Title

A multilevel mixed effects varying coefficient model with multilevel predictors and random effects for modeling hospitalization risk in patients on dialysis.

Journal

Biometrics, 76(3)

ISSN

0006-341X

Authors

Li, Yihao
Nguyen, Danh V
Kürüm, Esra
et al.

Publication Date

2020-09-01

DOI

10.1111/biom.13205

HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2020 September 12.

Published in final edited form as:

Biometrics. 2020 September ; 76(3): 924–938. doi:10.1111/biom.13205.

A Multilevel Mixed Effects Varying Coefficient Model with Multilevel Predictors and Random Effects for Modeling Hospitalization Risk in Patients on Dialysis

Yihao Li[1], **Danh V. Nguyen**[2], **Esra Kürüm**[3], **Connie M. Rhee**[2], **Yanjun Chen**[4], **Kamyar Kalantar-Zadeh**[2], **Damla Şentürk**[1,*]

[1]Department of Biostatistics, University of California, Los Angeles, CA 90095, U.S.A

[2]Department of Medicine, University of California, Irvine, Orange, CA 92868, U.S.A.

[3]Department of Statistics, University of California, Riverside, CA 92521, U.S.A.

[4]Institute for Clinical and Translational Science, Irvine, CA 92697, U.S.A.

Summary:

For patients on dialysis, hospitalizations remain a major risk factor for mortality and morbidity. We use data from a large national database, United States Renal Data System (USRDS), to model time-varying effects of hospitalization risk factors as functions of time since initiation of dialysis. To account for the three-level hierarchical structure in the data where hospitalizations are nested in patients and patients are nested in dialysis facilities, we propose a multilevel mixed effects varying coefficient model (MME-VCM) where multilevel (patient- and facility-level) random effects are used to model the dependence structure of the data. The proposed MME-VCM also includes multilevel covariates, where baseline demographics and comorbidities are among the patient-level factors, and staffing composition and facility size are among the facility-level risk factors. To address the challenge of high-dimensional integrals due to the hierarchical structure of the random effects, we propose a novel two-step approximate EM algorithm based on the fully exponential Laplace approximation. Inference for the varying coefficient functions and variance components is achieved via derivation of the standard errors using score contributions. The finite sample performance of the proposed estimation procedure is studied through simulations.

Keywords

End-stage renal disease; Hospitalization outcome; Multilevel varying coefficient models; United States Renal Data System

* dsenturk@ucla.edu.

Supporting Information

Web Appendices, referenced in Sections 2.2, 3 and 4.1, are available with this paper at the Biometrics website on Wiley Online Library. We provide R code for running our MME-VCM algorithm on simulated datasets on Github (https://github.com/dsenturk/MME-VCM).

1. Introduction

As of 2016, more than 726,000 individuals in the United States (US) were affected by end-stage renal disease (ESRD). Of those, about 458,000 were on dialysis, a life-sustaining treatment (United States Renal Data System, 2018). On average, a dialysis patient is hospitalized twice a year, and hospitalizations in this population contribute substantially to the cost of patient care. Studying the impact of leading risk factors on the probability of hospitalizations, particularly modifiable factors, will contribute to formulation of hospitalization risk reduction strategies that can improve patient care.

We analyze data from the United States Renal Data System (USRDS) where the data is structured hierarchically: longitudinal hospitalizations are nested within patients and patients are nested within dialysis facilities of which there are thousands across the US. Moreover, both patient-level (e.g., baseline demographics and comorbidities) and facility-level characteristics (e.g., staffing level) are potential risk factors that affect the likelihood of hospitalization. Hence, our modeling needs to account for the hierarchical structure in the data and quantify the effects of both patient- and facility-level risk factors. In addition, for ESRD patients, dialysis is a long-term life-sustaining treatment until death or kidney transplantation. Since patients' needs and clinical characteristics may change as they persist on dialysis, our desired model needs to model hospitalization risk over time and characterize the effects of multilevel risk factors as functions of time from initiation of dialysis.

Varying coefficient models have been used in the study of time-varying regression effects (Cleveland et al., 1991; Hastie and Tibshirani, 1993) and generalized varying coefficient models are useful tools in modeling generalized longitudinal outcomes, including binary and count data (Cai et al., 2000). Wu and Zhang (2002), Zhang (2004) and Chen and Wang (2011) all consider subject-specific random effects in varying coefficient models, which are applicable to only a two-level hierarchy where longitudinal outcomes are nested within subjects. Most of the existing works are not applicable to data with three or more levels in the hierarchy, such as longitudinal measurements nested in subjects and subjects nested in dialysis facilities. The few works that consider a three-level hierarchy do not propose correlation structures that are scalable to large data applications and do not have complex multilevel predictors (You et al., 2015). Li et al. (2018) considers a varying coefficient model for multilevel risk factors (VCM-MR) for three-level hierarchical data; however, only a single-level random effect at the subject-level is included. They account for within-facility correlations through facility-specific varying coefficient functions. Although useful, this leads to a large number of varying coefficient functions to be estimated, increasing the computational burden. As a novel departure from existing literature, we propose a multilevel mixed effects varying coefficient model (MME-VCM) for three-level hierarchical data using a two-level random effects structure. The hierarchical dependence is modeled via hierarchical random effects (at the patient- and facility-levels). In addition, we include multilevel (patient- and facility-level) predictors in the regression model. Inclusion of multilevel predictors allows direct comparisons of time-varying effects of multilevel risk factors. To address truncation of the longitudinal follow-up by death, we propose a partly conditional MME-VCM, modeling time-varying hospitalization risk conditional on the

patients being alive. Similar partly conditional varying coefficient models have been considered in the analysis of USRDS data (Estes et al. 2014, 2018).

Advances in estimation of multilevel models are limited by the severe challenge of high-dimensional random effects. For estimation in the MME-VCM, we propose a two-step approximate EM algorithm that resolves this computational challenge. Briefly, in the E-step, we compute the posterior means and variances of the patient- and facility-level random effects. In the M-step, we maximize the approximated global or local expected log-likelihoods to target the model parameters based on estimates from the E-step. Note that the integrals involved in the E-step need to be computed at the highest level of the hierarchy in the data, i.e., at the facility-level. More specifically, we need to integrate over the facility-level random effect as well as all patient-level random effects nested within a particular facility. That is, if a facility has N_i patients, we need to integrate over N_i patient-level random effects and one facility-level random effect, leading to a $(N_i + 1)$-dimensional integral where traditional numerical integration methods (e.g., Gauss quadrature) are not feasible. This is a rather significant challenge hindering the estimation in multilevel models in general with high-dimensional random effects. To address this problem, we use the Laplace approximation method which has less computational burden than other numerical integration methods such as Gauss quadrature or Monte Carlo approaches. Nevertheless, one limitation of the standard Laplace method is that the error associated with the approximation can get large in sparse longitudinal applications with small number of repeated measurements within subjects. Since the USRDS data has subjects with only a few (<5) repetitions during the follow-up period, we take advantage of the fact that in the proposed EM algorithm the posterior mean and variance of the random effects are in the form of a ratio between two integrals, and apply the fully exponential Laplace approximations (Tierney et al., 1989; Rizopoulos et al., 2009) which lead to lower order approximation errors for computing the integrals with respect to the facility- and patient-level random effects. For inference, the standard errors of the proposed multilevel varying coefficient functions and the variance components are derived using the inverse of the empirical Fisher information matrix computed by score contributions (Tutz and Kauermann, 2003).

Hence, the paper makes novel contributions both in the proposal of MME-VCM, a varying coefficient model with multilevel random effects and multilevel predictors, but also in the proposal of a computationally efficient EM algorithm based on fully exponential Laplace approximations to address the challenge of multilevel random effects. The paper is organized as follows. The proposed MME-VCM formulation, estimation and inference for analyzing multilevel risk factors on dialysis patients' hospitalization risk based on USRDS data are described in Section 2. Extensive simulations are presented in Section 3. We provide the cohort description and data analysis results with interpretations in Section 4, followed by a discussion in Section 5.

2. Proposed Multilevel Mixed Effects Varying Coefficient Model

2.1 Model Specification

Consider a cohort of incident dialysis patients followed over time from initiation of dialysis. Let $i = 1, \ldots, I$, index dialysis facilities; $j = 1, \ldots, N_i$, index subjects belonging to the ith

facility with N_i total subjects; and $k = 1, \ldots, N_{ij}$, index observations on the jth subject receiving dialysis at the ith facility, who has N_{ij} total observations. Subjects' follow-up time is grouped into three month intervals where the outcome is the binary indicator of an all-cause hospitalization in a three month interval. More specifically, the outcome $Y_{ijk} \equiv Y_{ij}(t_{ijk})$ equals one if the jth patient within facility i experiences one or more hospitalizations during the kth three month follow-up interval and equals zero otherwise. The index t denotes time after initiation of dialysis and t_{ijk} is defined as the midpoint of the kth three month interval in subject i's follow-up. The proposed MME-VCM includes both facility-level ($\gamma_i \sim \mathrm{N}(0, \sigma_\gamma^2)$) and subject-level ($b_{ij} \sim \mathrm{N}(0, \sigma_b^2)$) random effects:

$$g[E\{Y_{ij}(t) \mid X_{ij}, Z_{i(j)}, b_{ij}, \gamma_i, S_{ij} > t\}] \equiv g\{p_{ij}(t)\} = X_{ij}^{\mathsf{T}}\beta(t) + Z_{i(j)}^{\mathsf{T}}\theta(t) + b_{ij} + \gamma_i, \quad (1)$$

where g is the logit link function, S_{ij} is the death time of subject j in facility i, and $X_{ij} = (X_{1ij}, \ldots, X_{rij})^{\mathsf{T}}$ and $Z_{i(j)} = \{Z_{1i(j)}, \ldots, Z_{pi(j)}\}^{\mathsf{T}}$ denote the vector of r subject-level and p facility-level predictors with the varying coefficient functions $\beta(t) = \{\beta_1(t), \ldots, \beta_r(t)\}^{\mathsf{T}}$ and $\theta(t) = \{\theta_1(t), \ldots, \theta_p(t)\}^{\mathsf{T}}$, respectively. For our data application, the facility characteristics ($Z_{i(j)}$) are recorded at the end of each calendar year, and for a particular patient, their value will be determined by dialysis facility data from the prior calendar year to their initiation of dialysis. Hence these covariates are indexed not only by the dialysis facility index i, but also by the subject index j. The MME-VCM in (1) is a partly conditional model (Estes et al., 2014, 2016; Kurland and Heagerty, 2005), conditioning on the patients being alive $S_{ij} > t$, instead of their actual survival time. Motivated by the observation that for USRDS data missingness is mainly due to truncation by death, partly conditional target of inference has been considered previously by Estes et al. (2014, 2016) in the context of generalized linear varying coefficient models.

