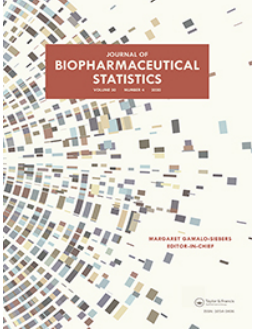

Journal of Biopharmaceutical Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lbps20

Subgroup analysis based on structured mixed-effects models for longitudinal data

Juan Shen & Annie Qu

To cite this article: Juan Shen & Annie Qu (2020) Subgroup analysis based on structured mixed-effects models for longitudinal data, Journal of Biopharmaceutical Statistics, 30:4, 607-622, DOI: 10.1080/10543406.2020.1730867

To link to this article: https://doi.org/10.1080/10543406.2020.1730867

Subgroup analysis based on structured mixed-effects mo longitudinal data

Juan Shen[a] and Annie Qu [b]

[a]Department of Statistics, Fudan University, Shanghai, China; [b]Department of Statistics, University c Irvine, Irvine, California, USA

ABSTRACT

In recent years, subgroup analysis has emerged as an important tool to identify unknown subgroup memberships. However, subgroup analysis is still under-studied for longitudinal data. In this paper, we propose a structured mixed-effects approach for longitudinal data to model subgroup distribution and identify subgroup membership simultaneously. In the proposed structured mixed-effects model, the heterogeneous treatment effect is modeled as a random effect from a two-component mixture model, while the membership of the mixture model is incorporated using a logistic model with respect to some covariates. One advantage of our approach is that we are able to derive the estimation of the treatment effects through an EM-type algorithm which keeps the subgroup membership unchanged over time. Our numerical studies and real data example demonstrate that the proposed model outperforms other competing methods.

ARTICLE H
Received 1
Accepted 2

KEYWORD
EM algorith
heterogene
mixed-effe
mixture mo
identificati

1. Introduction

Subgroup analysis has emerged as an important tool in various applications such as personalized medicine, and market segmentation. The earlier development of subg mainly focused on analysis with given subgroup memberships, such as in Song and C Altstein et al. (2011), among others. In recent years, there has been growing interest subgroups whose memberships are unknown. Su et al. (2009), Foster et al. (2011) a et al. (2011) utilize tree-based methods to categorize subgroups with an enhanced tre Cai et al. (2011) and Zhao et al. (2013) derive a parametric scoring system to facilit assignments to new patients. A Bayesian method for multiple subgroup analysis is Berger et al. (2014). Shen and He (2015) and Shen et al. (2017) also propose a struct mixture model that simultaneously models subgroup membership and treatment out each subgroup as functions of other covariates. Wu et al. (2016) and Fan et al. (201 model in Shen and He (2015) for Cox models and semi-parametric models.

However, not much work has been done in subgroup identification for longitudi which occur frequently in biomedical studies since responses are usually collected ove numerical studies provided by the aforementioned papers, they consider the data at o point only, even though more longitudinal data are actually available. This could lea information as not all time points are utilized. In this paper, our goal is to ident through a structured mixed-effects model for longitudinal data.

Our method is motivated by the Aids Clinical Trials Group 320 study (ACTG320) (1 1997) on the human immunodeficiency virus type 1 (HIV–1) infection where the in

CONTACT Annie Qu aqu2@uci.edu Department of Statistics, University of Illinois at Urbana-Champaig USA

Supplemental data for this article can be accessed here.

effects. The targeted outcome of the treatment is to increase or to inhibit the declin counts. Therefore, the observed response of this study is the change in CD4 counts a points.

For longitudinal data analyses, there are two main approaches in general. One is estimating equation method, such as in Liang and Zeger (1986) and Tang and Qu (20 the marginal estimating equations approach in Tang and Qu (2016) is designed for gr ship changing over time, which is not applicable as we require that the subgroup remain unchanged over time in clinical trials case here.

Another popular approach is to apply a random-effects model to incorporate the repeated measurements from the same subject. In addition, to capture the heterogenei population, the random effects are viewed to follow a mixture distribution with con tions. For example, Verbeke and Lesaffre (1996) apply a linear mixed-effects model w effects and random effects, and the random-effects are formulated from normal m model configurations are widely used in various applications, such as longitudinal clin and Hedeker (2001)), next-generation sequencing data (Cacho et al. (2018)), and r expression data (Celeux et al. (2005)). To study the association of the latent class mem the covariates, the group structure is further modeled through another level of model. Lin et al. (2002) propose a latent class approach to jointly model and estimate long where the latent class is determined by some covariates through a multi-logit model. In Asparouhov (2009), a two-level mixture model is proposed to investigate gender mathematical achievement from the US National Education Longitudinal Survey. Komarek et al. (2010) introduce a two-step method to classify subjects into multiple gr Lima et al. (2016) propose a joint model to analyze multiple longitudinal markers causes of progression simultaneously based on a latent process and latent classes. L (2014) develop a nonlinear mixed-effects mixture model for longitudinal data to ac geneity and skewness. The misspecified effects on classes are examined in multiva models through Monte Carlo simulations in Nylund-Gibson and Masyn (2016) .

In general, mixture models are widely applicable in many fields; for instance see T Lindsay (1995) and Fruhwirth-Schnatter (2006). Mixture models with covariates-depe are popular in social sciences for individual longitudinal profiles where latent classes a dependent are represented as the growth mixture framework (Hu et al. 2017; Huang and Harring 2017; Muthen et al. 2002; Proust-Lima et al. 2017). Such models are a "mixture of experts" models (Jordan and Jacobs 1994; Yuksel et al. 2012) in compute

In this paper, we adopt a mixed-effects model where the treatment effect is modelec effect from a two-component mixture model, and the membership of the mixture mc modeled using a logistic model of other covariates. There are four main advantages of method. First, we extract a subgroup with more efficient estimation of treatmen longitudinal data information is fully utilized, compared to existing ones using data time point only. The random-effects model also enables us to incorporate dependency subject. Second, the proposed method ensures that the constraint of unchanged subgr ship is satisfied. Third, the proposed method models the subgroup membership and th of each subgroup jointly, which provides better prediction and interpretation of Fourth, we provide a continuous score function for future subjects, and a higher s a higher chance of enhanced treatment effect. In addition, we provide an efficient algorithm in our estimation.

The rest of the paper is organized as follows. We elaborate the proposed methodolog where the structured mixed-effects model is introduced, and estimation and further

2. Methodology

2.1. *Structured mixed-effects models*

In this section, we illustrate the structured mixed-effects model for subgroup analysis. I
the random effects are assumed to follow a normal mixture model (Verbeke and Les
That is, for each observation, the response $y_i \in R^{n_i \times 1}$ is modeled as

$$y_i = X_i\alpha + Z_i b_i + \varepsilon_i, i = 1, \cdots, N,$$

where X_i and Z_i are vectors of the covariate, α is the fixed effect, and b_i is the vector c
specific random effect following a normal mixture distribution:

$$b_i \sim pN(\mu_1, D) + (1-p)N(\mu_0, D).$$

In Xu and Hedeker (2001), the random effect in (2.1) is assumed to follow a $k-g$
mixture model with equal covariance Σ but different means $\mu_1, \cdots, \mu_k$, and the prop
sponding to mixture components are constants $p_1, \cdots, p_k$. An EM algorithm is derived
estimation where the random effects are treated as missing. Based on the joint distribut
the posterior distribution of b_i is derived given y_i and the current parameter p, μ_1, μ_0, a
the E step, the posterior mean can be calculated. A similar strategy is also applied in H
(2005) and Ng et al. (2006) to model correlated data with unknown clusters.

