Classification of Distinct Trajectories in Longitudinal Data with Irregular Spaced Intervals: Heterogeneous Linear Mixed Model Vs Mixture Modeling of BLUPs from Linear Mixed Model

Md Jobayer Hossain[1,2], Benjamin E. Leiby[3]

[1]Biostatistics Core, Department of Biomedical Research, Nemours/Alfred I. duPont Hospital for Children, 1600 Rockland Road, Wilmington, DE 19803

[2]Department of Applied Economics and Statistics, University of Delaware, 213 Townsend Hall, Newark, DE 19716

[3]Division of Biostatistics, Thomas Jefferson University1015 Chestnut Street, Philadelphia, PA 19107

Corresponding:

Md Jobayer Hossain
Biostatistics Core, Department of Biomedical Research, Nemours/Alfred I. duPont Hospital for Children, 1600 Rockland Road, Wilmington, DE 19803
Email: jhossain@nemours.org

Abstract

Classifying heterogeneous trajectories into distinct groups has gained huge appeal in this precision medicine era when massive amounts of naturally occurring longitudinal data can be used to derive evidence-based knowledge for facilitating precise diagnosis, prevention, and tailored treatment. Methods on the topic are sparse. Mixed models with a mixture distribution of random effects (HLME) classify trajectories through estimating profiles and computing posterior probabilities of belonging to a class. Two forms of the method are implemented in R and M-plus among standard statistical software. The method becomes computationally complex with increased data size and level of unbalancedness. It also requires specification of the number of classes prior to analysis. Likelihood function may have multiple local maxima. We introduced a flexible approach that applies mixture models to empirical BLUPs from linear/ spline mixed models. The method reduces drawbacks of HLMEs and works well on large datasets. This study compares the classification ability of two methods using real and simulated datasets of complex temporal curves and identifies situations when one method outperforms the other.

Keywords: Classification; irregular spaced time; trajectories; linear mixed model ; heterogeneous; mixture distribution of random effects

1. **Background and Motivation**:

In the analysis of longitudinal data, recognizing subgroups of individuals with heterogeneous trajectories, identifying potential contributors to varying trends, and utilizing this data-driven knowledge to make decisions are areas of growing research interest in this data science era[1-2]. In recent research related to learning healthcare systems and personalized and precision medicine, massive amounts of naturally occurring longitudinal data are used to derive evidence-based knowledge by accounting for individual-level variations and recognizing distinct patterns of health and disease trajectories with the goal of facilitating precise diagnosis, appropriate strategies for prevention, and tailored treatment of health conditions[3-4].

While naturally occurring data such as electronic medical records have shown great utility in learning healthcare systems and personalized medical decisions, observation times in these datasets are often irregularly spaced and vary across individuals as patients make clinic visits idiosyncratically, thereby producing unique sequences of measurements across individuals. The repeated measures of this type are usually expressed as parametric or semiparametric functions of time and are described by levels and shapes of the curves. Linear mixed effects regressions model these data as a combination of population temporal trend that is shared by all individuals and subject-specific effects that describe how the trend over time of each individual differs from the population mean trend. The former are the fixed effects and the later are the random effects in the model. The random effects vary from one individual to another, thereby accounting for sources of heterogeneity in trajectories across individuals. The inclusion of time as random effects in the model allows us to express the covariance of repeated measures as a function of time. In the general form of the mixed effect model, random effects are assumed to be distributed as multivariate normal with mean 0 and constant variance-covariance matrix. This distributional assumption of the random effects implies that individual-level trajectories are homogeneous in shapes, thus the model under this assumption can be termed as a homogeneous linear mixed effects model. If the dataset contains heterogeneous individual-level trajectories, the above homogeneity assumption could be violated. With this violation, averaging over curves of heterogeneous shapes may result in missing important features of trajectories for individuals across heterogeneous subgroups causing misrepresentation of model fit. Therefore, it is important to identify potential hidden subgroups of individuals with distinct trajectories that may exist in a dataset. This requires appropriate capturing of the heterogeneity in levels and shapes of trajectories across individuals.

Verbeke et al (1996) introduced a heterogeneous linear mixed effects model where random coefficients were assumed to have a mixture of normals distribution. The method was implemented in a SAS macro, but the EM algorithm used was found to be computationally expensive and failed to provide good convergence criteria and direct estimates of the variance of the parameters. Proust-Lima et al (2017) recently extended the model of Verbeke et al by expressing mixture component-specific fixed and random effects. This extended model was implemented in R package ‘lcmm’. The package used Marquadrt algorithm in order to minimize the EM-algorithm related limitations found in Verbeke’s method. Immediately following Verbeke et al, Muthen et al (1999) introduced a similar method that combines multilevel mixed effects and mixture models and implemented the software Mplus. Muthen’s method gained wide popularity as the growth mixture model (GMM) and Mplus has been widely used, especially in the psychometric and socio-behavioral studies, for over a decade in classification of longitudinal data. Both of these methods are built on the same concept: specifically, the combination of linear mixed effects and mixture models and both methods have some inherent limitations. The method becomes mathematically and computationally complex fairly quickly. Computation time increases with sample size, degree of unbalancedness, complexity of the parametric curve, and number of random effects. The method requires pre-specification of the number of subgroups and typically uses information criteria (e.g., BIC) for the selection of number of classes, but definitive determination is exploratory. Log-likelihood functions may have local maxima; therefore, a careful choice of the initial values is crucial for ensuring convergence toward the global maxima. In Mplus, the programming structure lends itself to the situation of limited number of repeated observations per subject at a common set of measurement occasions across individuals. For computational feasibility, observations may need to be thinned and aligned to a common set of time points. In the R lcmm package, the algorithm may reach the highest number of iterations without convergence. Another serious limitation is that the methods are unavailable in mainstream statistical software packages other than R and Mplus. Taken together, it seems that although theoretically sound, methods implemented in R and Mplus do not fully resolve the problem of clustering unbalanced longitudinal data because of the computational complexities. In a recent application, we have determined that application of post-hoc finite mixture models to the empirical best linear unbiased predictor (eBLUP) from linear or piecewise linear mixed effects model can reasonably classify heterogeneous trajectories of hidden components in distinct subgroups. Theoretical ground of this functionality is obvious as vectors of BLUPs account for the heterogeneities in shapes across individual-level trajectories, and mixture models classify individuals based on these heterogeneities. This study uses 5 datasets of early childhood growth patterns consisting of 2-3 components of linear, quadratic and cubic trends with varying level of separability to compare the classification and evaluation performance of these three methods. The study would also identify the situations when one method performs better than others.