Multilevel modeling with non-time-varying coefficients has been studied extensively (Gelman, 2006) and is considered a tradeoff between complete pooling and no pooling. In the complete pooling framework, data from all facilities is analyzed without any facility-specific regression parameters. On the other hand, no pooling analyzes data within each facility separately and may lead to overfitting, especially for small facilities. In the proposed MME-VCM, the facility data is pooled through the facility-specific predictors $Z_{i(j)}$, leading to stable estimation through partial pooling of information.

2.2 Proposed Estimation and Inference

Estimation in the proposed MME-VCM is hindered by the high dimension of the random effects. Integrating the likelihood with respect to the random effects vector $u_i = (b_{i1}, \ldots, b_{iN_i}, \gamma_i)^{\mathsf{T}}$ within facility i, is a substantial computational challenge since the dimensionality of u_i grows with the number of subjects within a facility (N_i in the 100s for USRDS applications). We propose an approximate EM algorithm which considers the high-dimensional random effects missing and iterates between estimation of the expected value of the complete likelihood (E-step) and its maximization with respect to the multilevel model parameters (M-step). The challenge of the high-dimensional integration in the E-step is resolved by the use of the fully exponential Laplace approximations in estimation of the moments of the random effects. For estimation of the varying coefficient functions

associated with patient- and facility-level predictors in the M-step, a computationally feasible local linear smoothing procedure (Hoover et al., 1998; Wu et al., 1998; Fan and Zhang, 1999) is utilized which only uses local data available in sliding windows of follow-up time.

2.2.1 E-step and the Fully Exponential Laplace Approximations.—

Let $L_{Y_{ij}}\{\beta(t), \theta(t)\}$ denote the joint distribution of the outcome $(Y_{ij1}, \ldots, Y_{ijN_{ij}})$ of the jth subject observed at the time points $t_{ij} = (t_{ij1}, \ldots, t_{ijN_{ij}})$, conditional on u_i, X_{ij}, $Z_{i(j)}$ and $S_{ij} > t_{ij}$, $j = 1, 2, \ldots, N_i$. For mathematical convenience, we assume that the within-subject correlation among $(Y_{ij1}, \ldots, Y_{ijN_{ij}})$ is explained by two independent sources: the random effects u_i and the dependency of Y_{ijk}, $k = 1, \ldots, N_{ij}$, on the patient's death time S_{ij}. Using the independence between u_i and S_{ij}, the complete likelihood within facility i, corresponding to the joint distribution of $\{(Y_{ij1}, \ldots, Y_{ijN_{ij}}, u_i) : j = 1, \ldots, N_i\}$ conditional on X_{ij}, $Z_{i(j)}$ and $S_{ij} > t_{ij}$, can be given as

$$L_i\{u_i, \beta(t), \theta(t), \sigma_b, \sigma_\gamma\} = \prod_{j=1}^{N_i} \left[L_{Y_{ij}}\{\beta(t), \theta(t)\} \times \frac{\exp\{-b_{ij}^2 / (2\sigma_b^2)\}}{\sqrt{2\pi\sigma_b^2}} \right] \times \frac{\exp\{-\gamma_i^2 / (2\sigma_\gamma^2)\}}{\sqrt{2\pi\sigma_\gamma^2}},$$

for normally distributed random effects $u_i = (b_{i1}, \ldots, b_{iN_i}, \gamma_i)^\intercal$. Hence the total complete likelihood is $L\{u_1, \ldots, u_I, \beta(t), \theta(t), \sigma_b, \sigma_\gamma\} = \prod_{i=1}^{I} L_i\{u_i, \beta(t), \theta(t), \sigma_b, \sigma_\gamma\}$. In addition, viewing the random effects as missing, the incomplete likelihood available for estimation of $\{\beta(t), \theta(t), \sigma_b, \sigma_\gamma\}$ is $L\{\beta(t), \theta(t), \sigma_b, \sigma_\gamma\} = \prod_{i=1}^{I} [\int L_i\{u_i, \beta(t), \theta(t), \sigma_b, \sigma_\gamma\} du_i]$.

Assuming that the within-subject correlation introduced by the dependence on death time S_{ij} is weak compared to that introduced by the random effect u_i, we can approximate the joint likelihood $L_{Y_{ij}}\{\beta(t), \theta(t)\}$ conditional on u_i utilizing the working independence assumption:

$$L_{Y_{ij}}\{\beta(t), \theta(t)\} \approx \prod_{k=1}^{N_{ij}} \frac{\exp[\{X_{ij}^\intercal \beta(t_{ijk}) + Z_{i(j)}^\intercal \theta(t_{ijk}) + b_{ij} + \gamma_i\} Y_{ijk}]}{1 + \exp\{X_{ij}^\intercal \beta(t_{ijk}) + Z_{i(j)}^\intercal \theta(t_{ijk}) + b_{ij} + \gamma_i\}}.$$

The working independence assumption will also be used in the M-step of the proposed EM algorithm to estimate the partly conditional target. Kurland and Heagerty (2005) point out that a standard likelihood-based method or an estimating equation approach without a working independence structure will not lead to valid inference for a partly conditional target.

The variance components σ_b and σ_γ can be estimated via maximizing the incomplete likelihood directly, but the closed form solutions for maximizing the incomplete likelihood with respect to $\beta(t)$ and $\theta(t)$ are not available. Hence, we propose an approximate EM algorithm, where the expectation step targets the approximate conditional expectation of the complete likelihood by utilizing a Taylor's expansion. Then the maximization step optimizes the approximate expected likelihood with respect to $\beta(t)$ and $\theta(t)$.

For the expectation step, the posterior mean and variance of u_i, denoted by u_{i0} and v_{i0}, respectively, are

$$u_{i0} = \frac{\int u_i L_i\{u_i, \beta(t), \theta(t), \sigma_b, \sigma_\gamma\} du_i}{\int L_i\{u_i, \beta(t), \theta(t), \sigma_b, \sigma_\gamma\} du_i},\ v_{i0} = \frac{\int (u_i - u_{i0})(u_i - u_{i0})^\top L_i\{u_i, \beta(t), \theta(t), \sigma_b, \sigma_\gamma\} du_i}{\int L_i\{u_i, \beta(t), \theta(t), \sigma_b, \sigma_\gamma\} du_i} \quad (2)$$

To evaluate the high-dimensional integrals in (2), we will utilize the fully exponential Laplace approximation proposed by Tierney et al. (1989). Since the integrand in (2), namely u_i, may not always be positive, and the fully exponential Laplace approximation is applied only to strictly positive functions, we follow Rizopoulos et al. (2009) in targeting $E\{\exp(c^\top u_i)\}$, which is always positive. Here $c = (c_1, \ldots, c_{N_i+1})^\top$ denotes a $(N_i + 1) \times 1$ constant vector. We obtain the required expectations through differentiating the cumulant-generating function, defined by $\log[E\{\exp(c^\top u_i)\}]$, via $u_{i0} = \partial \log[E\{\exp(c^\top u_i)\}]/\partial c^\top|_{c=0}$ and $v_{i0} = \partial^2 \log[E\{\exp(c^\top u_i)\}]/\partial c^\top \partial c|_{c=0}$. The approximations are computed in two steps, where in the first step the complete likelihood is maximized via Newton-Raphson to obtain the mode of u_i. The second step uses the mode to obtain u_{i0} and v_{i0} through approximation of the cumulant-generating function.

Let $\ell_i(u_i, \alpha)$ denote the log of $L_i\{u_i, \beta(t), \theta(t), \sigma_b, \sigma_\gamma\}$ where α denotes the set of model parameters $\{\beta(t), \theta(t), \sigma_b, \sigma_\gamma\}$ for notational convenience. Further let $\hat{u}_i^{(c)} = \text{argmax}_{u_i}\{\ell_i(u_i, \alpha) + c^\top u_i\}$. The mode $\hat{u}_i = \hat{u}_i^{(c)}|_{c=0}$ is estimated by maximizing the log-likelihood $\ell_i(u_i, \alpha)$ with safeguarded Newton-Raphson algorithm according to $\hat{u}_i^{\text{it}+1} = \hat{u}_i^{\text{it}} - s(\Sigma_i^{\text{it}})^{-1} J(\hat{u}_i^{\text{it}})$, where 'it' denotes the iteration number, $J(\hat{u}_i^{\text{it}}) = -\partial \ell_i(u_i, \alpha) / \partial u_i^\top|_{u_i = \hat{u}_i^{\text{it}}}$, $\Sigma_i^{\text{it}} = \Sigma_i^{(c)}|_{(c, u_i) = (0, \hat{u}_i^{\text{it}})}$, $\Sigma_i^{(c)} = -\partial^2\{\ell_i(u_i, \alpha) + c^\top u_i\} / \partial u_i^\top \partial u_i = -\partial^2 \ell_i(u_i, \alpha) / \partial u_i^\top \partial u_i$ and s is the step size along the Newton-Raphson updating direction. Using the estimated mode $\hat{u}_i$, the approximated posterior mean and variance of u_i are obtained by differentiating the cumulant-generating function and evaluating at $c = 0$,