In our approach, we incorporate covariance information to characterize subgroup
where the mixture proportion also depends on covariates, such as the baseline med
ments in the ACTG320 study.

We first consider a simple case where there are two subgroups capturing two differ
random effects with different means. Let Y_{it} be a continuous response for $i = 1, \cdots, n$;
$X_i \in R^{q_2}$ are covariates associated with the subgroup membership; and $Z_{it} \in R^{q_1}$ are
covariates associated with the subgroup mean. We introduce a latent subgroup, $\delta_i \in$
can be modeled through a logistic regression of X_i. That is,

$$logit(P(\delta_i = 1 | X_i, Z_i)) = X_i^T \gamma.$$

Given δ_i, the treatment effect b_i is a random vector following a normal mixture model
if $\delta_i = 1$ and μ_0 if $\delta_i = 0$, and the covariance matrix $D_i = \sigma^2 I_{n_i \times n_i}$. Let $Y_i = (Y_{i1}, \cdots$
and $Z_i \in R^{q_1 \times n_i}$ be the matrix of $\{Z_{it} : t = 1, \ldots, n_i\}$. Given the subgroup indicator δ_i
covariate information, Y_i is normally distributed.

$$Y_i | (\delta_i, X_i, Z_i, T_i, b_i) = Z_i^T \alpha + T_i b_i + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \sigma^2 R_{n_i \times n_i})$, and $T_i \in \{0, 1\}$ with 1 for treatment and 0 for control group
effect $b_i \sim N(\mu_i, \sigma^2 I_{n_i \times n_i})$, independent of the noise ε_i. Then, conditional on X_i, Z_i, T_i an
are jointly normal:

$$\begin{pmatrix} Y_i \\ b_i \end{pmatrix} \sim N \begin{bmatrix} Z_i^T \alpha + T_i \mu_i \\ \mu_i \end{bmatrix}, \begin{bmatrix} \Sigma_i & T_i D_i \\ T_i D_i & D_i \end{bmatrix}$$

where $\mu_i = \mu_1 \delta_i + \mu_0 (1 - \delta_i)$, $D_i = \sigma^2 I_{n_i \times n_i}$, and $\Sigma_i = T_i^2 D_i + \sigma^2 R_i$. For balanced data, v
the i index for R_i and D_i. Let $Y_i = (Y_{i1}, \ldots, Y_{in_i})^T$, then we can also write the model as
For $t = 1, \ldots, n_i$,

Based on (2.4), the posterior mean of b_i given X_i, Z_i, T_i, δ_i, and Y_i is

$$\hat{b}_i = E[b_i|Y_i, \delta_i] = \mu_i + T_i D \Sigma_i^{-1}(y_i - Z_i^T \alpha - T_i \mu_i).$$

In addition, for each subgroup $k = 0, 1$, let

$$\hat{b}_{ik} = \mu_k + T_i D_i \Sigma_i^{-1}(y_i - Z_i^T \alpha - T_i \mu_k),$$

where the corresponding covariance matrix is

$$\hat{D}_i = D_i - T_i^2 D_i \Sigma_i^{-1} D_i.$$

Then, the second moment of random effects b_i conditional on latent variable δ_i and

$$B_{ik} = E[b_i b_i^T | \delta_i = k, Y_i] = \hat{b}_{ik} \hat{b}_{ik}^T + \hat{D}_i.$$

For simplicity, we first consider the balanced case where $n_i = n_0, R_i = R$, for all $i =$ parameters we intend to estimate are $\theta = (\gamma, \alpha, \mu_1, \mu_0, \sigma^2, R)$. When $\mu_1 \neq \mu_0$, the m defined. Therefore, without loss of generality, we assume that $\mu_1 > \mu_0$, and provide algorithm for parameter estimation in the next section.

2.2. *Estimation*

In this section, we treat the random effect b and the indicator δ as missing. The log-lik hypothetical complete data (Y, δ, b) is

$$\begin{aligned} Cl(\theta) = & \sum_{i=1}^{n} \{\delta_i \log \pi(x_i^T \gamma) + (1 - \delta_i) \log(1 - \pi(x_i^T \gamma))\} \\ & + \sum_{i=1}^{n} \{\delta_i \log \phi(y_i, Z_i^T \alpha + T_i b_i, \sigma^2 R) + (1 - \delta_i) \log \phi(y_i, Z_i^T \alpha + T_i b_i, \\ & + \sum_{i=1}^{n} \{\delta_i \log \phi(b_i, \mu_1, \sigma^2 I) + (1 - \delta_i) \log \phi(b_i, \mu_0, \sigma^2 I)\}, \end{aligned}$$

where $\pi(x) = 1/(1 + e^{-x})$ and $\phi(y, \mu, \Sigma)$ is a multivariate normal density at y with covariance matrix Σ.

We implement the EM-type algorithm as follows. Given current parameter estima next step, we have $Q(\theta|\theta^{(j)}) = I_1 + I_2$, where

$$I_1 = \sum_{i=1}^{n} \{E[\delta_i|Y] \log \pi(x_i^T \gamma) + E[(1 - \delta_i)|Y] \log(1 - \pi(x_i^T \gamma))\},$$

and

$$\begin{aligned} I_2 = & \sum_{i=1}^{n} \sum_{k=0}^{1} -\frac{P(\delta_i = k|Y)}{2\sigma^2} [(b_{ik} - \mu_k)^T (b_{ik} - \mu_k) + tr(\hat{D}_i))] \\ & + \sum_{i=1}^{n} \sum_{k=0}^{1} -\frac{P(\delta_i = k|Y)}{2\sigma^2} [(Y_i - Z_i^T \hat{\alpha} - T_i b_{ik})^T R^{-1} (Y_i - Z_i^T \hat{\alpha} - T_i b_{ik}) + T_i^2 \\ & - N \log(\sigma^2) - \frac{n}{2} \log |R|, \end{aligned}$$

where $\hat{D}_i$ is defined in (2.8), and $N = \sum n_i$ ($= nn_0$ for the balanced case).

We estimate θ at the $(j + 1)th$ step:

$$\theta^{(j+1)} = \text{argmax}_\theta Q(\theta|\theta^{(j)}).$$

and fit the linear model in (2.3) with b_i substituted by $\mu_1\delta_i + \mu_0(1-\delta_i)$ to obt
the initial parameters.

(2) (E Step) Given the current parameter $\theta^{(j)}$,

(a) For $i = 1, \cdots, n$, let

$$p_{i1} = P(\delta_i = 1|y_i, \theta^{(j)})$$
$$= \frac{\pi(x_i^T\gamma^{(j)})\phi(y_i, Z_i^T\alpha^{(j)} + T_i\mu_1^{(j)}, \Sigma^{(j)})}{\pi(x_i^T\gamma^{(j)})\phi(y_i, Z_i^T\alpha^{(j)} + T_i\mu_1^{(j)}, \Sigma^{(j)}) + (1-\pi(x_i^T\gamma^{(j)}))\phi(y_i, Z_i^T\alpha^{(j)} + T_i\mu_0^{(j)}, \Sigma}$$

and $p_{i0} = 1 - p_{i1}$, where $\phi(\cdot, \mu, \Sigma)$ is the probability density function c
vector with mean μ and covariance matrix Σ.