2. Overview of heterogeneous linear mixed effects model: The general form of the linear mixed effects model is, $Y_i = X_i\beta + Z_i b_i + \epsilon_i,\ i = 1,2,..,N$; where, β is the vector of fixed effects that describes the shapes of average trajectories over all individuals under study; $b_i \sim MN(0, D)$, is the vector of random coefficients related to the ith subject; $\epsilon_i \sim MN(0, R_i)$, is the vector of measurement or sampling errors associated with the responses of the ith subject. Vectors b_i and ϵ_i are assumed to be independent. The distributional assumption of b_i implies that individual-level trajectories are homogeneous in shapes, thus the model under this assumption is termed as homogeneous linear mixed effects model. Under this form of the model, $Y_i \sim MN(X_i\beta,\ Z'_i D Z_i + \sigma^2 I = V_i)$ and the empirical best linear unbiased predictor (BLUP) of random coefficients for given data (Y_i) is, $\widehat{b_\iota} = \hat{E}(b_i|Y_i) = \widehat{D}\, Z'_i \hat{V}_i^{-1}\left(Y_i - X_i\hat{\beta}\right)$. This normality assumption of $b_i, i = 1,2,..n$ could be violated in presence of the hidden subgroups or heterogeneous shapes of individual trajectories. In case of the presence of subgroups in shapes of trajectories, Verbeke et al (1996) suggested a Gaussian mixture distribution of b_i as $b_i \sim \sum_{g=1}^{G} \pi_g\, MN(\mu_g, D)$ with π_g is the component probability, $g = 1,2, \dots G$, μ_g is the component specific mean and D is the common variance-covariance matrix. Then the covariance matrix of b_i, $D^* = D + \sum_{g=1}^{G} \pi_g \mu_g \mu'_g - \sum_{l=1}^{G} \sum_{g=1}^{G} \pi_l \pi_g\, \mu_l \mu'_g$, and the distribution of $Y_i \sim \sum_{g=1}^{G} \pi_g\, MN\left(X_i\beta + Z_i\mu_g,\ V_i\right)$. It is obvious that a restriction of $\sum_{g=1}^{G} \pi_g \mu_g = 0$, makes $E(Y_i) = X_i\beta$ which is the mean trajectories under homogeneous model. Under the mixture model, the expression of the empirical BLUP becomes $\widehat{b_\iota} = E(\acute{b_\iota}|Y_\iota, \varphi) = \widehat{D}\, Z'_i \hat{V}_i^{-1}(Y_i - X_i\hat{\beta}) + A_i \sum_{g=1}^{G} \pi_{ig}\,(\varphi)\mu_g$; where $A_i = I - DZ_i' V_i^{-1} Z_i$; θ is the vector of parameters of β, σ, D and V_i, and $\pi_{ig} = \pi_{ig}(\varphi) =$

$\frac{\pi_g f(Y_i|\theta)}{\sum_{g=1}^{G}\pi_g\, f(Y_i|\theta)}$; φ'=(π', θ') is the posterior probability for the ith individual to belong to the gth component of the mixture. The second part of the expression of $\hat{b_i}$ is the correction term toward the component means, proportional to the posterior probability of belonging to each component. The model described above is referred to as the heterogeneous mixed effects model and has been used for the classification of longitudinal data (Verbeeke, 1996). Proust-Lima et al. (2017) defined an extension of this model in which both fixed and random effects can be mixture component specific with $b_i \sim \sum_{g=1}^{G}\pi_g\, MNV(\mu_g, D_g)$; $D_g = w_g D$, where w_g is the class-specific intensity of individual variability. They also replaced EM algorithm by Marquardt algorithm to improve computational efficiency and implemented to minimize the EM algorithm-related limitations; and implemented in the R package 'lcmm'.

Immediate following Verbeeke et al introduced to HLME model, Muthen et al (1999) expanded the concept through latent variable mixed effects models from the structural equation modeling approach and accommodated many linear and nonlinear mean functions over time. The method was implemented in Mplus and received popularity as growth mixture model (GMM).

For a simple linear growth curve implemented in Mplus, Muthen et al described the responses of ith individual who belongs to latent class $c_i = g$ as

$$Y_{ij}|_{c_i=g} = \eta_{0i} + \eta_{1i}t_{ij} + \epsilon_{ij}$$

$$\eta_{ki}|_{c_i=g} = \alpha_{kg} + \gamma_{kg}x_i + \xi_{ki}, k = 0, 1$$

where η_{0i} and η_{1i} are random intercepts and slopes, α_{0g} and $\alpha_{1g}$are the average intercept and slope of time varying variables, γ_{0g} and γ_{1g}intercepts and slopes of time invariant variables associated with latent class g, and c_i is the latent categorical random variable with probability of the unobserved class membership of the ith subject, $\Pr(c_i = g) = \pi_{ig}$. This probability follows the multinomial logistic regression with respect to time invariant covariates X_{ci} associated with ith subject as

$$\pi_{ig} = P(ci = g|X_{ci}) = \frac{e^{\gamma_{0g}+\, X_{ci}^T\gamma_{1g}}}{\sum_{l=1}^{G} e^{\gamma_{0l}+\, X_{ci}^T\gamma_{1l}}}\ , \pi_{ig} \geq 0.$$