$$u_{i0} = \hat{u}_i - \frac{1}{2}\text{tr}(\mathscr{V}), \quad v_{i0} = \Sigma_i^{-1} - \frac{1}{2}\text{tr}\left\{-\mathscr{V}\mathscr{V}^\top + \Sigma_i^{-1}\frac{\partial^2 \Sigma_i^{(c)}}{\partial c^\top \partial c}|_{(c, u_i) - (0, \hat{u}_i)}\right\}, \quad (3)$$

where $\mathscr{V} = \Sigma_i^{-1}\{\partial \Sigma_i^{(c)} / \partial c^\top\}|_{(c, u_i) = (0, \hat{u}_i)}$, $\Sigma_i = \Sigma_i^{(c)}|_{c=0}$ and the values of $\hat{u}_i$ and Σ_i^{-1} in (3) are obtained from the last iteration of the Newton-Raphson algorithm for finding the mode (i.e. $\hat{u}_i^{\text{it}}$ and $(\Sigma_i^{\text{it}})^{-1}$). Details of derivatives of $\ell_i(u_i, \alpha)$ with respect to u_i as well as the explicit forms of the correction terms $\text{tr}(\mathscr{V}) / 2$ and $\text{tr}\{-\mathscr{V}\mathscr{V}^\top + \Sigma_i^{-1}\partial^2 \Sigma_i^{(c)} / \partial c^\top \partial c|_{(c, u_i) = (0, \hat{u}_i)}\} / 2$ in (3) are deferred to Web Appendix A.

For approximating the conditional expectation of the complete likelihood in the E-step, let α^* denote the set of current parameter estimates $\{\beta^*(t), \theta^*(t), \sigma_b^*, \sigma_\gamma^*\}$, and

$u_{i0}^* = (b_{i10}^*, \ldots, b_{iN_i0}^*, \gamma_{i0}^*)^\top$, $v_{b,ij0}^*$, $v_{\gamma,i0}^*$ and r_{ij0}^* denote the estimated posterior mean of u_i, posterior variances of b_{ij} and γ_i, and the posterior covariance of b_{ij} and γ_i based on the current parameter estimates, respectively. Because the closed form for $\sum_{i=1}^{I} E\{\ell_i(u_i, \alpha) \mid Y_i, \alpha^*, S_{ij} > t_{ij}\}$ is not available, we use the second order Taylor series expansion to approximate the expected log-likelihood around u_{i0}^* by

$$\begin{aligned}
&\sum_{i=1}^{I} \ell_i(u_{i0}^*, \alpha^*) + \ell_i'(u_{i0}^*, \alpha^*) E(u_i - u_{i0}^*) - \frac{1}{2} E(u_i - u_{i0}^*)^\top \Sigma_i^* (u_i - u_{i0}^*) \\
&= \sum_{i=1}^{I} \left\{ \sum_{j=1}^{N_i} \left(\sum_{k=1}^{N_{ij}} \left[Y_{ijk}\{g(p_{0,ijk}^*)\} + \log(q_{0,ijk}^*) - \frac{v_{b,ij0}^* + 2r_{ij0}^* + v_{\gamma,i0}^*}{2} p_{0,ijk}^* q_{0,ijk}^* \right] \right. \right. \\
&\left. \left. - \frac{(b_{ij0}^*)^2 + v_{b,ij0}^*}{2(\sigma_b^*)^2} - \frac{1}{2}\log\{2\pi(\sigma_b^*)^2\} \right) - \frac{(\gamma_{i0}^*)^2 + v_{\gamma,i0}^*}{2(\sigma_\gamma^*)^2} - \frac{1}{2}\log\{2\pi(\sigma_\gamma^*)^2\} \right\},
\end{aligned} \tag{4}$$

where $p_{0,ijk}^* = g^{-1}\{X_{ij}^\top \beta^*(t_{ijk}) + Z_{i(j)}^\top \theta^*(t_{ijk}) + b_{ij0}^* + \gamma_{i0}^*\}$, $q_{0,ijk}^* = 1 - p_{0,ijk}^*$, $\Sigma_i^* = \Sigma_i \mid_{u_i = u_{i0}^*, \alpha = \alpha^*}$, $\ell'(u_{i0}^*, \alpha^*) = \partial \ell_i(u_i, \alpha) / \partial u_i^\top \mid_{u_i = u_{i0}^*, \alpha = \alpha^*}$, and $E(u_i - u_{i0}^*) = 0$.

2.2.2 M-step and Estimation of the Standard Errors.—

The steps of the proposed EM algorithm are as follows:

1. Initialize the estimates for model parameters and denote them by $\beta^{(0)}(t)$, $\theta^{(0)}(t)$, $\sigma_b^{(0)}$ and $\sigma_\gamma^{(0)}$.
2. (E-step) In the mth iteration, update posterior means and variances of random effects u_i using the fully exponential Laplace approximation and current parameter estimates $\beta^{(m-1)}(t)$, $\theta^{(m-1)}(t)$, $\sigma_b^{(m-1)}$ and $\sigma_\gamma^{(m-1)}$. Let $b_{ij0}^{(m)}$, $v_{b,ij0}^{(m)}$, $\gamma_{i0}^{(m)}$, $v_{\gamma,i0}^{(m)}$ and $r_{ij0}^{(m)}$ denote the estimated posterior mean and variance of b_{ij}, γ_i and the posterior covariance of b_{ij} and γ_i, respectively.
3. (M-step) Maximize the incomplete log-likelihood with respect to σ_b and σ_γ to obtain $\sigma_b^{(m)}$ and $\sigma_\gamma^{(m)}$ based on $b_{ij0}^{(m)}$, $\gamma_{i0}^{(m)}$, $v_{b,ij0}^{(m)}$ and $v_{\gamma,i0}^{(m)}$ from step 2.
4. (M-step) At each time point t_0, define $\phi(t) = [\{\beta(t)\}^\top, \{\theta(t)\}^\top]^\top$ and expand it by $\phi(t) \approx \phi_0 + \phi_1(t - t_0)$, where $\beta_0 = (\beta_{01}, \ldots, \beta_{0r})^\top$, $\beta_1 = (\beta_{11}, \ldots, \beta_{1r})^\top$, $\theta_0 = (\theta_{01}, \ldots \theta_{0p})^\top$, $\theta_1 = (\theta_{11}, \ldots, \theta_{1p})^\top$, $\phi_0 = (\beta_0^\top, \theta_0^\top)^\top$ and $\phi_1 = (\beta_1^\top, \theta_1^\top)^\top$. Maximize the approximated expected local log-likelihood with respect to (ϕ_0, ϕ_1) to obtain $(\phi_0^{(m)}, \phi_1^{(m)})$ and set $\phi^{(m)}(t_0) = \phi_0^{(m)}$
5. Iterate between steps 2 and 4 until $\max_{i,j,k} \mid p_{0,ijk}^{(m)} - p_{0,ijk}^{(m-1)} \mid < \epsilon$, where ϵ is a predefined tolerance level and

$p_{0,ijk}^{(m)} = g^{-1}\{X_{ij}^{\top}\beta^{(m)}(t_{ijk}) + Z_{i(j)}^{\top}\theta^{(m)}(t_{ijk}) + b_{ij}^{(m)} + \gamma_i^{(m)}\}$. (For our application, $\epsilon = .001$.)

In the first step described above, the initial values for $\beta(t)$, $\theta(t)$, σ_b and σ_γ are estimated through fitting a non-time-varying multilevel generalized linear model $g[E\{Y_{ijk} \mid X_{ij}, Z_{i(j)}, b_{ij}, \gamma_i\}] = X_{ij}^{\top}\beta + Z_{i(j)}^{\top}\theta + b_{ij} + \gamma_i$ using the glmer function from R package lme4. The $\sigma_b^{(m)}$ and $\sigma_\gamma^{(m)}$ in step 3 are obtained by setting the score functions based on incomplete log-likelihood to zero. Let $\ell(\alpha)$ denote the log of the incomplete likelihood $L(\alpha) = L\{\beta(t), \theta(t), \sigma_b, \sigma_\gamma\}$, where α is used to denote the set of parameters $\{\beta(t), \theta(t), \sigma_b, \sigma_\gamma\}$ as before. The score functions with respect to σ_b and σ_γ can be given as:

$$V(\sigma_b^2) = \frac{\partial \ell(\alpha)}{\partial \sigma_b^2} = \sum_{i=1}^{I} \frac{\partial}{\partial \sigma_b^2} \log\left\{\int L_i(u_i, \alpha) du_i\right\} = \sum_{i=1}^{I} \int \sum_{j=1}^{N_i} \left(\frac{b_{ij}^2}{2\sigma_b^4} - \frac{1}{2\sigma_b^2}\right) D_i(u_i) du_i \equiv \sum_{i=1}^{I} V_i(\sigma_b^2)$$

and

$$V(\sigma_\gamma^2) = \frac{\partial \ell(\alpha)}{\partial \sigma_\gamma^2} = \sum_{i=1}^{I} \frac{\partial}{\partial \sigma_\gamma^2} \log\left\{\int L_i(u_i, \alpha) du_i\right\} = \sum_{i=1}^{I} \int \left(\frac{\gamma_i^2}{2\sigma_\gamma^4} - \frac{1}{2\sigma_\gamma^2}\right) D_i(u_i) du_i \equiv \sum_{i=1}^{I} V_i(\sigma_\gamma^2),$$

where $D_i(u_i) = L_i(u_i, \alpha) / \int L_i(u_i, \alpha) du_i$ is the posterior density of u_i. Setting the score functions $V(\sigma_b^2)$ and $V(\sigma_\gamma^2)$ equal to zero leads to $\sigma_b^{(m)} = ((\sum_{i=1}^{I} N_i)^{-1} \sum_{i=1}^{I} \sum_{j=1}^{N_i} [\{b_{ij0}^{(m)}\}^2 + v_{b,ij0}^{(m)}])^{1/2}$ and $\sigma_\gamma^{(m)} = (I^{-1} \sum_{i=1}^{I} [\{\gamma_{i0}^{(m)}\}^2 + v_{\gamma,i0}^{(m)}])^{1/2}$. For inference, we rely on standard errors obtained from the inverse of the appropriate empirical Fisher information matrix for all model parameters. For $\hat{\sigma}_b^2$ and $\hat{\sigma}_\gamma^2$, the standard errors are equal to the square root of diagonal elements of $(\sum_{i=1}^{I} V_i V_i^{\top})^{-1}$ where $V_i = \{V_i(\hat{\sigma}_b^2), V_i(\hat{\sigma}_\gamma^2)\}^{\top}$.