(b) For each subgroup $k = 0, 1$, calculate b_{ik} following (2.7), where the c
covariance matrix $\hat{D}_i$ is calculated via (2.8) and B_{ik} is obtained via (2.9).

(3) (M Step)

(a) Calculation of the subgroup membership parameter γ:

$$\gamma^{(j+1)} = \text{argmax}_\gamma p_{i1}\log\pi(x_i^T\gamma) + p_{i0}\log(1-\pi(x_i^T\gamma)).$$

(b) Calculation of the subgroup mean parameters: for $k = 0, 1$, we have

$$\mu_k^{(j+1)} = \frac{1}{\sum p_{ik}}\sum_{i=1}^{n} p_{ik}b_{ik}.$$

(c) Calculation of the parameter α associated with covariates Z:

$$\hat{\alpha}^{(j+1)} = \left\{\sum Z_i^T R^{-1} Z_i\right\}^{-1}\left\{\sum_{i=1}^{n}\sum_{k=0}^{1} p_{ik}Z_i^T R^{-1}(Y_i - T_i b_{ik})\right\}.$$

(d) Calculation of the variance parameters:

$$(\hat{\sigma}^2)^{(j+1)} = \frac{A_1 + A_2}{2N},$$

where $A_1 = \sum_{i=1}^{n}\sum_{k=0}^{1} p_{ik}[(b_{ik} - \mu_k)^T(b_{ik} - \mu_k)) + tr(\hat{D}_i)]$ and $A_2 = \sum$
$[(Y_i - Z_i\alpha - T_i b_{ik})^T R^{-1}(Y_i - Z_i\alpha - T_i b_{i1}) + T_i^2 tr(R^{-1}\hat{D}_i))]$, and

$$R^{(j+1)} = \frac{1}{n\sigma^2}\sum_{i=1}^{n}\sum_{k=0}^{1} p_{ik}[(Y_i - Z_i\hat{\alpha} - T_i b_{ik})(Y_i - Z_i\hat{\alpha} - T_i b_{ik})^T + T_i^2\hat{D}_i].$$

(4) (Stopping Criterion) Iterate E-step and M-step until $|\theta^{(j+1)} - \theta^{(j)}| < 10^{-3}$.

Since the EM-type algorithm only guarantees a local optimum, in practice we
multiple starting values and searching for the best one among all local optima. In c
studies, we try about 10 different starting values by applying different δ_0's in the initi

In the EM-type algorithm, the standard error of the estimator can be obtained fo
(1982). That is, for the i-th observation, let S_i and B_i be the first derivative and the ne
derivative of the complete log-likelihood in (2.10), respectively. Let $\hat{\theta}$ be the estimat

$$- \sum_{i \neq j}^{n} (\mathbb{E}_{(\delta_i, b_i | W_i; \hat{\theta})} S_i)(\mathbb{E}_{(\delta_j, b_j | W_i; \hat{\theta})} S_j)^T.$$

From (2.10), the information matrix I is a block diagonal for γ and the rest of the para the block diagonal matrix associated with γ is

$$B_i(\gamma) = \pi(X_i^T \gamma) \cdot \{1 - \pi(X_i^T \gamma)\} X_i X_i^T,$$

and the corresponding score function associated with γ is

$$S_i(\gamma) = \{\delta_i - \pi(X_i^T \gamma)\} X_i.$$

Therefore, the standard errors for $\hat{\gamma}$ and the other estimators can be calculated se standard error calculation for $\hat{\gamma}$ only involves $B_i(\hat{\gamma})$, $S_i(\hat{\gamma})$, and Equation (2.16). errors of $\hat{\gamma}$ are useful to identify the subgroup membership. In addition, we can Wald test or construct confidence intervals to choose the covariates which are re subgroup membership. For example, if the absolute coefficient estimator corr a covariate is more than twice its standard error, then this covariate is likely i subgroup identification. The calculation of standard errors for the rest of the provided in Web Appendix A.

2.3. *Extension to model with time interaction*

In Section 2.1, we assume that the treatment effect is normally distributed with a c However, in general, the treatment effect could change over time. In this subsection, w mean parameter to be associated with time. Given the subgroup indicator variable $\delta_i =$

$$b_i \sim N(\mu_{k1} \text{Time} + \mu_{k0}, D).$$

The model of other parameters is similar as in Section 2.1.

Given X_i, Z_i, T_i and δ_i, Y_i and b_i are jointly normal distributed as i $\mu_i = (\mu_{11} \text{Time} + \mu_{10})\delta_i + (\mu_{01} \text{Time} + \mu_{00})(1 - \delta_i)$; for each subgroup $k = 0, 1$, the p of b_i given X_i, Z_i, T_i, δ_i, and Y_i is

$$\hat{b}_{ik} = \mu_{k1} \text{Time} + \mu_{k0} + T_i D \Sigma^{-1} \{y_i - Z_i^T \alpha - T_i(\mu_{k1} \text{Time} + \mu_{k0})\}.$$

We apply the EM-type algorithm similarly as in Section 2.2 to estimate the parameters. we compute p_i as in (2.11); the conditional mean of the random effect is calculated $k = 0, 1$, and the conditional variance of the random effect is obtained via (2.8).

In the M step, the subgroup membership parameter γ is estimated similarly as in that the subgroup mean parameters (μ_{11}, μ_{10}) satisfy the following equations:

$$\left(\sum p_i \text{Time}^T \text{Time}\right)\mu_{11} + \left(\sum p_i 1_{n_i \times 1}^T \text{Time}\right)\mu_{10} = \sum p_i b_{i1}^T \text{Time},$$
$$\left(\sum p_i \text{Time}^T 1_{n_i \times 1}\right)\mu_{11} + \left(\sum p_i 1_{n_i \times 1}^T 1_{n_i \times 1}\right)\mu_{10} = \sum p_i b_{i1}^T 1_{n_i \times 1}.$$

Similarly, the mean parameters (μ_{01}, μ_{00}) for the other subgroup are defined as th equations similar to (2.21), where p_i, b_{i1}, μ_{11}, and μ_{10} are replaced by $q_i, b_{i0},$ respectively. For the rest of the parameter estimation and standard error cal computation follows in a similar fashion as in Section 2.2, and the details are Web Appendix A.

we may have unbalanced data where the subjects are measured at different time points. Th
to impose a specific working correlation structure with some unknown correlation parame
correlation matrix $R_i(\rho)$ is the exchange structure or the first-order autoregressive with di
sizes n_i, and this leads to the estimation of ρ instead of estimating R in the M step. For sim
apply the moment estimator instead of maximizing the Q function to obtain ρ.

From (2.4), given $\delta_i = k(k = 0, 1)$, we have $E(Y_i) = Z_i^T\alpha + T_i\mu_i$ and $Var(Y_i) = \tau^2 I_{n_i}$
Let $\hat{\epsilon}_i = y_i - Z_i^T\hat{\alpha} - T_i\hat{\mu}_{ik}$ where $k = 1$ if $P(\delta_i = 1) > 0.5$ and otherwise 0. In addition, b
$\hat{\epsilon}_i \sim N_{n_i}(0, \tau^2 I_{n_i \times n_i} + \sigma^2 R_i(\rho))$. Then, given the estimated residual $\hat{\epsilon}_{ij}, i = 1, \ldots, n, j = 1$
be estimated as follows.