To obtain the desired clustering given a pre-specified number of classes, subjects can be assigned to their most likely class based on the posterior probabilities of class membership as,

$$P(ci = g|X_{ci}, Y_i) = \frac{\pi_{ig}P(Y_i|c_i = g)}{\sum_{l=1}^{G}\pi_{il}P(Y_i|c_i = l)}.$$

Model parameters are estimated using an EM algorithm from the likelihood function of mixture distributions for a given pre-specified number of classes. [An overview of the model framework and estimation procedure is available in a book of Taylor & Francis Groups book's "Longitudinal Data Analysis" (edited by G Fitzmaurice, M Davidian, G Verbeke, G Molenberghs)

3. Post-hoc Mixture Modeling of BLUPs: This method is built on capturing the heterogeneity in individual trajectories through the empirical BLUP (eBLUP) from the fit of a suitable linear mixed effects model and then applying the conventional mixture model on the captured eBLUP as a post-hoc analysis. Fixed effects in the linear mixed effects model describe the shape of the average trajectories over all individuals under study and the random coefficients explains the heterogeneity between trends in average and individual trajectories. Thus, the random coefficients convey heterogeneities across trajectories of all individuals respective to the shape of average trajectory. For individuals with trajectories of similar shapes would be expected to have similar pattern in heterogeneities, and thereby could safely be used as a primary sources of information for classification of trajectories. Without loss of generality fixed effects can be ignored for this purpose. Under the general form of linear mixed effects model, the random coefficients $b_i, i = 1,2,..n,$ are normally distributed, $b_i \sim MN(0, D)$. If there exists G mixture components, then $b_i | g_i = g \sim MN(\gamma_g, D_g)$, and the marginal distribution of b_i can be given as $f(b_i) = \sum_{g=1}^{G} f(b_i, g_i = g) = \sum_{g=1}^{G} f(b_i | g_i = g) P(g_i = g)$ = $\sum_{g=1}^{G} \pi_k f(b_i | g_i = g) = \sum_{g=1}^{G} \pi_g MN(\gamma_g, D_g)$. The likelihood function of $\theta = (\gamma_1, \ldots \gamma_G,$ $D_1, \ldots D_G, \pi_1 \ldots \pi_{G-1})$ for given b_i is $L(\theta | b_1, b_2, \ldots. b_n) = \coprod_{i=1}^{n} \sum_{g=1}^{G} \pi_g MN(b_i; \gamma_g, D_g)$. In reality, we observe data Y_i, not b_i. However, we use the empirical BLUP of b_i for given Y_i. That is, we use $\widehat{b_i} = \hat{E}(b_i | Y_i) = \widehat{D} Z'_i \hat{V}_i^{-1}(Y_i - X_i \hat{\beta})$ as data for classification. In addition to the empirical BLUP, scores from suitable linear transformations of BLUP such as principal component analysis, factor analysis and canonical analysis can also be used for classification purpose. Transformations may be beneficial to pre-process BLUP for classification. It may be helpful to handle outliers also. Of course, the accuracy of the classification may depend on the accuracy of capturing heterogeneities between individual and average trajectories. Prediction error, $\widehat{Var}(\widehat{b_i} \; \text{-} b_i)$ $= \widehat{D} - \widehat{D} Z'_i \hat{V}_i^{-1} Z_i \widehat{D} + \widehat{D} Z'_i \hat{V}_i^{-1} X_i \left[\sum_{i=1}^{n} X_i' \hat{V}_i^{-1} X_i\right]^{-1} X_i' \hat{V}_i^{-1} Z_i \widehat{D}$ can be investigated for this purpose. In this study, we used Gaussian finite mixture models implemented R package mclust for classifications BLUP and principal components of BLUP (BLUPLNT).

4. Real Data Applications: We have used five datasets consisting of trajectories of 2-3 distinct components with varying level of separability to compare classification performance of three methods. The datasets of mixture distributions are created by combining components of plausible homogeneous patterns of linear, quadratic and cubic trends of early childhood growth trajectories identified from a large dataset of 3,365 children. The dataset consists of the standardized scores of weight-for-length (at ages < 2 years) and body mass index (BMI) (at ages ≥ 2 years), collected on clinic visits during their first 5 years of life. In the United States, weight-for-length and BMI are common measures of the somatic growth of children aged < 2 years and ≥ 2 years, respectively. Because the same quantile cutoffs of the two variables are used to classify the weight status of children, the standardized score of this variable, denoted BMIz, has been used as an early childhood growth indicator. The data were retrospectively retrieved from electronic health records.

Using the identified plausibly homogeneous subsets, we generated 5 datasets of two or three components with varying extent of separability. We generated classifications of 2-4 groups using four methods: HLME in Mplus (GMM), HLME in R (HLME), post-hoc mixture model of BLUPs using R-package mclust (BLUP), and post-hoc mixture model of principal component analysis of BLUPs (BLUPLNT). We used Bayesian information criteria (BIC) from the following three methods for evaluations: Mplus GMM fit (Mplus), R HLME fit (RHLME), the goodness of fit of the linear mixed effects model. The linear fixed effects model was applied on the solutions of each method, namely Mplus GMM

(GMM), R HLME (HLME), BLUP and BLUPLNT. The description of the datasets and classification and evaluation performance results for all methods are presented below:

Dataset 1 consists of three well isolatable components (Figure 1) of trajectories with linear and opposite quadratic trends of BMIz. Component 1 of this dataset consists of 254 children with a linear trend, component 2 consists of 83 children with concave downtrend, and component 3 consists of 50 children with a concave up trend. Figure 1 shows classifications of 2, 3, and 4 groups of this dataset using GMM, HLME, BLUP and BLUPLNT. In two-group classification, GMM, BLUP and BLUPLNT produced similar results, combining components 1 and 3 into one group. By contrast, HLME split all three components to form two groups. In three-group classifications, all methods were able to identify the three components as distinct groups. In four-group classifications, all methods retained components 2 and 3 as distinct groups and split component 1 into two groups. GMM and HLME retained most trajectories of the component in the same group while a BLUP and BLUPLANT divided the group more evenly. In terms of the evaluations of cluster solutions, all methods uniquely identified three-group classifications as the best (Figure 1). To evaluate the same number of classes, GMM and BLUPLNT in two-group classifications, HLME and BLUPLNT in three-group classifications, and BLUP and BLUPLNT emerged as optimum in four-group classifications, respectively using mixed effects model.

Dataset 2 also consists of three components with distinct trajectories (Figure 2). Components 1 and 2 of this dataset are the same as those in dataset 1. Component 3 comprises a cubic mean trajectory trend. Once again, all methods achieved similar performance to identify three components as distinct groups. In two-group classifications, GMM combines components 1 and 2 into 1 group, HLME combines 1 and 3, and BLUP and BLULNT combine 2 and 3 into one group. All methods retained two components distinct and split one component in to two groups in four-group classifications. GMM split component 3, while other methods split component 1, but in different sizes. In terms of the evaluation, Mplus showed a tendency of picking higher number of groups as the better solution, HLME picked the three-group classification as the optimum. Linear mixed effects model fit chose three-group classifications as the best for GMM and HLME solutions and two-group classifications for BLUP and BLUPLNT. Using this method of evaluation to compare classifications of the same number of groups, BLUP and BLUPLNT for two-group, GMM for three-group and BLUP for four-group classifications appeared to be the best.

Dataset3 comprises three less separable components (Figure 3). Component 1 consists of 156 children with a weak cubic mean trend in BMIz, components 2 and 3 contains 102 and 126 children with mean trends of opposite patterns at the beginning of the life. Methods differed substantially in both classifications and evaluations of this dataset. In two-group classifications, GMM, BLUP, and BLUPLNT merged children of components 1 and 2 into the same group. Component 3 remained as a distinct group. HLME split components 1 and 3 to form two groups. One group contained children from component 2 and splits of components 1 and 3, and the group contained children splits of components 1 and 3. In three-group classifications of this dataset, GMM and HLME merged components 1 and 2 and splits component 3; BLUP split component 1 to merge 43 children with component children in component 2, and retained component 3 as a distinct group. BLUPLNT was almost the same as BLUP except that it merged 11 children from component 1 with children of component 2. In four-group classifications, HLME, BLUP, and BLUPLNT appeared to be similar. All three methods split components 1 and 3, merged a split of component 1 with component 2, and split component 3 to two distinct groups. GMM merged components 1 and 2 into a group and split component 3 into three distinct groups. For evaluations of the cluster solutions, Mplus suggested an increased number of classifications, HLME suggested its three group classifications, and linear mixed effects model suggested BLUP's three-group classifications as the optimum solutions. We can also evaluate classifications of the same group numbers across four methods using mixed effects model. In two-group classifications, GMM, BLUP, and BLUPLNT have the optimum solution. In three- and four-group classifications, BLUP solutions are the best. Overall in this method of evaluation, BLUP's three-group classification is the best.

Dataset 4 is composed of two clearly opposite quadratic trends (Figure 4). GMM, BLUP and BLUPLNT identified two components perfectly in two-group classifications while HLME split both components. In three- and four-group classifications, all methods performed similarly. Specifically, all methods retained component 1 as distinct and split component 2. Mplus and RHLME picked four-group classifications as the optimum, while linear mixed effects model picked two-group classifications using GMM, BLUP and BLUPLNT as the optimum. For the HLME solutions, this method picked three-group classification as the optimum. According this method, BLUP perform the best for two-, three-, and four-group classifications among four methods.

Dataset 5 is made of two components (Figure 5). Trajectories in component 1 showed trend to become heterogeneous with increased age. GMM, BLUP, and BLUPLNT performed similarly once again for two-group classifications. HLME combined almost all children (97 out of 101) of component 1 with the children of component 2 for this classification. For three-group classifications, GMM and HLME split component 2 in two different groups, and component 1 remained distinct. BLUP and BLUPLNT still retained two-components as two distinct groups and third group contains only one child. For four-group classifications, GMM and HLME split both components to form four groups. BLUP and BLUPLNT still retained component 1 as distinct group and split component 2 in three groups with one group contained only one child.

5. Discussions and conclusions: GMM and HLME combine the concepts of linear mixed effects and mixture models to classify hidden subgroups in the longitudinal unbalanced data. Theoretically, both methods are firm footed and both methods apply mixture models on the heterogeneity in trajectories captured by random coefficients of the mixed effects model. The process of combining these two concepts becomes computationally complicated as it requires estimating increased number of parameters. None of the methods considered in this study performed consistently in all datasets. As discussed, GMM in Mplus generally works well on relatively small datasets with less degree of unbalancedness. The HLME in R can handle large datasets, but was very inconsistent in regards to the accuracy of the classification. Also, computational time can increase greatly with the size of the data and number of random coefficients. Application of post-hoc mixture models on the empirical random coefficients may reduce complexities drastically as it does not involve in estimating extra parameters for combining random coefficients and mixture models. In running programs in all three methods, we need a linear mixed effects model that captures the heterogeneity in trajectories accurately. In other words, classification performance of all these methods largely depends on the accuracy of the model fit. The classification performance also depends on the separability of the components in the datasets. We identified a piecewise linear mixed effects model for each dataset that best fit to the corresponding dataset and then applied the same model across all methods. In regards to the classification performance, we investigated the strength of the methods in classifying datasets of known number of components with varying separability into the same number as well as more or fewer number of groups.