In step 4, the approximated expected local log-likelihood at t_0 can be given as:

$$\begin{aligned}
&\sum_{i=1}^{I} \sum_{j=1}^{N_i} \sum_{k=1}^{N_{ij}} \Big(Y_{ijk}[X_{ij}^{\top}\{\beta_0 + \beta_1(t_{ijk} - t_0)\} + Z_{i(j)}^{\top}\{\theta_0 + \theta_1(t_{ijk} - t_0)\} + b_{ij0}^{(m)} + \gamma_{i0}^{(m)} \\
&] \\
&+ \log\{\tilde{q}_{\phi,ijk}^{(m)}\} - \frac{v_{b,ij0}^{(m)} + 2r_{ij0}^{(m)} + v_{\gamma,i0}^{(m)}}{2} \tilde{p}_{\phi,ijk}^{(m)} \tilde{q}_{\phi,ijk}^{(m)} - \frac{\{b_{ij0}^{(m)}\}^2 + v_{b,ij0}^{(m)}}{2N_{ij}\{\sigma_b^{(m)}\}^2} - \frac{1}{2N_{ij}} \log \qquad (5) \\
&[2\pi\{\sigma_b^{(m)}\}^2] \\
&- \frac{\{\gamma_{i0}^{(m)}\}^2 + v_{\gamma,i0}^{(m)}}{2(\sum_{j=1}^{N_i} N_{ij})\{\sigma_\gamma^{(m)}\}^2} - \frac{1}{2\sum_{j=1}^{N_i} N_{ij}} \log[2\pi\{\sigma_\gamma^{(m)}\}^2] \Big) K_h(t_{ijk} - t_0),
\end{aligned}$$

where $\tilde{p}_{\phi,ijk}^{(m)} = g^{-1}[X_{ij}^{\top}\{\beta_0 + \beta_1(t_{ijk} - t_0)\} + Z_{i(j)}^{\top}\{\theta_0 + \theta_1(t_{ijk} - t_0)\} + b_{ij0}^{(m)} + \gamma_{i0}^{(m)}]$, $\tilde{q}_{\phi,ijk}^{(m)} = 1 - \tilde{p}_{\phi,ijk}^{(m)}$ and $K_h(\cdot) = K(\cdot/h)/h$ with $K(\cdot)$ denoting the kernel function and h denoting the bandwidth. For selection of bandwidth, we use cross-validation methods (Hoover et al., 1998; Wu et al., 1998). We use a safeguarded one-step Newton-Raphson iteration to maximize the above approximated expected local log-likelihood. The updated estimator $\phi^{(m)} = [\{\phi_0^{(m)}\}^{\top}, \{\phi_1^{(m)}\}^{\top}]^{\top}$ can be given as $\phi^{(m)} = \phi^{(m-1)} + s\{I_{\phi}^{(m)}(t_0)\}^{-1}V_{\phi}^{(m)}(t_0)$, where the score function with respect to ϕ is equal to

$$V_{\phi}^{(m)}(t_0) = \sum_{i=1}^{I} V_{\phi i}^{(m)}(t_0) = \sum_{i=1}^{I}\sum_{j=1}^{N_i}\sum_{k=1}^{N_{ij}} a_{\phi,ijk}^{(m)} K_h(t_{ijk} - t_0)\{X_{ij}^{\top}, Z_{i(j)}^{\top}, (t_{ijk} - t_0)X_{ij}^{\top}, (t_{ijk} - t_0)Z_{i(j)}^{\top}\}^{\top}$$

with $a_{\phi,ijk}^{(m)} = Y_{ijk} - p_{\phi,ijk}^{(m)} - \{v_{b,ij0}^{(m)} + 2r_{ij0}^{(m)} + v_{\gamma,i0}^{(m)}\}\{p_{\phi,ijk}^{(m)}(q_{\phi,ijk}^{(m)})^2 - q_{\phi,ijk}^{(m)}(p_{\phi,ijk}^{(m)})^2\}/2$, $p_{\phi,ijk}^{(m)} = g^{-1}[X_{ij}^{\top}\{\beta_0^{(m-1)} + \beta_1^{(m-1)}(t_{ijk} - t_0) + Z_{i(j)}^{\top}\{\theta_0^{(m-1)} + \theta_1^{(m-1)}(t_{ijk} - t_0)\} + b_{ij0}^{(m)} + \gamma_{i0}^{(m)}\}]$ and $q_{\phi,ijk}^{(m)} = 1 - p_{\phi,ijk}^{(m)}$. The explicit expression for $I_{\phi}^{(m)}(t_0)$ and further discussions on the safeguarded Newton-Raphson algorithm are deferred to Web Appendix B. For inference on the varying coefficient functions, we propose pointwise standard errors for $\hat{\beta}(t)$ and $\hat{\theta}(t)$. At a fixed time point t_0, the standard errors of $\hat{\beta}(t_0)$ and $\hat{\theta}(t_0)$ can be obtained from the inverse of the empirical Fisher information matrix $\sum_{i=1}^{I} V_i(t_0)V_i(t_0)^{\top}$ where $V_i(t_0)$ is equal to $V_{\phi i}^{(m)}(t_0)$ from the last iteration.

Note that standard errors based on information matrices have been reported to potentially underestimate their targeted values in the EM algorithm framework due to not taking into account the variability in the estimation of the random effects (Kass and Steffey, 1989). We study the accuracy of the proposed standard errors in simulation studies outlined in the next section and observe that standard errors for only $\beta_0(t)$ are underestimated in the proposed MME-VCM, which does not directly affect the inference on the time-varying effects of subject- and facility-level risk factors as well as the variance components.

3. Simulation

We conduct simulations to study the finite sample performance of the proposed estimators and standard errors based on the proposed EM algorithm. The robustness of the proposed estimators under violations of the distribution assumptions of the multilevel random effects is also studied. Finally, the effects of ignoring facility-level dependence is considered via a direct comparison of the proposed MME-VCM with a multilevel varying coefficient model including single-level random effects, denoted by MVCM. We defer details on simulation design to Web Appendix C.

3.1 Finite Sample Performance and Robustness to Violation of Distribution Assumptions

Two cases with $I = 100$ and $I = 500$ total number of facilities are considered. The bandwidths used for estimating the subject- and facility-level varying coefficient functions $\beta(t)$ and $\theta(t)$, respectively, are chosen in a preliminary simulation study with 50 Monte Carlo runs using 10-fold cross-validation and are kept fixed for the main simulation. The bandwidths selected for $I = 100$ and $I = 500$ are 1.1 and .6 years, respectively. Mean squared error (MSE) and

relative mean squared deviation error (MSDE), $\text{MSDE}_{\hat{f}} = [\int \{\hat{f}(t) - f(t)\}^2 dt] / \int f^2(t) dt$ (for a generic function $f(t)$) are used to assess the estimation of the time-invariant model parameters σ_b^2 and σ_γ^2 and time-varying coefficient functions, $\beta(t)$ and $\theta(t)$, respectively. MSDE is a commonly used measure to assess overall estimation accuracy of functional parameters. It can be thought of as a standardized MSE measure for functions, combining information on bias and standard deviation. The simulation results presented are based on 200 Monte Carlo runs.

Figure 1 displays the estimated time-varying coefficient functions of the multilevel risk factors $\beta(t)$ and $\theta(t)$ along with their pointwise confidence intervals (± 2 SEs) from the simulation runs with the median MSDE based on $I = 100$ total facilities. The estimates (solid curves) track the true functions (dashed curves) which lie within the pointwise confidence intervals (± 2 SEs, shaded) for all varying coefficient functions except the y-intercept $\beta_0(t)$. As explained before in Section 2.2.2, the standard errors have been reported to potentially underestimate their targeted values in an EM framework due to not taking into account the variability in the estimation of the random effects (Kass and Steffey, 1989). Note that this underestimation is only observed for $\beta_0(t)$ in MME-VCM and does not affect the inference on time-varying effects of subject- and facility-level covariates, as well as the variance components. The (25th, 50th, 75th) percentiles of the MSDEs for the varying coefficient functions and of the MSEs for the variance components σ_b^2 and σ_γ^2 from both simulation cases are summarized in Table 1. The two error measures both get smaller with increasing number of facilities, as expected.

The performance of the proposed standard error estimates of the varying coefficient functions $\beta(t)$ and $\theta(t)$ and the variance components σ_b^2 and σ_γ^2 are also studied. The bias, sample average (denoted by SE) and sample standard deviation (denoted by SD_{SE}) of the estimated standard errors at three time points are given in Table 2. Also given in Table 2 are the standard deviations of the estimates $\hat{\beta}(t)$ and $\hat{\theta}(t)$ (denoted by SD) which can be regarded as the true standard errors. Note that the estimation bias is less than SD, implying that the proposed estimator targets the true function. In addition, for subject- and facility-level varying coefficient functions and the variance components, although the proposed standard error formula slightly overestimates the actual one (similar to the results reported in (Tutz and Kauermann, 2003)), the differences between SE and SD are typically smaller than twice SD_{SE}, except for $\beta_0(t)$, showing that the proposed standard error formula works reasonably well. Also note that reported SE gets closer to SD and the reported bias gets smaller with increasing number of facilities, as expected.