(1) For exchangeable correlation structure with parameter ρ,

$$\hat{\rho} = \frac{1}{\hat{\sigma}^2 \sum_{i=1}^{n} n_i(n_i - 1)} \sum_{i=1}^{n} \sum_{j \neq k} \hat{\epsilon}_{ij}\hat{\epsilon}_{ik};$$

(2) For AR(1) correlation structure with parameter ρ,

$$\hat{\rho} = \frac{1}{\hat{\sigma}^2 \sum_{i=1}^{n} (n_i - 1)} \sum_{i=1}^{n} \sum_{j=1}^{n_i - 1} \hat{\epsilon}_{ij}\hat{\epsilon}_{i(j+1)}.$$

3. Simulation studies

In this section, we perform simulation studies to compare the proposed method to
under various settings. First, we generate data based on the proposed models, that is
and (2.3) with time-invariant treatment effect, where the coefficients are specified in T
with covariance structures specified as exchangeable and AR(1). Next, we genera
a misspecified model with an error in (2.3) following a t distribution to study the
our method. We also study the unbalanced case and report the results in Table 4. We
mean-squared errors of the proposed estimators under different correlation structures,
true and working correlation structures, and the empirical estimator from the algorith

In Tables 1 and 2, we provide the mean and standard deviations of the estimates u
scenarios. The true parameter is: $\gamma = (-1, 1), \alpha = (-1, 1), \sigma = 1, \mu_1 = (10, \ldots, 10), \mu_0$
and the sample size is 100. The estimations of all the parameters under various setting
ably close to the true parameters. In particular, the standard errors calculated by (16) a
standard deviations for $\hat{\gamma}$. For example, in the first setting of Table 1, when the mod
specified, the standard deviations for $\hat{\gamma}$ from 250 simulations are 0.55 and 0.40, while
errors by (2.16) are 0.52 and 0.40, respectively. Additional estimation results for d
based on Model (2.2) and (2.19) with time-varying treatment effects are shown in W
Web Appendix B.

Next, we compare the estimation performances for different sample sizes. For mod
time effect, we generate data with sample sizes of 50, 100, and 200, and obtain the est
the parameters. In Figure 1, we show the box plots for the estimations of the paramet
and α_1, where the true values are 1, 2, 1, and 1, respectively. Figure 1 indicates clearly
sample sizes increase from 50 to 200, the variances of the estimators get smaller, and
the estimators get closer to the true parameters. This implies that when the mode
specified, the proposed estimators are consistent.

est	9.99	10.02	9.98	9.99	9.99	−1.11	1.08
sd	0.29	0.27	0.31	0.29	0.31	0.55	0.40
s.e.	0.29	0.30	0.29	0.30	0.30	0.52	0.40
			μ_0				α
true	0.00	0.00	0.00	0.00	0.00	−1.00	1.00
est	−0.01	0.01	0.00	−0.03	0.00	−0.99	1.00
sd	0.31	0.30	0.28	0.29	0.29	0.09	0.06
s.e.	0.29	0.30	0.30	0.30	0.30	0.09	0.05
				AR(1) with $\rho = 0.8$			
			μ_1				γ
true	10.00	10.00	10.00	10.00	10.00	−1.00	1.00
est	9.99	10.01	9.98	9.99	9.99	−1.11	1.08
sd	0.30	0.30	0.31	0.30	0.32	0.55	0.40
s.e.	0.34	0.32	0.33	0.33	0.33	0.59	0.47
			μ_0				α
true	0.00	0.00	0.00	0.00	0.00	−1.00	1.00
est	−0.01	0.00	0.01	−0.03	−0.00	−0.99	1.00
sd	0.31	0.32	0.30	0.30	0.30	0.15	0.09
s.e.	0.27	0.27	0.27	0.26	0.27	0.09	0.05
				Exchangeable with $\rho = 0.2$			
			μ_1				γ
true	10.00	10.00	10.00	10.00	10.00	−1.00	1.00
est	9.99	10.01	9.98	9.99	9.99	−1.11	1.08
sd	0.29	0.27	0.31	0.29	0.31	0.55	0.40
s.e.	0.28	0.27	0.27	0.27	0.27	0.42	0.39
			μ_0				α
true	0.00	0.00	0.00	0.00	0.00	−1.00	1.00
est	−0.01	0.01	0.01	−0.03	0.00	−0.99	1.00
sd	0.31	0.30	0.28	0.29	0.29	0.11	0.07
s.e.	0.29	0.29	0.29	0.30	0.30	0.09	0.06
				Exchangeable with $\rho = 0.8$			
			μ_1				γ
true	10.00	10.00	10.00	10.00	10.00	−1.00	1.00
est	9.98	10.01	9.98	9.99	9.99	−1.11	1.08
sd	0.30	0.29	0.31	0.30	0.32	0.55	0.40
s.e.	0.32	0.31	0.31	0.31	0.32	0.45	0.32
			μ_0				α
true	0.00	0.00	0.00	0.00	0.00	−1.00	1.00
est	−0.02	0.00	0.01	−0.03	−0.01	−0.99	1.00
sd	0.32	0.32	0.31	0.30	0.30	0.16	0.10
s.e.	0.31	0.32	0.32	0.32	0.32	0.08	0.06

In Table 3, the true error in (2.3) is generated from t distributions with degrees of fre respectively, or lognormal distributions. However, we assume that the errors follow a no tion. The estimation for μ's,γ, and α still remains reasonably close to the true parameter when the error follows a t distribution with degrees of freedom 3, we multiply $\sqrt{1/3}$ to e variance remains 1. For the same reason, when the error follows a t distribution with degr 4, we multiply by $\sqrt{1/2}$. For the error of t_3, the estimation of σ^2 assuming a normal dis has mean 1.02; while for the error of t_4, the estimation of σ^2 has mean 0.98. In general, t of the parameters remain reasonable as long as the true error is a t distribution with a su degree of freedom. Meanwhile, when the error follows a log-normal distribution with pa

true	10.00	10.00	10.00	10.00	10.00	−1.00	1.00
est	9.96	9.99	9.96	9.97	9.98	−1.11	1.08
sd	0.31	0.30	0.32	0.30	0.32	0.55	0.40
s.e.	0.33	0.32	0.32	0.32	0.32	0.47	0.29
			μ_0				α
true	0.00	0.00	0.00	0.00	0.00	−1.00	1.00
est	0.00	0.02	0.02	−0.01	0.01	−0.99	1.00
sd	0.32	0.33	0.32	0.31	0.31	0.16	0.10
s.e.	0.30	0.30	0.31	0.31	0.30	0.09	0.06
				AR(1) with $\rho = 0.8$			
			μ_1				γ
true	10.00	10.00	10.00	10.00	10.00	−1.00	1.00
est	9.99	10.01	9.98	9.99	9.99	−1.11	1.08
sd	0.30	0.30	0.31	0.30	0.32	0.55	0.40
s.e.	0.33	0.34	0.33	0.34	0.33	0.54	0.44
			μ_0				α
true	0.00	0.00	0.00	0.00	0.00	−1.00	1.00
est	−0.01	0.00	0.01	−0.03	−0.00	−0.99	1.00
sd	0.31	0.32	0.30	0.30	0.30	0.15	0.09
s.e.	0.30	0.30	0.30	0.29	0.29	0.16	0.10
				Exchangeable with $\rho = 0.2$			
			μ_1				γ
true	10.00	10.00	10.00	10.00	10.00	−1.00	1.00
est	9.99	10.01	9.98	9.99	9.99	−1.11	1.08
sd	0.29	0.27	0.31	0.29	0.31	0.55	0.40
s.e.	0.31	0.30	0.30	0.31	0.30	0.49	0.37
			μ_0				α
true	0.00	0.00	0.00	0.00	0.00	−1.00	1.00
est	−0.01	0.01	0.00	−0.03	0.00	−0.99	1.00
sd	0.31	0.30	0.28	0.29	0.29	0.11	0.07
s.e.	0.30	0.29	0.29	0.28	0.28	0.12	0.08
				Exchangeable with $\rho = 0.8$			
			μ_1				γ
true	10.00	10.00	10.00	10.00	10.00	−1.00	1.00
est	9.96	9.99	9.96	9.97	9.98	−1.11	1.08
sd	0.31	0.30	0.32	0.30	0.32	0.55	0.40
s.e.	0.35	0.36	0.37	0.35	0.36	0.56	0.51
			μ_0				α
true	0.00	0.00	0.00	0.00	0.00	−1.00	1.00
est	0.00	0.02	0.02	−0.01	0.01	−0.99	1.00
sd	0.32	0.33	0.32	0.31	0.31	0.16	0.10
s.e.	0.36	0.34	0.35	0.35	0.35	0.18	0.11

