathbf{V}_i\left(\boldsymbol{\alpha}\right)\right| + \sum_{i=1}^{n}\left(\mathbf{Y}_i-\mathbf{X}\boldsymbol{\beta}\right)^{t}\mathbf{V}_i^{-1}\left(\boldsymbol{\alpha}\right)\left(\mathbf{Y}_i-\mathbf{X}\boldsymbol{\beta}\right)$$

General parameter estimation in the mixed model

- For the standard GLM, there are the regression coefficients (**β**) and one covariance parameter (σ^2) to estimate, which can be carried out using matrix algebra.

- Due to the inclusion of more covariance parameters in the model (in either **G** or **R**), parameter estimation in the mixed model is not as straightforward and generally requires at least some numerical analysis.

- Before describing these techniques in more detail, we will first discuss the most common estimation approaches, *maximum likelihood (ML)* estimation and *restricted maximum likelihood (REML)* estimation. There is also the *MIVQUE0* estimation approach, which is seldom used.

Maximum Likelihood (ML) Estimation

- The ML estimators of $\boldsymbol{\beta}$ are obtained by maximizing the likelihood L or minimizing ℓ (both given on previous page) based on the marginal distribution of $\mathbf{Y}$. This can be accomplished by first noting that maximizing the likelihood with respect to $\boldsymbol{\beta}$, conditional on $\boldsymbol{\alpha}$, yields

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^{n} \mathbf{X}_i^t \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \mathbf{X}_i \right)^{-} \sum_{i=1}^{n} \mathbf{X}_i^t \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \mathbf{Y}_i \quad \text{(subject-specific form)}$$

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X} \right)^{-} \mathbf{X}^t \mathbf{V}^{-1} \mathbf{Y} \quad \text{(complete-data form)} \qquad (1)$$

where $\mathbf{V}_i = \mathrm{Var}(\mathbf{Y}_i) = \mathbf{Z}_i \mathbf{G}_i \mathbf{Z}_i^t + \mathbf{R}_i$ (subject-specific form).

- Notice that we need values of $\boldsymbol{\alpha}$ in order to solve (1). To accomplish this, we can replace $\boldsymbol{\beta}$ in the likelihood function with its MLE in (1). Now we have a likelihood expressed in terms of $\boldsymbol{\alpha}$ only. Such a likelihood is sometimes referred to as a *profile likelihood*.

- Now we can maximize the profile likelihood function in order to obtain $\hat{\boldsymbol{\alpha}}$ using a numerical technique such as a ridge-stabilized Newton-Raphson algorithm (common in SAS). We can then go back and determine $\hat{\boldsymbol{\beta}}$ using (1) by replacing $\boldsymbol{\alpha}$ with its ML estimates.

- The MLE solution we obtain with this approach is the same as what we would obtain if we were able to maximize the likelihood simultaneously with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Notice that the estimator of $\boldsymbol{\beta}$ in (1) is identical to the weighted least-squares estimates with $\mathbf{V}^{-1}$ as the weighting matrix.

- One drawback of ML estimation is that associated estimators of covariance parameters tend to be biased. REML offers one way to remove or reduce bias.

- Note that (1) uses a generalized inverse in case $\mathbf{X}$ does not have full rank. If $\mathbf{X}$ does have full rank, then we can replace $\left(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}\right)^{-}$ with $\left(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}\right)^{-1}$. Issues of model parameterization and estimation here are analogous to those discussed in the GLM review.

Restricted maximum likelihood (REML) estimation

- To first introduce REML estimation, consider estimating the population variance based on a random sample from the population of interest.

- We know the sample variance (s^2, which uses 'n–1' in the denominator) is unbiased for the population variance in the case that the population mean is unknown and estimated (i.e., the usual case).

- But the ML estimator of σ^2 has n in the denominator. This demonstrates that ML estimates may not necessarily be unbiased estimators. REML estimation offers an alternative to ML estimation which helps to circumvent this problem. [Note: some call s^2 the adjusted MLE estimator.]