Methods performed equally in identifying components for datasets with clear separability, and varied substantially for datasets with less separable components. All methods performed uniformly for classifications and evaluations to identify three components of the dataset 1 as three distinct groups. This is a dataset in which components of this dataset are unambiguously well-separable. It is also obvious from graphical inspection that the three components are distinct with sound homogeneity among within component trajectories. When squeezed three components into two classess, HLME split all components to form two groups which was different from three other methods. In mixed model evaluation, this classification was worst among four in terms of BIC. When classified three components in to four classes, BLUP and BLUPLNT appeared to perform much better according to BIC using linear mixed effects model.
The dataset 2 differ from the dataset 1 by only component 3. This component of the dataset 2 is slightly more complex in terms of the degree of polynomial and heterogeneity than that of the component 3 in the dataset 1. This makes components of the dataset 2 slightly less separable than that of the dataset 1. Once again, all methods performed similarly to identify three components. But, the evaluation was different across methods. GMM in Mplus tended to show that the higher the number of groups the better is the classification, while HLME in R showed three-groups as the optimum. For the classifications using GMM and HLME, the linear mixed effects model suggested three groups as the optimum classification. Using the linear mixed effects model, BIC is the same for HLME, BLUP and BLUPLNT, however, if we consider all model fits, it is the smallest for two-group BLUP classification. Once again, BLUP performed better in two- and four-groups classification of this three-component dataset using the linear mixed effect evaluation.

Components of dataset 3 are least separable among the three-component datasets. To group into three classes of this three-component dataset, GMM and HLME worked similarly and merged component 1 with component 2 and split component 3. Likewise linear mixed effects model, visual inspection of the components of this dataset seems don't quite support these classifications. Once again, GMM in Mplus supported a higher number of clusters, while RHLME supported a three-group classification. Linear mixed effects indicated two groups in GMM and four in HLME as the optimum for the corresponding methods. Three-group classifications using BLUP and BLUPLNT are quite different, and linear mixed effects models identified this classification using BLUP as the optimum among all groupings. In each of the two-, three-and four-group classifications, BLUP performed best according to the linear mixed effects model.
Components of dataset 4 are highly separable from each other. GMM, BLUP and BLUPLNT identified perfectly both components of this dataset in two-group classifications. However, GMM in Mplus identified 4 group-classifications as optimum, while linear mixed effect models supported this two-group classification as the best. HLME performed poorly in identifying two components of this dataset. In three- and four-group classifications of this two-component dataset, BLUP performed better in terms of the evaluation by linear mixed effects model.

Components of dataset 5 are less separable. GMM and BLUP and BLUPLNT performed similarly to identify two components of this dataset. GMM in Mplus supported this as the optimum which is suggested by linear mixed effects model also. However, for overall groupings, the linear mixed effects model suggested four clusters by the HLME as the optimum followed by the same numbers of BLUP classification. However, HLME once again completely failed to recognize two components as distinct groups.

Looking across all datasets, when the number of groups in the modeling approach is the same as the number of components in the dataset, BLUP and BLUPLNT performed well in identifying components in all datasets, GMM was able to identify the correct clusters in 4 datasets, and HLME only performed well in 2 3-cluster datasets with separable components. HLME performed poorly in classifying two-component datasets irrespective of separability. Evaluations using linear mixed effects model appeared to be more realistic than fit statistics from the GMM in Mplus and RHLME. Choosing the number of clusters based on the BIC from the GMM in Mplus tended to pick higher number of groups as the optimum classifications. HLME showed inconsistent performance in classification and then evaluation was often different from others. In evaluations of classifications involving fewer or more number of groups than the number of components, fit statistics BIC for BLUP and BLUPLNT classes was lower in most cases than that for classes of GMM and HLME as per evaluations of linear mixed effects models. This may suggestive that BLUP and BLUPLNT are likely to be more reasonable than two other methods in these situations.

The other strengths of the post-hoc mixture modeling of BLUP or BLUPLNT are that, unlike HLME and GMM, this method can provide more freedom in terms of software use by allowing to perform mixed effects model and mixture models in different statistical software packages, doesn’t impose serious computational burden for the increase in size and the level of unbalancedness of the dataset, and is less likely to be affected by the problem related to multiple maxima of the likelihood as the model is less complex than two others.

In summary, this study indicates that post-hoc mixture modeling of BLUP and BLUPLNT can be used to identify the hidden components of unbalanced longitudinal data. Classification performance heavily relies on the accuracy of the model fit in all methods. Methods perform classification and evaluation with similar level of accuracy for datasets with clearly separable components, and differ in performance for datasets with relatively less separability. Post-hoc mixture modeling of BLUPs or BLUPLNTs showed better accuracy than two other methods in classification and evaluation in this situation. Overall, post-hoc mixture modeling methods perform better than two other methods on these five datasets. Future appropriate simulation studies could provide further insight in comparing the strength and weakness of the three methods in classifying longitudinal data with irregular spaces between measurement times.

Acknowledgments
The first author received partial salary support from the National Institutes of Health (NIH) COBRE Grant 8P20GM103464-9 (PI: Shaffer) and an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the NIH under Grant no. U54-GM104941 (PI: Binder-Macleod).

References:

Jung, T. and Wickrama, K.A.S., 2008. An introduction to latent class growth analysis and growth mixture modeling. Social and personality psychology compass, 2(1), pp.302-317.

Wang, M. and Bodner, T.E., 2007. Growth mixture modeling: Identifying and predicting unobserved subpopulations with longitudinal data. Organizational Research Methods, 10(4), pp.635-656.