where the value of the parameter (μ, σ) is specified in the table, we rescale the error first
its mean $\exp(\mu + \sigma^2/2)$ and dividing its standard variation $\sqrt{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma}$
ing the data. The slope of γ from 250 datasets has a mean 1.06 and a standard deviation o
the results when the error is generated from a normal distribution in Table 1 of Web S
Material where the same parameter has a mean 1.09 and a standard deviation 0.45 from
The estimations of other parameters are also sound in Table 3 when a lognormal distributi
on the error term. This confirms that the estimation is also robust against different error

Next, we evaluate the performance of our method under the unbalanced design. V
same simulation setting as in Table 2 for Model (2.2) and (2.3) but allow the data

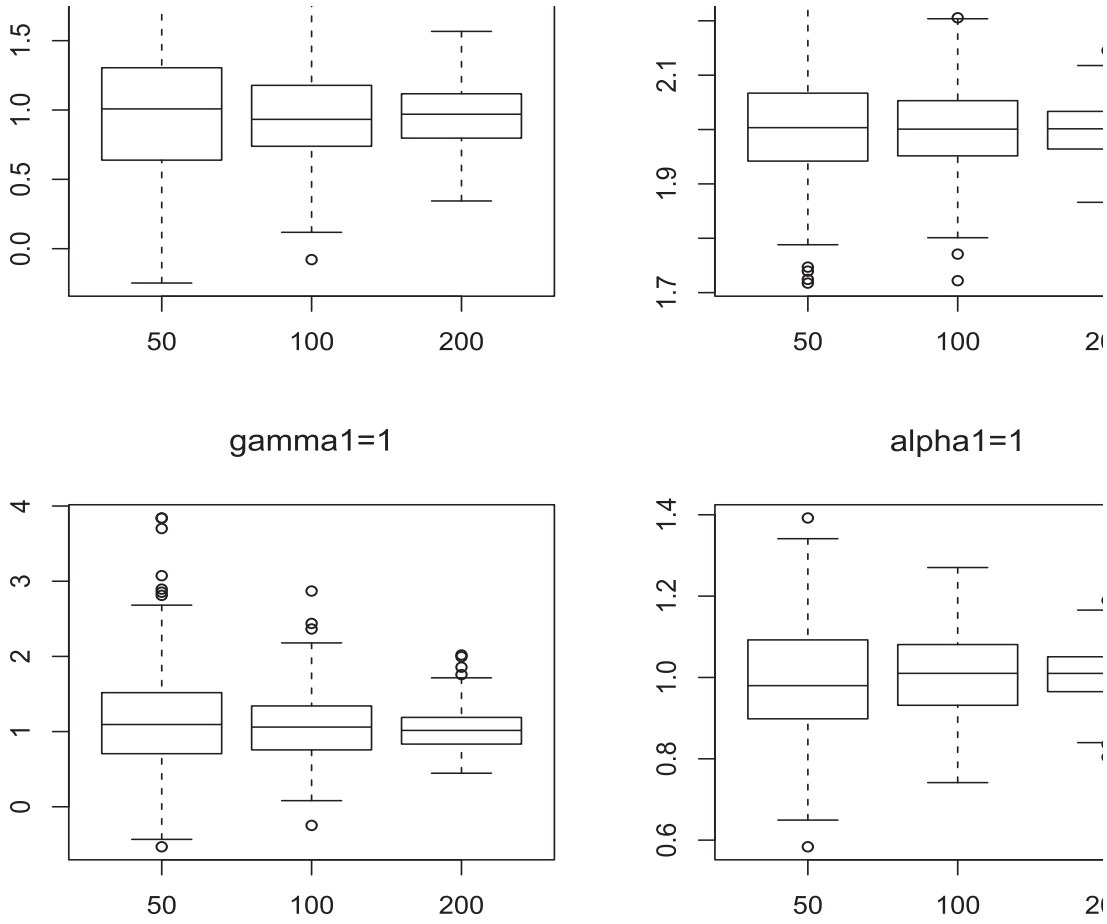

Figure 1. Box plots of the estimates for sample sizes 50, 100, and 200 for model (2.2) and (2.19).

Table 3. The mean and standard deviations of all the parameters from 250 experiments for model (2.2) and the random error follows a t or lognormal distributions.

	μ_{10}	μ_{11}	μ_{00}	μ_{01}	γ		α	
				Random error follows $t_4 \times \sqrt{1/2}$				
true	1	2	0	1	−1	1	−1	1
est	0.94	2.01	0.06	0.99	−1.11	1.10	−1.01	1.01
sd	0.35	0.09	0.37	0.09	0.56	0.46	0.16	0.10
				Random error follows $t_3 \times \sqrt{1/3}$				
true	1	2	0	1	−1	1	−1	1
est	0.95	2.00	0.13	0.98	−1.09	1.08	−1.02	1.01
sd	0.36	0.10	2.15	0.40	0.67	0.45	0.16	0.11
				Random error follows centered lognormal(0,1)				
true	1	2	0	1	−1	1	−1	1
est	0.94	1.99	−0.00	0.98	−1.03	1.06	−1.00	1.01
sd	0.32	0.09	0.33	0.08	0.57	0.42	0.13	0.10
				Random error follows centered lognormal(1,1)				
true	1	2	0	1	−1	1	−1	1
est	0.94	1.99	−0.00	0.98	−1.03	1.06	−1.00	1.01
sd	0.32	0.09	0.33	0.08	0.57	0.42	0.13	0.10

true	10.00	10.00	10.00	10.00	10.00	−1.00	1.00
est	9.98	9.95	9.97	10.03	9.98	−1.12	1.12
sd	0.38	0.38	0.39	0.44	0.55	0.57	0.43
			μ_0			α	
true	0.00	0.00	0.00	0.00	0.00	−1.00	1.00
est	0.04	0.04	0.06	0.01	0.02	−1.00	1.00
sd	0.52	0.57	0.56	0.57	0.71	0.12	0.08
	Exchangeable with $\rho = 0.2$						
			μ_1			γ	
true	10.00	10.00	10.00	10.00	10.00	−1.00	1.00
est	10.00	9.99	9.94	9.93	9.97	−1.07	1.08
sd	0.58	0.57	0.58	0.64	0.67	0.54	0.41
			μ_0			α	
true	0.00	0.00	0.00	0.00	0.00	−1.00	1.00
est	0.01	0.05	0.07	0.06	0.09	−1.00	1.00
sd	0.49	0.51	0.49	0.47	0.56	0.12	0.08

randomly. Specifically, for each subject, we generate 5 observations but only ke observations randomly. The mean and standard deviations of all the parameter est 250 experiments are reported in Table 4, where correlation ρ is estimated via (2.22) anc exchangeable or AR(1) structure, respectively. Table 4 shows that the proposed met reasonably well under the unbalanced data setting.