In the context of generalized mixed effects models with one-level random effects, it has been shown that estimation of the fixed effects parameters, except for the y-intercept, are quite robust to misspecification of the random effects distribution (Neuhaus et al., 1992; Heagerty and Kurland, 2001; McCulloch and Neuhaus, 2011). To study the impact of misspecification of the random effect distributions at the facility- or the subject- or at both the facility- and the subject-levels on estimation of MME-VCM, we conduct additional simulation studies, where deviations from normality are induced by assuming a gamma distribution for the

random effects. Data are generated using the the same setup as described in Web Appendix C with the exception that the random effects equal $\gamma_i = \sigma_\gamma(a_i - \lambda) / \sqrt{\lambda}$ and $b_{ij} = \sigma_b(w_{ij} - \lambda) / \sqrt{\lambda}$ where a_I and w_{ij} are generated from gamma distributions with shape parameter λ and rate parameter 1. We explore simulation setups with varying values of λ, where smaller λ correspond to further deviations from normality. Since the random effects are multilevel, we consider four scenarios: (I) the distribution of only the subject-level random effects is misspecified, (II) the distribution of only the facility-level random effects is misspecified, (III) the distributions of both subject- and facility-level random effects are misspecified and (IV) no violations in the random effects distributions. For the first three scenarios, three different values of λ are considered: λ = .5, λ = 2 and λ = 4.

Table 3 reports the average bias $\int|\bar{f}^*(t) - f(t)|dt$, where $\bar{f}^*(t)$ is the mean of the estimated functions $\hat{f}(t)$ targeting a generic function $f(t)$) for varying coefficient functions associated with subject-level covariates ($\beta(t)$) and facility-level covariates ($\theta(t)$. Also reported is the bias for variance components σ_b^2 and σ_γ^2. The results show that the proposed EM algorithm is quite robust to misspecification of the random effects distribution at the facility-level. The proposed estimators are also robust to violations at the subject-level with the exception of the intercept term $\beta_0(t)$. The bias of the intercept estimator ($\hat{\beta}_0(t)$) decreases as deviations from normality decrease (corresponding to larger λ values in the gamma distribution), as expected. These results are consistent with the results of Neuhaus et al. (1992), Heagerty and Kurland (2001), and McCulloch and Neuhaus (2011) on robustness to misspecification of the random effects distribution in generalized mixed effects models. In addition, the fact that the bias is observed only when the distribution of the subject-level random effects is misspecified can be explained by the fact that in the proposed EM algorithm, misspecified facility-level random effect affects only one element of the random effects vector $u_i = (b_{i1}, \ldots, b_{iN_i}, \gamma_i)^\intercal$ within facility i, when the rest of the subject-level random effects are still normal. Hence the deviation of the distribution of the random effects vector from multivariate normality is minimal. However, misspecification of the distribution of the subject-level random effects, affects almost all the elements of the random effects vector, leading to significant deviation from normality.

3.2 Effects of Ignoring Facility-level Dependence

Finally, we conduct simulations to study the effects of ignoring facility-level correlation in the data. We compare the proposed MME-VCM with a varying coefficient model with multilevel covariates and only subject-level random effects (referred to as MVCM). Data are generated using the setup of Web Appendix C under three different variance ratios of the random effects: (1) $\sigma_\gamma^2 / \sigma_b^2 = 1$, (2) $\sigma_\gamma^2 / \sigma_b^2 = .25$ and (3) $\sigma_\gamma^2 / \sigma_b^2 = .07$. Results from the third case with $\sigma_\gamma^2 / \sigma_b^2 = .07$, mimicking the random effects variance ratio from the USRDS data application, are deferred to Supporting Information Table S1 and S2. Results show no differences in the bias of the estimated multilevel varying coefficient functions between the two models, however there are differences observed in the true (SD) and estimated (SE) standard errors of the facility-level varying coefficient functions ($\hat{\theta}(t)$) (Table 4, Table S1). Table 4 and Table S1 show the bias, the empirical standard deviations (SD) and the sample

average (SE) and sample standard deviation (SD_{SE}) of the estimated standard errors of the varying coefficient functions at three time points from both models. Note that the empirical standard deviations (which can be regarded as the true standard errors) of the facility-level varying coefficient functions $\hat{\theta}(t)$, are smaller in MME-VCM compared to MVCM, implying that the MME-VCM is more efficient in estimation of the effects of facility-level risk factors than the single-level random effects model. This is also reflected in the MSDE results shown in Table 5 and Table S2, with MME-VCM leading to reduced MSDE for facility-level varying coefficient functions. In addition, for the facility-level varying coefficient functions, the differences between SE and SD are typically larger than twice SD_{SE} in MVCM except for $\sigma_\gamma^2 / \sigma_b^2 = .07$, indicating that the estimated standard errors do not target the true standard errors when the within-facility correlation is ignored. Note that the differences observed in standard error estimates only apply to facility-level varying coefficient functions, and are not observed for subject-level estimates. Also, the difference between standard error estimation of the two models in targeting facility-level varying coefficient functions decreases as the variance of the facility-level random effects gets smaller compared to the variance of the subject-level random effects, as expected. However, even with facility-level random effects variance one fourth or 7% of the subject-level random effects variance, gains from accounting for the facility-level correlation via MME-VCM are visible (Table 4, Table S1).

4. Modeling Hospitalization Risk Among Patients on Dialysis

4.1 Description of the USRDS Study Cohort and Patient- and Facility-level Predictors

Our study utilizes the United States Renal Data System (USRDS), a national database that collects data on nearly all patients with end-stage renal disease (ESRD) in the US. Patient demographics, hospitalizations, as well as comorbidities at initiation of dialysis are all included in the USRDS. The cohort in our study includes patients of age 18 years or older who initiated dialysis between January 1, 2006 and December 31, 2008. Patients are followed up for a maximum period of five years where the last date of follow-up is December 31, 2013. The final study cohort includes 102,342 patients and 2,618 facilities with an overall three month hospitalization risk of 27.14%. We defer detailed descriptions of the study cohort and exclusion rules to Web Appendix D.

Time-varying effects of 27 patient-level and three facility-level covariates on patients' hospitalization risk are studied with the proposed MME-VCM. Patients' age, gender, body mass index (BMI), whether diabetes is the cause of ESRD and 23 additional comorbidities, ranging from chronic obstructive pulmonary disease (COPD), septicemia, ulcers, drug and alcohol disorders, end-stage liver disease, severe cancer, psychiatric disorders to arthritis are included in the patient-level covariates. The facility-level covariates include total number of patients, nurse-to-patient ratio, and patient care technician (PCT)-to-patient ratio. For ease of interpretation, we transform facility-level covariates nurse-to-patient ratio and PCT-to-patient ratio into percentages by multiplying the ratios by 100 and truncating percentages larger than 100 at 100 (12 facilities in total). In addition, all continuous covariates (age, BMI, total number of patients, nurse-to-patient ratio, and PCT-to-patient ratio) are mean-centered before modeling.

4.2 Results

The variances of the subject- and facility-level random effects are estimated to be $\hat{\sigma}_b^2 = 1.277$ and $\hat{\sigma}_\gamma^2 = 0.089$ with standard errors 0.011 and 0.005, respectively. Hence, a large portion of the total variation is explained at the subject-level, nevertheless, both subject- and facility-level random effects explain parts of the variation in the multilevel modeling. For selection of the bandwidth of the varying coefficient functions corresponding to the multilevel risk factors, we utilize a 10-fold cross-validation. The prediction error used for cross-validation is $\sum_{i=1}^{I}\{\sum_{j=1}^{N_i}\sum_{k=1}^{N_{ij}}(Y_{ijk}-\hat{p}_{ijk})^2 / \sum_{j=1}^{N_i} N_{ij}\}$, where $\hat{p}_{ijk} = g^{-1}\{X_{ij}^{\mathsf{T}}\hat{\beta}(t_{ijk}) + Z_{i(j)}^{\mathsf{T}}\hat{\theta}(t_{ijk})\}$. The predictions $\hat{p}_{ijk}$ for facilities left out, are estimated without the subject- and facility-level random effects where $\hat{\beta}(t)$ and $\hat{\theta}(t)$ are estimated using data on all facilities, except for facilities in the leave-out group. A bandwidth of 3.25 years in the five year follow-up is chosen for estimation of the varying coefficient functions $\beta(t)$ and $\theta(t)$.

Among the 27 patient-level risk factors, 25 are found to have significant effects on patient hospitalization risk. The two nonsignificant patient-level risk factors are fibrosis of the lung or other chronic lung disorders and respirator dependence. As expected, all of the comorbidities (prior to dialysis) that are significant are associated with an increase in hospitalization risk. Figure 2 displays the estimated varying coefficient functions for (a) the y-intercept and a selection of several patient-level risk factors which are found to be significant: (b) gender (female), (c) age, (d) BMI, (e) whether diabetes is the cause of ESRD, (f) COPD, (g) septicemia and (h) arthritis. The pointwise confidence intervals (± 2 SEs) are shaded in gray. The estimated y-intercept $\hat{\beta}_0(t)$ is increasing over t, showing that the hospitalization risk of a male patient initiating dialysis at mean age 65, with mean BMI of 29, no comorbidities, not having diabetes as the cause of ESRD, and treated at an average facility (with mean size of 93 patients, nurse-to-patient ratio of 7.5% and PCT-to-patient ratio of 9.5%), is increasing over time on dialysis, conditional on survival. Females have higher estimated hospitalization risk than males, but this difference in hospitalization risk gets smaller as patients stay longer on dialysis. Older age at initiation of dialysis is associated with higher hospitalization risk except during the first year on dialysis. Diabetes as the cause of ESRD is associated with higher hospitalization risk with the effect getting stronger in the later years of dialysis treatment (e.g., > 2 years). Higher BMI is associated with lower hospitalization risk but the protective effect is getting weaker in later years of dialysis. This protective effect of BMI is also found in other studies on adverse events such as cardiovascular risk (Kalantar-Zadeh et al., 2003) and mortality (Kalantar-Zadeh et al., 2005) among patients on dialysis. All three comorbidities displayed (COPD, septicemia and arthritis) are associated with higher hospitalization risk, with different time-varying effects throughout the course of dialysis. Chronic conditions seem to have longer lasting effects on hospitalization, with effects getting stronger over time on dialysis. For example, the association between COPD and hospitalization risk gets stronger as patients stay longer on dialysis, whereas the association between septicemia and hospitalization risk gets weaker.