Jha, A.K., DesRoches, C.M., Campbell, E.G., Donelan, K., Rao, S.R., Ferris, T.G., Shields, A., Rosenbaum, S. and Blumenthal, D., 2009. Use of electronic health records in US hospitals. New England Journal of Medicine, 360(16), pp.1628-1638.

Wei, W.Q. and Denny, J.C., 2015. Extracting research-quality phenotypes from electronic health records to support precision medicine. Genome medicine, 7(1), p.41.

Fitzmaurice, G.M., Laird, N.M. and Ware, J.H., 2012. Applied longitudinal analysis (Vol. 998). John Wiley & Sons.

Laird, N.M. and Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics, 38(4), pp.963-974.

Verbeke, G. and Lesaffre, E., 1996. A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association, 91(433), pp.217-221.

Komárek, A., Verbeke, G. and Molenberghs, G., 2002. A SAS-Macro for Linear Mixed Models with Finite Normal Mixtures as Random-Effects Distribution, version 1.1. *Katholieke Universiteit Leuven Biostatistisch Centrum*. http://www.med.kuleuven.ac.be/biostat/research/ software.htm, 2001.

Spiessens, B., Verbeke, G. and Komárek, A., 2002. A SAS-macro for the classification of longitudinal profiles using mixtures of normal distributions in nonlinear and generalised linear mixed models. http://www.med.kuleuven.ac.be/ biostat/research/software.htm, 2002.

Proust C, Jacqmin-Gadda H. Estimation of linear mixed models with a mixture of distribution for the random effects. Computer methods and programs in biomedicine. 2005 May 1;78(2):165-73.

Proust-Lima, C., Philipps, V. and Liquet, B., 2015. Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. arXiv preprint arXiv:1503.00890.

Dempster, A.P., Laird, N.M. and Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39(1), pp.1-22.

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. Journal of the society for Industrial and Applied Mathematics, 11(2), pp.431-441.

Proust, C. and Jacqmin-Gadda, H., 2005. Estimation of linear mixed models with a mixture of distribution for the random effects. Computer methods and programs in biomedicine, 78(2), pp.165-173.

Muthén, B. and Shedden, K., 1999. Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics, 55(2), pp.463-469.

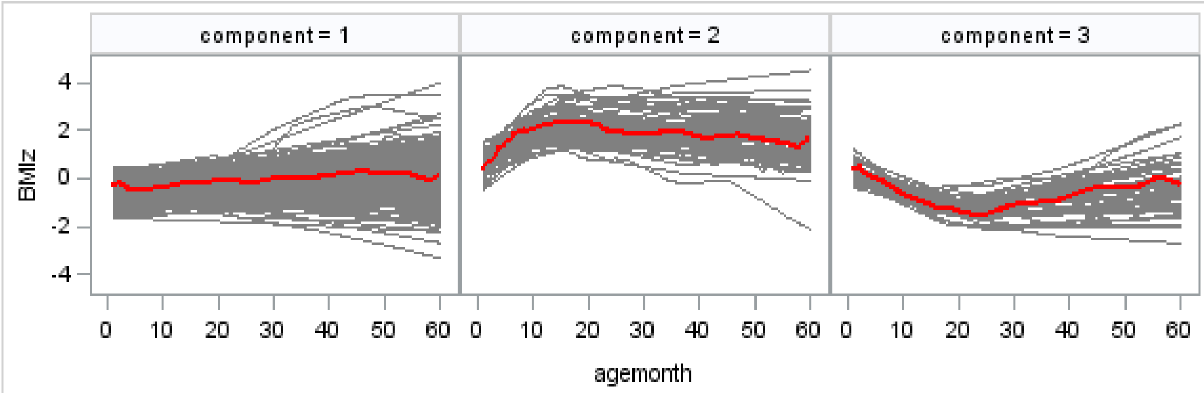
Muthén, B. and Asparouhov, T., 2008. Growth mixture modeling: Analysis with non-Gaussian random effects.

hossain, M. J. 2016. Clustering Longitudinal Unbalanced Data: An Application to the Early Childhood Growth Pattern, Section on Statistics in Epidemiology. Alexandria, VA: American Statistical Association, pp.

Arminger, G., Clogg, C.C. and Sobel, M.E. eds., 2013. Handbook of statistical modeling for the social and behavioral sciences. Springer Science & Business Media.

Garrett, E.S. and Zeger, S.L., 2000. Latent class model diagnosis. Biometrics, 56(4), pp.1055-1067.

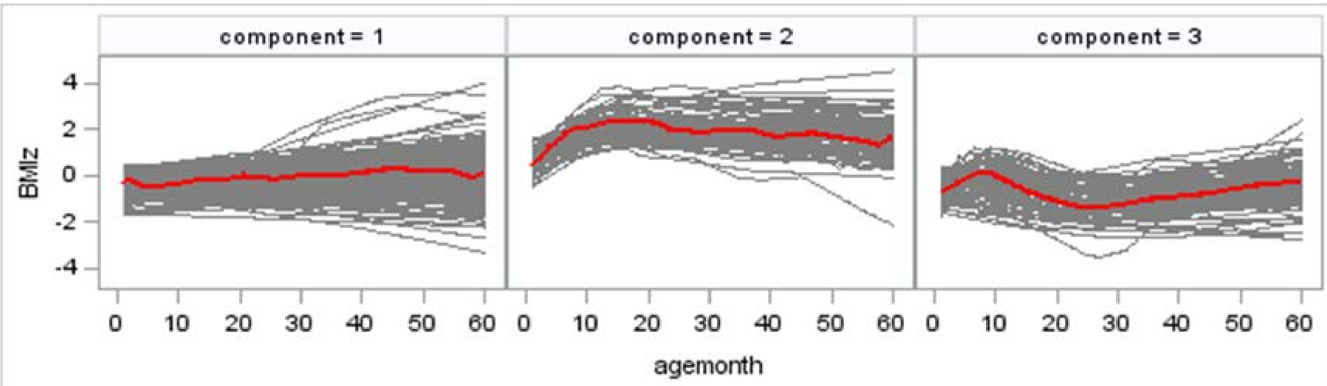
Figure 1: Classifications and Evaluations of Dataset 1