In longitudinal data analysis, the true correlation structure is often unknown, and it assume a working correlation structure instead. In the following, we compare our esti we apply the true and the working correlation structures, in addition to an empir correlation from the M step. Table 5 provides the means, standard deviations, and errors (MSE) from 250 simulations for the model without time interaction under the tr structures of AR(1) or exchangeable with $\rho = 0.8$. In general, the MSE is the smallest correlation structure is incorporated, and the next most efficient estimator is based on estimated correlation structure. For example, when the true correlation structure is AI of the slope estimator $\hat{\gamma}$ is 0.221 under the true correlation structure while the MSE of $\hat{\gamma}$ empirical correlation estimator is applied.

In Table 5, we also provide estimation results if the correlation information is ignored. EM algorithm to obtain parameter estimators under the independent assumption, denot as in the table. Specifically, the MSE's of the parameter estimators assuming independe than twice those incorporating correlation structures. For example, when the true correla is AR(1), the MSE of $\hat{\gamma}$ assuming independence increases to 0.379. Therefore, it is incorporate correlation information to achieve consistent and efficient estimation for lon

4. Data analysis

In this section, we analyze longitudinal data from the Aids Clinical Trials Grou (ACTG320). In particular, we are interested in identifying a subpopulation which from the treatment. The CD4 cell counts are measured at the baseline, and the 4, 8 week. The response variable is the CD4 change at the 4, 8, 24, and 40th week, an cd4.4, cd4.8, cd4.24, and cd4.40, respectively. The relevant covariates include baselir acid (*rna*.0), which is similar to DNA and plays essential roles in coding, decoding, re expression of genes. Other baselines are CD4 (cd4.0) and age. The summary statistic

		true	1	2	0	1	−1	1	−1	
					AR(1) with $\rho = 0.8$					
TRU	est		0.967	1.997	0.017	1.000	−1.120	1.095	−0.993	1
	sd		0.350	0.095	0.349	0.090	0.584	0.443	0.152	0
	MSE		0.147	0.034	0.147	0.033	0.366	0.221	0.048	0
Est	est		0.969	1.997	0.013	1.001	−1.123	1.096	−0.991	1
	sd		0.351	0.095	0.348	0.090	0.587	0.447	0.152	0
	MSE		0.149	0.035	0.148	0.034	0.371	0.226	0.049	0
Ind	est		1.025	1.988	−0.035	1.007	−1.097	1.073	−0.995	1
	sd		0.352	0.097	0.347	0.092	0.586	0.441	0.150	0
	MSE		0.142	0.027	0.138	0.026	0.361	0.213	0.041	0
MAR	est		1.223	1.942	−0.216	1.052	−0.926	0.917	−0.991	0
	sd		0.415	0.148	0.390	0.150	0.597	0.489	0.151	0
	MSE		0.312	0.162	0.292	0.162	0.496	0.379	0.163	0
					Exchangeable with $\rho = 0.8$					
TRU	est		0.953	2.001	0.029	0.999	−1.121	1.094	−0.996	1
	sd		0.326	0.079	0.328	0.075	0.585	0.447	0.166	0
	MSE		0.133	0.033	0.134	0.032	0.369	0.227	0.054	0
Est	est		0.954	2.001	0.026	0.999	−1.121	1.094	−0.995	1
	sd		0.326	0.079	0.329	0.075	0.586	0.448	0.167	0
	MSE		0.133	0.033	0.135	0.033	0.370	0.228	0.055	0
Ind	est		1.037	1.986	−0.056	1.012	−1.086	1.063	−0.994	1
	sd		0.327	0.083	0.323	0.078	0.583	0.445	0.164	0
	MSE		0.125	0.025	0.122	0.024	0.357	0.215	0.044	0
MAR	est		1.230	1.956	−0.212	1.039	−1.013	0.999	−0.984	0
	sd		0.386	0.083	0.388	0.078	0.599	0.448	0.156	0
	MSE		0.275	0.133	0.277	0.132	0.485	0.327	0.150	0

sizes of data are provided in Web Table 2 in Web Appendix C. Specifically, there are 1(the beginning of the study, and the sample size drops to 564 by the end of the stuc include the patients who have completed observations during the entire study, the samp However, our method allows unbalanced data, therefore, we include the patients with time-point observations, that is, those who drop out at least after the 24th week, and t increasing sample size of 819. The sample sizes from the control and treatment gro similar among the patients. In addition, we apply log transformation on the RNA and in the model and standardize the response variable.

We denote X as covariates of baselines, such as age, cd4.0, and rna.0. The covariate the intercept and time variable. Through the EM-type algorithm described in Section we investigate the time-invariant model based on the complete data (sample size unbalanced data (sample size 819). The corresponding estimators of the parameters ar Web Table 3 in Web Appendix C.

Under the time-invariant model assumption of the random effects, the subgrou depends on the covariates through a logistic model, and the corresponding estimated c provided as follows:

$$\text{logit}(\text{P}(\delta_\text{i}= 1|\text{X}_\text{i}, \text{Z}_\text{i}))= -10.820 - 0.154\text{Age} + 0.563 \log \text{ cd4.0} + 1.410 \log_{10} \text{rna.}$$

Given $\delta_i = 1$, the treatment effect b_i is normal with means of $1.749, 2.152, 2.204$, and 2 $4, 8, 24$ and 40; and given $\delta_i = 0$, the treatment effect b_i is normal with means of 0.055 and 0.563, respectively, which are significantly smaller than those given $\delta_i = 1$. In add and b_i,

$$Y_i|(\delta_i, X_i, Z_i, T_i, b_i) = -0.138 - 0.086\text{Time} + T_i b_i + \varepsilon_i,$$

and rna.0 corresponding to the logistic model (4.1) are 0.183, 0.181, and 0.393, respe
logistic model (4.1), the estimated coefficients of cd4.0 and rna.0 are 0.563 and 1.4
more than twice their corresponding standard errors 0.181 and 0.393, while the estimat
of age is -0.154 with standard error 0.183, indicating that cd4.0 and rna.0 are mo
determining subgroup memberships. Based on (4.1), those with higher baseline scores
rna.0 are more likely to be categorized in the subgroup with more enhanced treatm
compared to other patients. In particular, the treatment effects from the beneficial
1.749, 2.152, 2.204, and 2.092 at weeks $4, 8, 24$ and 40, respectively.