To visualize the combined effects of patient-level covariates on hospitalization risk, the estimated risk trajectories $\hat{p}_{ij}(t) = g^{-1}\{X_{ij}^{\mathsf{T}}\hat{\beta}(t) + Z_{i(j)}^{\mathsf{T}}\hat{\theta}(t)\}$ for three hypothetical subjects are

displayed in Figure 3. The three subjects compared are chosen to be female patients who initiated dialysis at 65 years of age, with diabetes as the cause of ESRD and having BMI of 29, receiving dialysis at a typical ('median') facility (with 86 patients, 6.8% nurse-to patient ratio and 9.4% PCT-to-patient ratio). The three subjects differ in their comorbidities at initiation of dialysis, where one of the subjects has no comorbidities (representing a 'healthy' patient), one has only arthritis ('moderate') and the last patient has both septicemia and a lung disorder ('severe'). The largest difference in hospitalization risk appears in the first year of dialysis, since the effects of septicemia are the largest at initiation of dialysis. The differences between the three risk trajectories are found significant within the first two years. However, the pointwise confidence intervals (± 2 SEs) start to overlap for later years on dialysis. While differences in hospitalization risk are typically not significant for later years on dialysis for patients with nonchronic comorbidities (such as septicemia), for patients with chronic comorbidities at initiation of dialysis such as COPD, the significant increase in hospitalization risk is found to be sustained during all five years of follow-up.

At the facility-level, the time-varying effects of the three risk factors considered are displayed in Figure 4: (a) nurse-to-patient ratio (in percent), (b) patient care technician (PCT)-to-patient ratio (in percent) and (c) total number of patients. The number of total patients (i.e., facility size) is not found to be significantly correlated with hospitalization risk. However, both nurse-to-patient and PCT-to-patient ratios are found significant in the first year of dialysis, where higher percent of staff are correlated with a lower risk of hospitalization. This effect is not found significant in the later years of dialysis. Note also that the magnitude of the significant effects on hospitalization in the first year are comparable for nurse-to-patient and PCT-to-patient ratios. Nevertheless, the magnitude of facility-level effects are small relative to the effects of patient-level factors, as expected.

Time-varying effects of multilevel risk factors on hospitalizations of dialysis patients have been studied before by Li et al. (2018) where a varying coefficient model for multilevel risk factors (VCM-MR) was proposed for three-level hierarchical data. The MME-VCM and VCM-MR consider the same set of patient- and facility-level risk factors in modeling hospitalization risk using the USRDS data, where the results from MME-VCM largely agree with findings from the VCM-MR. Both models identify the same patient-level demographics and comorbidities as having significant effects on hospitalization (with similar time-varying trends). Both models find that the large proportion of variation is explained at the subject-level and that the magnitude of effects of facility-level factors are small relative to the effects of patient-level factors. However, the two models differ in their findings on effects of facility-level covariates. While the VCM-MR finds all three facility-level risk factors significant throughout the first five years of dialysis, MME-VCM does not find facility size to have significant effects on hospitalization, and finds that nurse-to-patient and PCT-to-patient ratios are significant (negatively correlated) for hospitalization risk only in the first year of dialysis.

Rigorous studies of dialysis facility-level effects on patients' hospitalization risk are sparse and time-varying effects of these risk factors have not been studied before in literature except Li et al. (2018). Nevertheless, Chen et al. (2019) studied the association of dialysis facility staffing factors with profiling results with respect to yearly 30-day unplanned

hospitalization readmission rates. Both nurse-to-patient and total staff-to-patient ratios have been considered and the study found that dialysis facilities with significantly worse 30-day readmission rates had lower nurse-to-patient ratios and total staff-to-patient ratios, with only the disparities in nurse-to-patient ratio from 2010 reported as significant. Hence the association identified between the two facility-level staffing ratios (nurse-to-patient and PCT-to-patient ratios) considered in MME-VCM and VCM-MR agree with the previous findings of Chen et al. (2019).

The differences found in the time-varying effects of facility-level risk factors are likely due to the main difference between the two models, that is while VCM-MR includes only patient-level random effects, the proposed MME-VCM includes random effects at both the patient- and facility-levels. The VCM-MR models within-facility correlation through facility-specific fixed effects instead of random effects, which adds flexibility in modeling facility-specific risk trajectories when the main goal centers around inference for facility-specific time-varying effects. There is extensive discussion on modeling facility effects via fixed or random effects in the time-static facility profiling literature where the goal is to identify facilities with significantly worse (or better) performance than a reference standard (Kalbfleisch and Wolfe, 2013). For profiling purposes, i.e. for identifying facility performance deviating from a norm, fixed facility effects have been reported to be more effective, where models with facility-level random effects shrink estimates to an overall mean leading to more reliable estimation of the facility effects near the center of the distribution but not away from the center (He et al., 2013).

For the main goal of the current paper, which is to study effects of multilevel risk factors on patient hospitalizations, multilevel random effects in MME-VCM lead to more stable and reliable inference, making facility-level effects found in applications to the USRDS data using MME-VCM more trustworthy, due to the stabilization through the random effects. Moreover, multilevel random effects lead to additional stabilization for estimation of effects of small facilities (with a low number of patients), since estimating facility-specific fixed effects based on data from a small facility is challenging (Kalbfleisch and Wolfe, 2013). In addition, while the inference for MME-VCM is based on the inverse of the empirical information matrix, Li et al. (2018) proposes inference for VCM-MR via bootstrap. This leads to additional computational savings for inference in MME-VCM, where for simulations with 100 facilities, the run times of the estimation and inference procedures proposed for MME-VCM and VCM-MR (inference based on 200 bootstrap samples), on a DELL XPS 8910 PC (3.4 GHz CPU, 16 GB RAM), are 2.5 minutes and 4 hours and 50 minutes, respectively.

5. Discussion

We proposed a generalized multilevel mixed effects varying coefficient model (MME-VCM) to study time-varying effects of multilevel risk factors for longitudinal data. Due to the three-level hierarchical structure in the USRDS data where longitudinal measurements are nested in patients and patients are nested in dialysis facilities, we model the hierarchical dependence via a two-level random effects structure. In addition, both patient- and facility-level predictors are included in the regression model to characterize their effects on

hospitalization risk as functions of time that patients are on dialysis. To handle the high-dimensional integration resulting from the hierarchical random effects structure, we utilize a fully exponential Laplace approximation approach which leads to lower order approximation errors than the standard Laplace method without a substantial increase in the computational burden. For inference on the multilevel varying coefficient functions and the variance components, we derive standard errors based on the inverse of the empirical Fisher information matrices. In the USRDS data application, MME-VCM identifies significant multilevel risk factors for patient hospitalizations, providing insights into health care strategies for the reduction of patient hospitalization risk.

Note that even though the application to the USRDS data considered in Section 4 only includes baseline covariates, the proposed estimation and inference procedures for MME-VCM can easily be extended to accommodate time-varying covariates. Finally, the proposed model targets the hospitalization risk of a dynamic cohort of survivors through a partly conditional modeling approach, conditional on the patients being alive. The model can be extended to include the patients' death as part of the outcome, leading to time-dynamic joint modeling of survival and multilevel longitudinal data. This extension requires further research and is identified as a future direction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This study was supported by research grants from the National Institute of Diabetes and Digestive and Kidney Diseases (R01 DK092232 - DS, DVN, KK, CMR, YC; K23 DK102903 - CMR, KK, DVN). The interpretation and reporting of the data presented here are the responsibility of the authors and in no way should be seen as an official policy or interpretation of the United States government.

References

Cai Z, Fan J, and Li R (2000). Efficient estimation and inferences for varying-coefficient models. Journal of the American Statistical Association 95, 888–902.

Chen H and Wang Y (2011). A penalized spline approach to functional mixed effects model analysis. Biometrics 67, 861–870. [PubMed: 21155747]

Chen Y, Rhee C, Senturk D, Kurum E, Campos L, Li Y, Kalantar-Zadeh K, and Nguyen D (2019). Association of US dialysis facility staffing with profiling of hospital-wide 30-day unplanned readmission. Kidney Diseases 5, 153–162. [PubMed: 31259177]

Cleveland WS, Grosse E, and Shyu WM (1991). Local regression models In Chambers JM and Hastie TJ, editors, Statistical Models in S, pages 309–376. Wadsworth & Brooks, Pacific Grove.

Estes JP, Nguyen DV, Chen Y, Dalrymple LS, Rhee CM, Kalantar-Zadeh K, and Şentürk D (2018). Time-dynamic profiling with application to hospital readmission among patients on dialysis. Biometrics 74, 1383–1394. [PubMed: 29870064]

Estes JP, Nguyen DV, Dalrymple LS, Mu Y, and Şentürk D (2014). Cardiovascular event risk dynamics over time in older patients on dialysis: A generalized multiple-index varying coefficient model approach. Biometrics 70, 754–764. [PubMed: 24766178]

Estes JP, Nguyen DV, Dalrymple LS, Mu Y, and Şentürk D (2016). Time-varying effect modeling with longitudinal data truncated by death: conditional models, interpretations, and inference. Statistics in Medicine 35, 1834–1847. [PubMed: 26646582]

Fan J and Zhang W (1999). Statistical estimation in varying coefficient models. The Annals of Statistics 27, 1491–1518.