Classifications													
		Comp1	Comp2	Comp3	Comp1	Comp2	Comp3	Comp1	Comp2	Comp3	Comp1	Comp2	Comp3
		N=254	N=83	N=50	N=254	N=83	N=50	N=254	N=83	N=50	N=254	N=83	N=50
No. of Classes	Groups	GMM			HLME			BLUP			BLUPLNT		
2	1	254	0	50	131	39	41	251	0	50	254	0	50
	2	0	83	0	123	44	9	3	83	0	0	83	0
3	1	0	83	0	0	83	0	0	0	50	0	0	50
	2	254	0	1	254	0	0	251	0	0	254	0	0
	3	0	0	49	0	0	50	3	83	0	0	83	0
4	1	6	0	0	13	0	0	0	0	50	0	0	50
	2	0	0	48	241	0	0	65	0	0	128	0	0
	3	0	83	0	0	83	0	188	0	0	126	0	0
	4	248	0	2	0	0	50	1	83	0	0	83	0

Evaluations						
	BIC					
			Mixed Effects Model			
Groups	Mplus	RHLME	GMM	HLME	BLUP	BLUPLNT
2	7952	9895	9241	9562	9263	9241
3	**7815**	**9418**	**9106**	**9094**	**9107**	**9094**
4	7819	9433	9708	9678	9358	9351

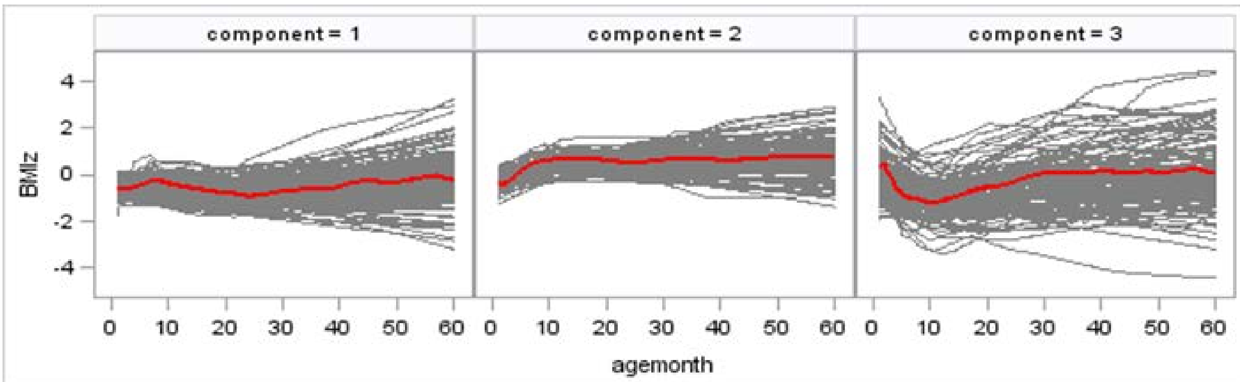
Figure 2: Classifications and Evaluations of dataset 2



No. of Classes	Groups	Classifications											
		Comp1	Comp 2	Comp 3	Comp 1	Comp 2	Comp 3	Comp1	Comp2	Comp3	Comp1	Comp2	Comp 3
		N=254	N=83	N=84	N=254	N=83	N=84	N=254	N=83	N=84	N=254	N=83	N=84
		GMM			HLME			BLUP			BLUPLNT		
2	1	253	83	3	254	0	82	254	0	8	0	83	79
	2	1	0	81	0	83	2	0	83	76	254	0	5
3	1	253	0	5	254	0	1	254	0	1	0	83	0
	2	0	83	0	0	0	83	0	0	83	254	0	1
	3	1	0	79	0	83	0	0	83	0	0	0	83
4	1	0	83	0	0	0	83	159	0	0	0	83	0
	2	0	0	45	243	0	1	95	0	2	81	0	0
	3	254	0	2	0	83	0	0	0	82	0	0	84
	4	0	0	37	11	0	0	0	83	0	173	0	0

Evaluations						
Groups	BIC					
	Mplus	RHLME	Mixed Effects Model			
			GMM	HLME	BLUP	BLUPLNT
2	9197	11157	10807	10890	10742	10785
3	9002	10920	10801	10842.6	10843	10843
4	8956	10940	11259	11051.2	11019	11060

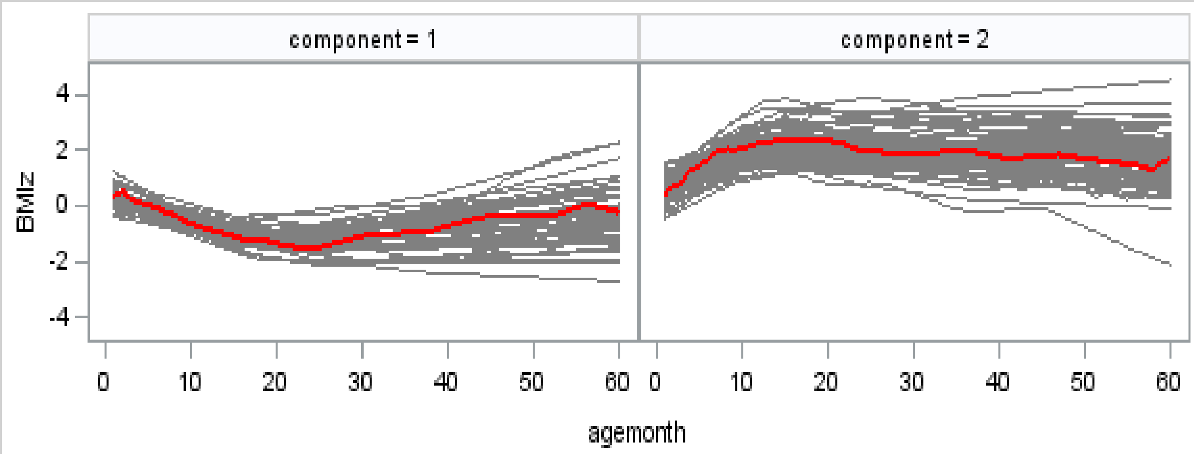
Figure 3: Classifications and Evaluations of Dataset 3