When we add the incomplete samples with the sample size of 819, the estimated m

$$\text{logit}(\text{P}(\delta_i = 1|X_i, Z_i)) = -16.185 - 0.538\text{Age} + 0.551 \log \text{cd4.0} + 2.171 \log_{10} \text{rna}$$

Given $\delta_i = 1$, the treatment effect b_i is normal with means of 2.671, 2.923, 2.992,
weeks $4, 8, 24$ and 40; and given $\delta_i = 0$, the treatment effect b_i is normal wi
0.238, 0.356, 0.624, and 0.717, respectively, which are significantly smaller than
$\delta_i = 1$. In addition, given δ_i and b_i,

$$Y_i|(\delta_i, X_i, Z_i, T_i, b_i) = -0.212 - 0.027\text{Time} + T_i b_i + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, 0.500R)$, where R is treated as a working correlation matrix, similar
standard errors of the coefficient estimators are summarized in Web Table 3 in Web

For both models, we conclude that the subjects with higher baseline CD4 and R
likely to benefit from the treatment. To examine our method more thoroughly, we ran
with respect to the probability with which subjects would benefit more from the treat
the top-benefitting subjects to form a targeted group, and calculate the treatment e
score functions S_1 as in (4.1), and S_2 as in (4.3) from the logistic models utilizing all tin
S_3 as in Shen and He (2015) where only the responses at the 24th week are used. Tha

$$S_i = \beta_{1i}\text{Age} + \beta_{2i} \log \text{cd4.0} + \beta_{3i} \log_{10} \text{rna.0}, \quad i = 1, 2,$$

where the estimators of β_{1i}, β_{2i} and β_{3i} in S_i's and their standard errors are listed in W
Web Appendix C.

A higher score on S_i corresponds to having a better chance to receive benefit from t
Consequently, we calculate the scores in (4.5) for all subjects and rank them by the
different quantiles $q = 0.1, 0.11, \ldots, 0.9$, we select the subjects among the top $(1 - q)$ >
to form a target subgroup, and calculate the treatment effect for this selected subgrou
the same time point from where S_3 is calculated. Figure 2 presents the treatment
selected subgroups for different quantile q's using three different score functions
Figure 2 shows that when q is less than 0.5, that is, when we select a relatively large-si
with enhanced treatment effect, the three methods provide subgroups with simi
treatment effects. As q gets larger, the treatment effects of the target subgroups form
circles) and S_2 (blue crossings), are clearly above those of S_3 (green solid circles). In
score functions from S_1 and S_2 from the proposed model using longitudinal data cho
with better treatment effects than S_3. When $q > 0.5$, the subgroups formed by S_2 (blue c
higher treatment effects than those formed by S_1 (red circles) mostly (about 80% times)
more samples.

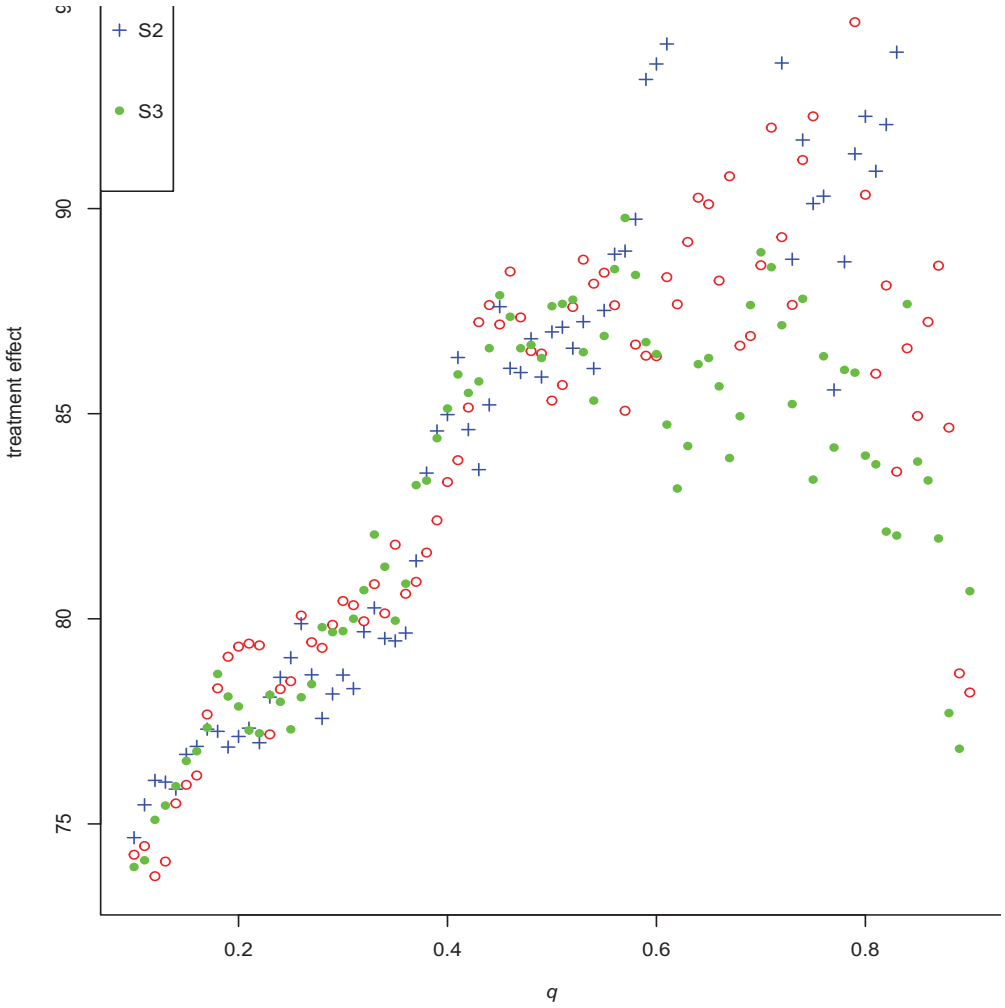

Figure 2. Treatment effects in selected subgroups from top scores by three different score functions S_1, S_2 anc

5. Discussion

In this paper, we propose a structured mixed-effects model for subgroup analysis or observations. In particular, we model the treatment effect as a random effect from a tw mixture normal model, where the proportions of the mixture model are modeled us model depending on some covariates. Through this structured model, we can simultar the subgroup membership and the distribution of the response within the subgroups. EM-type algorithm to obtain parameter estimation.

For balanced data, we can estimate the correlation matrix R from the M step directl iteration. For unbalanced data, we assume correlation structures with an unknown param AR(1) or exchangeable, where ρ is obtained through moment estimation in the iterations, each individual correlation matrix can be calculated based on the estimated ρ. The pro address problems concerning both time-invariant treatment effect as in Section 2.1, a varying treatment effect as in Section 2.3. In addition, the time-varying model can be cases where the treatment effect is linearly associated with other covariates besides time.

In our approach, we only consider the normal distribution of response for ea However, our method is capable of extending to binomial or Poisson distribution in e

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (1150
11690012), and the National Science Foundation (DMS 1613190 and DMS 182119). In addition,
very grateful for constructive comments and suggestions from Associate Editor and referees.

Funding

This work was partially supported by the National Natural Science Foundation of China (1150
11690012), and the National Science Foundation (DMS 1613190 and DMS 182119).

ORCID

Annie Qu http://orcid.org/0000-0002-8396-7828

References

Altstein, L. L., G. Li, and R. M. Elashoff. 2011. A method to estimate treatment efficacy among late
a randomized clinical trial. *Statistics in Medicine* 30 (7):709–717. doi:10.1002/sim.v30.7.

Berger, J. O., X. Wang, and L. Shen. 2014. A Bayesian approach to subgroup identificat
Biopharmaceutical Statistics 24 (1):110–129. doi:10.1080/10543406.2013.856026.