Gelman A (2006). Multilevel (hierarchical) modeling: what it can and cannot do. Technometrics 48, 432–435.

Hastie T and Tibshirani R (1993). Varying-coefficient models. Journal of the Royal Statistical Society. Series B (Methodological) 55, 757–796.

He K, Kalbfleisch JD, Li Y, and Li Y (2013). Evaluating hospital readmission rates in dialysis facilities; adjusting for hospital effects. Lifetime Data Analysis 19, 490–512. [PubMed: 23709309]

Heagerty PJ and Kurland BF (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. Biometrika 88, 973–985.

Hoover DR, Rice JA, Wu CO, and Yang L-P (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika 85, 809–822.

Kalantar-Zadeh K, Abbott KC, Salahudeen AK, Kilpatrick RD, and Horwich TB (2005). Survival advantages of obesity in dialysis patients. The American Journal of Clinical Nutrition 81, 543–554. [PubMed: 15755821]

Kalantar-Zadeh K, Block G, Humphreys MH, and Kopple JD (2003). Reverse epidemiology of cardiovascular risk factors in maintenance dialysis patients. Kidney International 63, 793–808. [PubMed: 12631061]

Kalbfleisch JD and Wolfe RA (2013). On monitoring outcomes of medical providers. Statistics in Biosciences 5, 286–302.

Kass RE and Steffey D (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). Journal of the American Statistical Association 84, 717–726.

Kurland BF and Heagerty PJ (2005). Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. Biostatistics 6, 241–258. [PubMed: 15772103]

Li Y, Nguyen DV, Chen Y, Rhee CM, Kalantar-Zadeh K, and Şentürk D (2018). Modeling time-varying effects of multilevel risk factors of hospitalizations in patients on dialysis. Statistics in Medicine 37, 4707–4720. [PubMed: 30252153]

McCulloch CE and Neuhaus JM (2011). Prediction of random effects in linear and generalized linear models under model misspecification. Biometrics 67, 270–279. [PubMed: 20528860]

Neuhaus JM, Hauck WW, and Kalbfleisch JD (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. Biometrika 79, 755–762.

Rizopoulos D, Verbeke G, and Lesaffre E (2009). Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71, 637–654.

Tierney L, Kass RE, and Kadane JB (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. Journal of the American Statistical Association 84, 710–716.

Tutz G and Kauermann G (2003). Generalized linear random effects models with varying coefficients. Computational Statistics & Data Analysis 43, 13–28.

United States Renal Data System (2018). 2018 USRDS annual data report: “Epidemiology of kidney disease in the United States” Technical report, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD.

Wu CO, Chiang C-T, and Hoover DR (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. Journal of the American Statistical Association 93, 1388–1402.

Wu H and Zhang J-T (2002). Local polynomial mixed-effects models for longitudinal data. Journal of the American Statistical Association 97, 883–897.

You J, Wan ATK, Liu S, and Zhou Y (2015). A varying-coefficient approach to estimating multi-level clustered data models. Test 24, 417–440.

Zhang D (2004). Generalized linear mixed models with varying coefficients for longitudinal data. Biometrics 60, 8–15. [PubMed: 15032768]

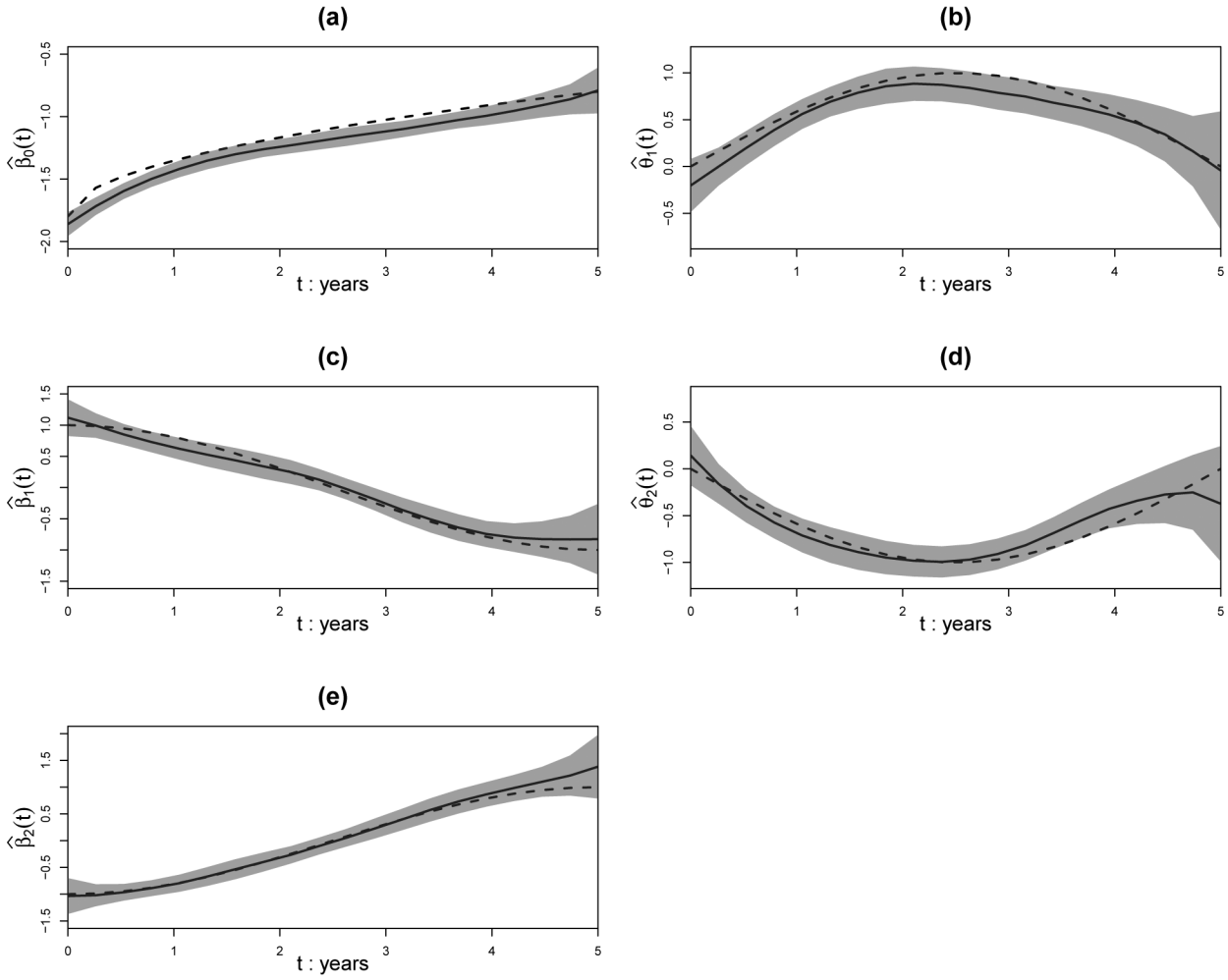

Figure 1.

The varying coefficient function estimates (solid) from the runs with the median MSDE among 200 Monte Carlo runs for $I = 100$ facilities. Also plotted are the pointwise confidence intervals (± 2 SEs, shaded) and the true varying coefficient functions (dashed).

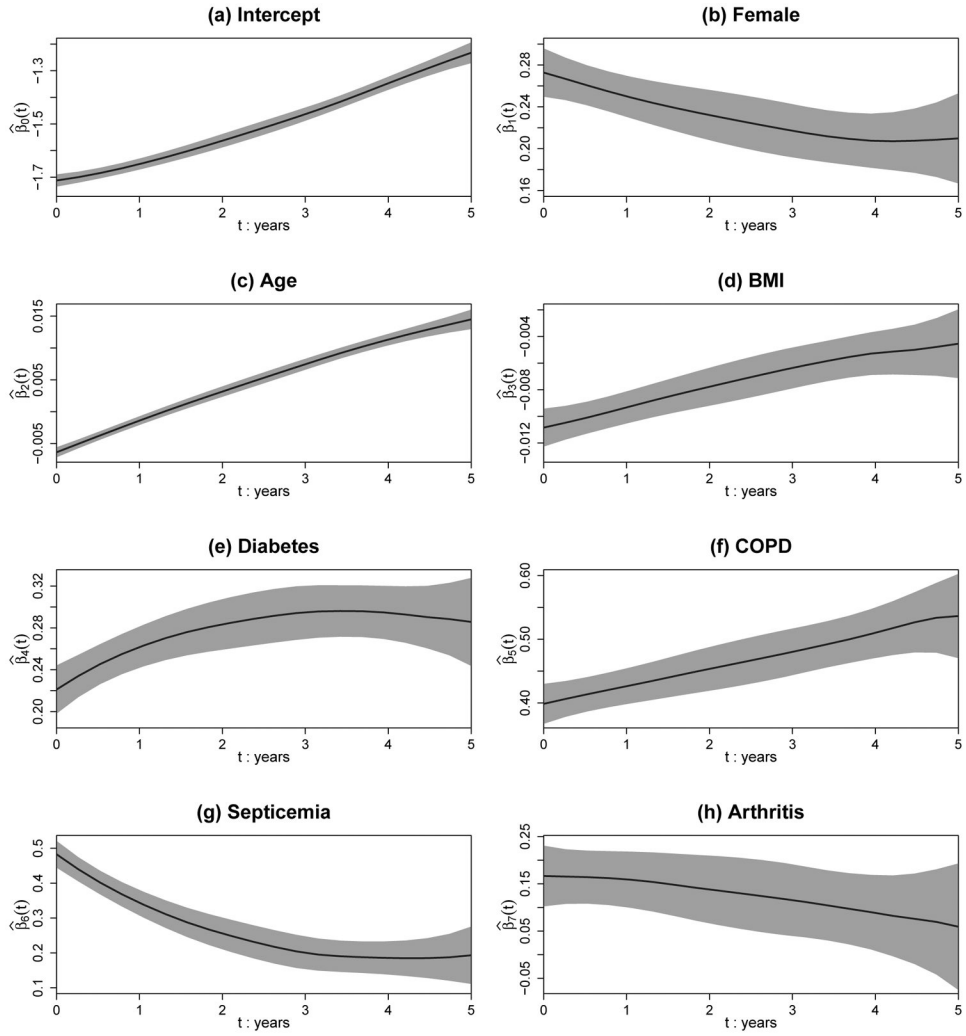

Figure 2.