REML estimation for σ^2

- Let $\mathbf{J}_n = \underset{n\times 1}{\mathbf{J}}$, $\mathbf{I}_n = \underset{n\times n}{\mathbf{I}}$, and let $\mathbf{Y} = \left(Y_1,\ Y_2,\ \ldots,\ Y_n\right)^t$, where $\mathbf{Y} \sim N\left(\mu\,\mathbf{J}_n,\ \sigma^2\mathbf{I}_n\right)$

- Let **A** = any matrix with *n*–1 independent columns orthogonal to **J***ₙ*. E.g.:

$$\mathbf{A}=\begin{pmatrix} 1 & 0 & ... & 0 \\ -1 & 1 & ... & ... \\ 0 & -1 & ... & ... \\ ... & 0 & ... & 0 \\ ... & ... & ... & 1 \\ 0 & 0 & & -1 \end{pmatrix}$$

- Let $\mathbf{U} = \mathbf{A}^t\mathbf{Y}$ be “error contrasts.” Note that $\mathbf{U} \sim N\left(\mathbf{0},\ \sigma^2\mathbf{A}^t\mathbf{A}\right)$ and that σ^2 is the only parameter in the distribution for **U**. Maximizing the likelihood for **U** with respect to σ^2 yields: $\hat{\sigma}^2 = [\mathbf{Y}^t\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t\mathbf{Y}]/(n-1){=}s^2$.

- In a similar fashion, it can be shown that the REML estimator of σ^2 in a GLM is $[1/(n-k)]\mathbf{Y}^t\left(\mathbf{I}-\mathbf{P}_\mathbf{X}\right)\mathbf{Y}$. Can you do this?

REML estimation in the linear mixed model

- Let **A** be a full rank matrix with columns orthogonal to the columns of **X**. Then $\mathbf{U}=\mathbf{A}^t\mathbf{Y}\sim N\left(\mathbf{0},\ \mathbf{A}^t\mathbf{V}\left(\boldsymbol{\alpha}\right)\mathbf{A}\right)$, which does not depend on $\boldsymbol{\beta}$. The associated likelihood is

$$L=\left(2\pi\right)^{-(r_{tot}-k)/2}\left|\sum_{i=1}^{n}\mathbf{X}_i^t\mathbf{X}_i\right|^{1/2}\left|\sum_{i=1}^{n}\mathbf{X}_i^t\mathbf{V}_i^{-1}\mathbf{X}_i\right|^{-1/2}\prod_{i=1}^{n}\left|\mathbf{V}_i\right|^{-1/2}e^{-1/2\sum_{i=1}^{n}(\mathbf{Y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}})^t\mathbf{V}_i^{-1}(\mathbf{Y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}})},$$

  where *k*=*rank*(**X**).

- Note that this restricted *L* does not involve $\boldsymbol{\beta}$ parameters ($\hat{\boldsymbol{\beta}}$ is a function of $\boldsymbol{\alpha}$, as are $\mathbf{V}_i$ matrices) and is not a profile likelihood, as before. This is why some software (e.g., SAS) does not penalize for β terms in the AIC.

- The restricted likelihood can be maximized to yield $\hat{\boldsymbol{\alpha}}$. The problem is that this method really only offers a way to estimate parameters in $\boldsymbol{\alpha}$, not $\boldsymbol{\beta}$.

- The common approach to estimate $\boldsymbol{\beta}$ is then to plug the REML estimators of $\boldsymbol{\alpha}$ back into equation (1). But equation (1) was derived using ML methods, so this estimation of $\boldsymbol{\beta}$ is really based on a hybrid of ML and REML methods. Specifically, estimators of $\boldsymbol{\beta}$ use the ML form, but employ REML estimators of the variance components in that form.

- Verbeke denotes these as "REML" estimators of $\boldsymbol{\beta}$ (quotes emphasized). Since estimation of $\boldsymbol{\beta}$ is not based on one clear method, some statisticians prefer ML estimation. On the other hand, this estimation method does offer a way to reduce bias in variance component estimators. Some might argue that this is more important than the methodological issue.

Choosing the estimation method in SAS

- In the PROC MIXED statement, an option can be added: method = ML <or> REML <or> MIVQUE0 (no slash to separate the method option from the rest of the statement) if ML estimates are of interest. Note that the default method is REML; if no option is specified, then REML will be used.

Properties of estimators in the LMM: BLUE, BLUP, EBLUE, EBLUP