Classifications													
No. of Classes	Groups	Comp1	Comp2	Comp3	Comp1	Comp2	Comp3	Comp1	Comp2	Comp3	Comp1	Comp2	Comp3
		N=156	N=102	N=126	N=156	N=102	N=126	N=156	N=102	N=126	N=156	N=102	N=126
		GMM			HLME			BLUP			BLUPLNT		
2	1	156	102	0	141	0	94	0	0	126	0	0	126
	2	0	0	126	15	102	32	156	102	0	156	102	0
3	1	156	102	0	0	0	65	0	0	126	145	0	0
	2	0	0	77	156	102	0	113	0	0	11	102	0
	3	0	0	49	0	0	61	43	102	0	0	0	126
4	1	156	102	0	50	102	0	0	0	56	0	0	62
	2	0	0	71	106	0	0	0	0	70	0	0	63
	3	0	0	29	0	0	65	113	0	0	145	0	1
	4	0	0	26	0	0	61	43	102	0	11	102	0

Evaluations						
Groups	BIC					
	Mplus	RHLME	Mixed Effects Model			
			GMM	HLME	BLUP	BLUPLNT
2	9687	11829	11304	11417	11304	11304
3	9594	11472	11491	11651	11176	11293
4	9562	11496	11638	11261	11252	11334

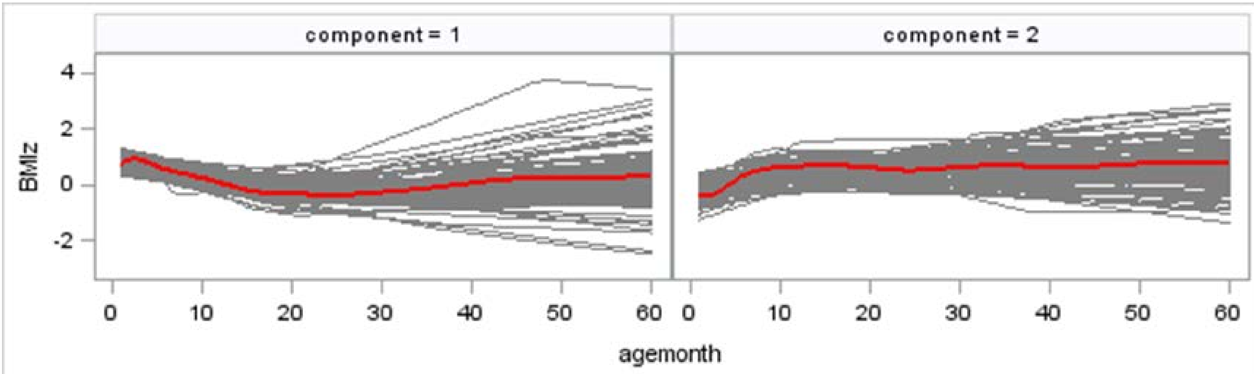
Figure 4: Classifications and Evaluations of the dataset 4



No. of Classes	Groups	GMM Comp1 N=50	GMM Comp2 N=83	HLME Comp1 N=50	HLME Comp2 N=83	BLUP Comp1 N=50	BLUP Comp2 N=83	BLUPLNT Comp1 N=50	BLUPLNT Comp2 N=83
2	1	50	0	31	40	0	83	50	0
	2	0	83	19	43	50	0	0	83
3	1	0	79	50	0	0	47	50	0
	2	0	4	0	72	0	36	0	30
	3	50	0	0	11	50	0	0	53
4	1	0	58	0	61	0	39	50	0
	2	0	4	0	18	0	30	0	27
	3	50	0	0	4	0	14	0	31
	4	0	21	50	0	50	0	0	25

Groups	BIC Mplus	BIC RHLME	Mixed Effects Model GMM	Mixed Effects Model HLME	Mixed Effects Model BLUP	Mixed Effects Model BLUPLNT
2	2688	3416	3049	3299	3049	3049
3	2678	3200	3129	3139	3160	3173
4	2655	3182	3406	3249	3144	3223

Figure 5: Classifications and Evaluations of Dataset 5



		Classifications							
No. of Classes	Groups	Comp1 N=101	Comp2 N=102	Comp1 N=101	Comp2 N=102	Comp1 N=101	Comp2 N=102	Comp1 N=101	Comp2 N=102
		GMM		HLME		BLUP		BLUPLNT	
2	1	0	101	97	102	101	1	100	0
	2	101	1	4	0	0	101	1	102
3	1	0	30	0	31	101	0	101	0
	2	101	1	101	0	0	1	0	101
	3	0	71	0	71	0	101	0	1
4	1	89	1	0	33	101	0	100	0
	2	12	0	0	69	0	1	1	50
	3	0	34	8	0	0	70	0	1
	4	0	67	93	0	0	31	0	51

	Evaluations					
Groups	BIC					
	MPlus	RHLME	Mixed Effects Model			
			GMM	HLME	BLUP	BLUPLNT
2	3819	4771	4629	4718	4621	4656
3	3823	4565	4708	4690	4624	4624
4	3831	4578	4722	4523	4583	4604