Cacho, A., W. Yao, and X. Cui. 2018. Base-calling using a random effects mixture model on
sequencing data. *Statistics in Biosciences* 10:3–19. doi:10.1007/s12561-017-9190-3.

Cai, T., L. Tian, P. H. Wong, and L. J. Wei. 2011. Analysis of randomized comparative clinic
personalized treatment selections. *Biostatistics* 12 (2):270–282. doi:10.1093/biostatistics/kxq060.

Celeux, G., O. Martin, and C. Lavergne. 2005. Mixture of linear mixed models for clustering gene ex
from repeated microarray experiments. *Statistical Modelling* 5:1–25. doi:10.1191/1471082X05st096

Fan, A., R. Song, and W. Lu. 2017. Change-plane analysis for subgroup detection and sample size cal
of the American Statistical Association 112 (518):769–778. doi:10.1080/01621459.2016.1166115.

Foster, J. C., J. M. Taylor, and S. J. Ruberg. 2011. Subgroup identification from randomized clinical tri
in Medicine 30 (24):2867–2880. doi:10.1002/sim.4322.

Fruhwirth-Schnatter, S. 2006. *Finite mixture and markov switching models*. New York, NY: Springer.

Hall, D. B., and L. Wang. 2005. Two-component mixtures of generalized linear mixed effects mo
correlated data. *Statistical Modelling* 5:21–37. doi:10.1191/1471082X05st090oa.

Hammer, S. M., K. E. Squires, M. D. Hughes, J. M. Grimes, L. M. Demeter, J. S. Currier, J. J. Eron
H. H. Balfour, L. R. Deyton et al. 1997. A controlled trial of two nucleoside analogues plus indinavir
human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. *T
Journal of Medicine*. 337(11):725–733. doi:10.1056/NEJM199709113371101.

Hu, J., W. L. Leite, and M. Gao. 2017. An evaluation of the use of covariates to assist in class enum
growth mixture modeling. *Behavior Research Methods* 49:1179–1190. doi:10.3758/s13428-016-0778

Huang, D., M.-L. Brecht, M. Hara, and Y.-I. Hser. 2010. Influences of covariates on growth mixture m
of Drug Issues 40:173–194. doi:10.1177/002204261004000110.

Jordan, M. I., and R. A. Jacobs. 1994. Hierarchical mixtures of experts and the EM algorithm. *Neura
(2):181–214. doi:10.1162/neco.1994.6.2.181.

Komarek, A., B. E. Hansen, E. M. Kuiper, H. R. van Buuren, and E. Lesaffre. 2010. Discriminan
a multivariate linear mixed model with a normal mixture in the random effects distribution. *Statis
29:3267–3283. doi:10.1002/sim.3849.

Li, M., and J. R. Harring. 2017. Investigating approaches to estimating covariate effects in growth mi
A simulation study. *Educational and Psychological Measurement* 77:766–791. doi:10.1177/0013164

Liang, K. Y., and S. L. Zeger. 1986. Longitudinal data analysis using generalized linear models. *Biom
doi:10.1093/biomet/73.1.13.

Statistics.
Lipkovich, I., A. Dmitrienko, J. Denne, and G. Enas. 2011. Subgroup identification based on differentia A recursive partitioning method for establishing response to treatment in patient subpopulatic *Medicine* 30 (21):2601–2621. doi:10.1002/sim.4289.
Louis, T. A. 1982. Finding the observed information matrix when using the EM algorithm. *Jourr Statistical Society, Series B: Statistical Methodology* 44 (2):226–233.
Lu, X., and Y. Huang. 2014. Bayesian analysis of nonlinear mixed-effects mixture models for longitu heterogeneity and skewness. *Statistics in Medicine* 33:2830–2849. doi:10.1002/sim.6136.
Muthen, B., C. H. Brown, K. Masyn, B. Jo, S.-T. Khoo, -C.-C. Yang, C.-P. Wang, S. G. Kellam, J. B. Ca 2002. General growth mixture modeling for randomized preventive interventions. *Biostati.* doi:10.1093/biostatistics/3.4.459.
Muthen, B., and T. Asparouhov. 2009. Multilevel regression mixture analysis. *Journal of the Royal Si Series A: Statistics in Society* 172 (3):639–657. doi:10.1111/j.1467-985X.2009.00589.x.
Ng, S. K., G. J. McLachlan, K. Wang, L. B. Jones, and S. Ng. 2006. A mixture model with random-effe for clustering correlated gene-expression profiles. *Bioinformatics* 22 (14):1745–1752. doi:10.1093 btl165.
Nylund-Gibson, K., and K. E. Masyn. 2016. Covariates and mixture modeling: Results of a simulation the impact of misspecified effects on class enumeration. *Structural Equation Modeling: A Multidisc* 23:782–797. doi:10.1080/10705511.2016.1221313.
Proust-Lima, C., J.-F. Dartigues, and H. Jacqmin-Gaddaa. 2016. Joint modeling of repeated multiv measures and competing risks of dementia and death: A latent process and latent class approa *Medicine* 35:382–398. doi:10.1002/sim.v35.3.
Proust-Lima, C., V. Philipps, and B. Liquet. 2017. Estimation of extended mixed models using latent c processes: The R package LCMM. *Journal of Statistical Software* 78:1–56. doi:10.18637/jss.v078.i02
Shen, J., and X. He. 2015. Inference for subgroup analysis with a structured logistic-normal mixture n *the American Statistical Association* 110 (509):303–312. doi:10.1080/01621459.2014.894763.
Shen, J., Y. Wang, and X. He. 2017. Penalized likelihood for logistic-normal mixture models with un *Statistica Sinica* 27:711–731.
Song, Y., and G. Y. Chi. 2007. A method for testing a pre-specified subgroup in clinical trials. *Statistic* (19):3535–3549. doi:10.1002/sim.2825.
Su, X., C. L. Tsai, H. Wang, D. M. Nickerson, and B. Li. 2009. Subgroup analysis via recursive partitic *Machine Learning Research* 10:141–158.
Tang, X., and A. Qu. 2016. Mixture modeling for longitudinal data. *Journal of Computational and Gr* 25:1117–1137. doi:10.1080/10618600.2015.1092979.
Turner, T. R. 2000. Estimating the propagation rate of a viral infection of potato plants via mixture *Journal of the Royal Statistical Society, Series C: Applied Statistics* 49:371–384. doi:10.1111/1467-98
Verbeke, G., and E. Lesaffre. 1996. A linear mixed-effects model with heterogeneity in the random-eff *Journal of the American Statistical Association* 91 (433):217–221. doi:10.1080/01621459.1996.10476
Wu, R., M. Zheng, and W. Yu. 2016. Subgroup analysis with time-to-event data under a logistic-cox *Scandinavian Journal of Statistics* 43 (3):863–878. doi:10.1111/sjos.v43.3.
Xu, W., and D. Hedeker. 2001. A random-effects mixture model for classifying treatment response clinical trials. *Journal of Biopharmaceutical Statistics* 11:253–273. doi:10.1081/BIP-120008848.
Yuksel, S. E., J. N. Wilson, and P. D. Gader. 2012. Twenty years of mixture of experts. *IEEE Transa Networks and Learning Systems* 23 (8):1177–1193. doi:10.1109/TNNLS.2012.2200299.
Zhao, L., L. Tian, T. Cai, B. Claggett, and L. J. Wei. 2013. Effectively selecting a target populati comparative study. *Journal of the American Statistical Association* 108 (502):527–53 01621459.2013.770705.