Estimated patient-level effects $\hat{\beta}(t)$ (black) corresponding to (a) the intercept, (b) gender (female), (c) age, (d) BMI, (e) whether diabetes is the cause of ESRD, (f) COPD, (g) septicemia, and (h) arthritis along with their pointwise confidence intervals (± 2 SEs, shaded).

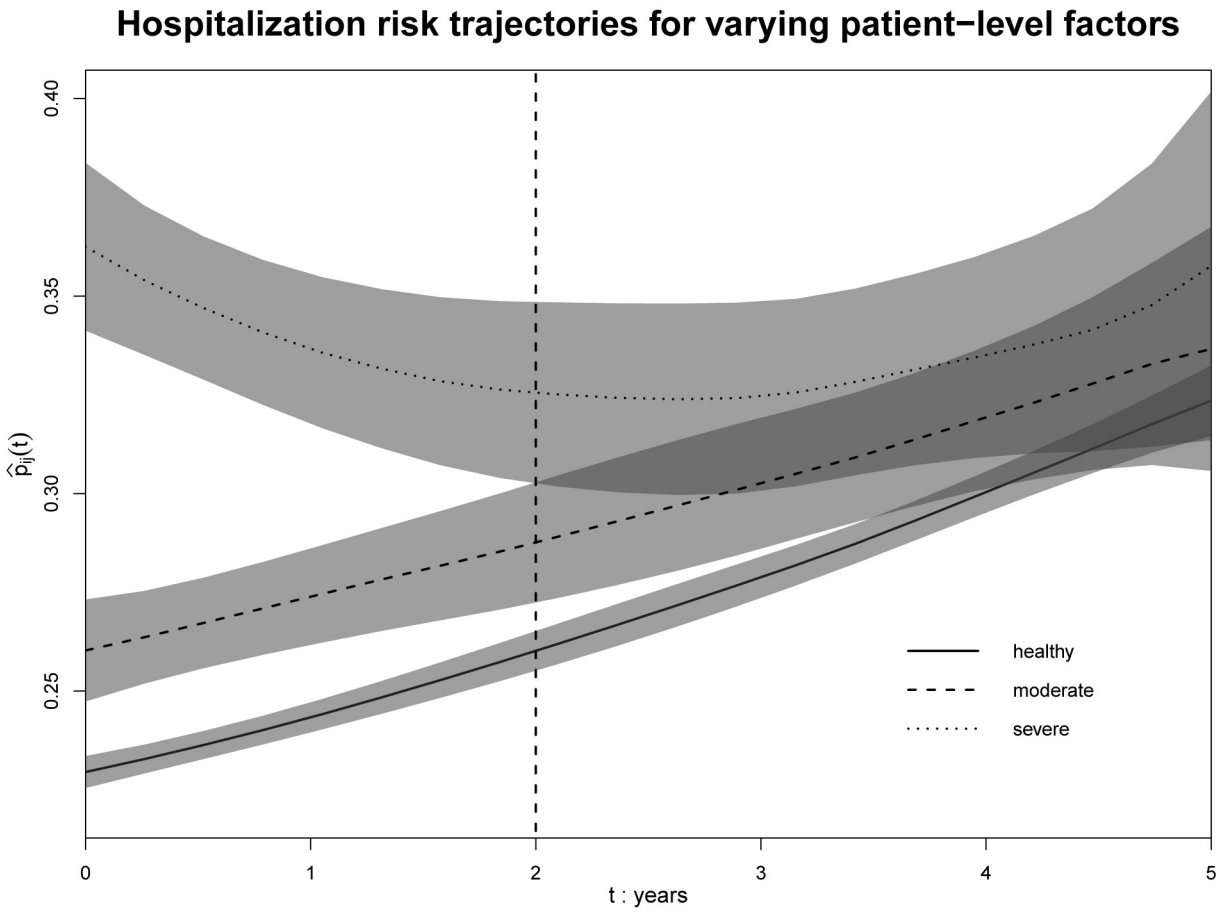

Figure 3.

Estimated hospitalization risk trajectories for three patients receiving dialysis at a typical ('median') facility who are 'healthy', 'moderate' and 'severe' at initiation of dialysis, given in solid, dashed and dotted, respectively, along with their pointwise confidence intervals (± 2 SEs, shaded).

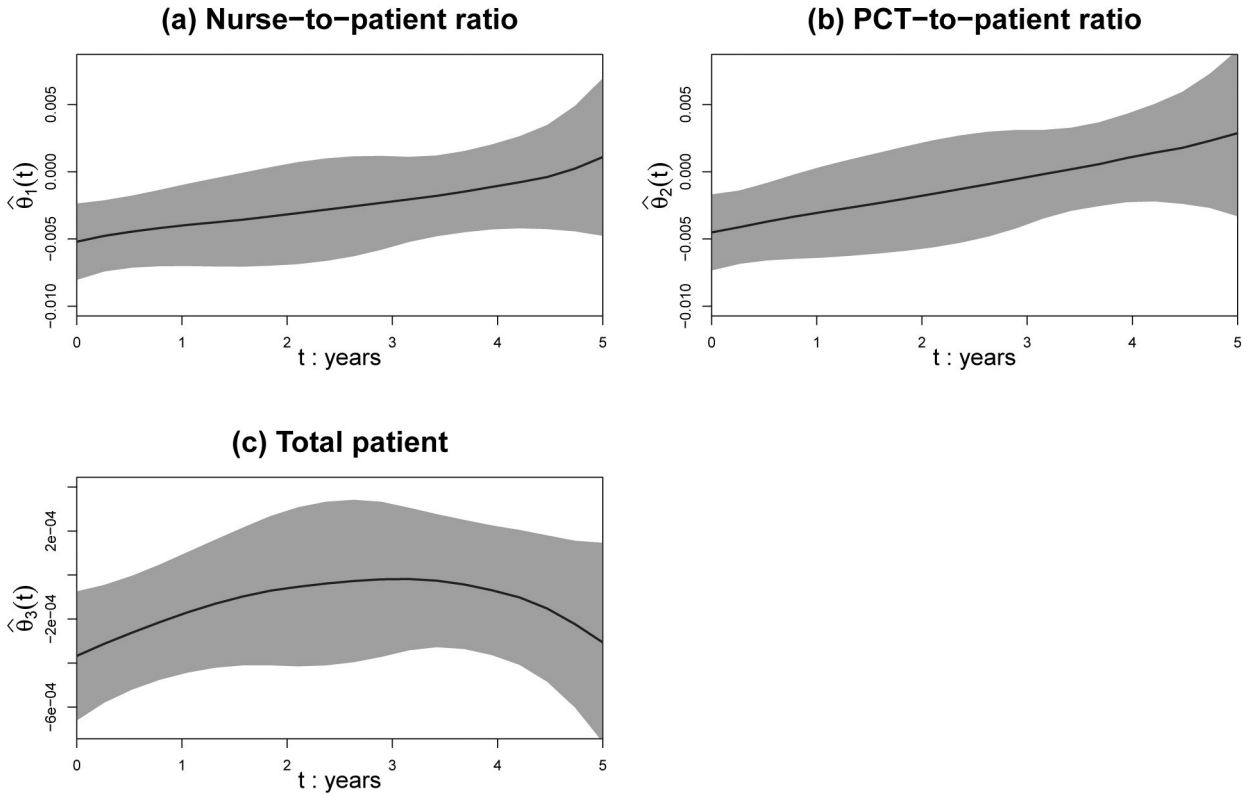

Figure 4.

Estimated facility-level effects $\hat{\theta}(t)$ corresponding to (a) nurse-to-patient ratio, (b) patient care technician (PCT)-to-patient ratio and (c) total number of patients, along with their pointwise confidence intervals (± 2 SEs, shaded).

Table 1

Percentiles of the mean squared deviation error (MSDE) for the time-varying coefficient estimates of effects of subject-level covariates $\beta(t)$ and effects of facility-level covariates $\theta(t)$ based on 200 Monte Carlo runs. Percentiles of mean squared error (MSE) are reported for variance components σ_b^2 and σ_γ^2.

	I=100			I=500		
MSDE	**25%**	**50%**	**75%**	**25%**	**50%**	**75%**
$\hat{\beta}_0(t)$	.002	.005	.012	.001	.002	.003
$\hat{\beta}_1(t)$	.012	.019	.031	.003	.005	.007
$\hat{\beta}_2(t)$	.012	.018	.032	.003	.005	.008
$\hat{\theta}_1(t)$	.014	.024	.043	.004	.007	.011
$\hat{\theta}_2(t)$	.015	.026	.046	.004	.007	.012
MSE						
$\hat{\sigma}_b^2$	<.001	<.001	.003	<.001	<.001	<.001
$\hat{\sigma}_\gamma^2$	.002	.010	.029	<.001	.002	.006

Table 2

The bias, true standard errors (SD), sample average (SE) and sample standard deviation (SD_{SE}) of the estimated standard errors at three time points (beginning, middle and end of follow-up) for varying coefficient functions (VCFs) associated with subject-level covariates $\beta(t)$ and facility-level covariates $\theta(t)$, as well as the variance components σ_b^2 and σ_γ^2.

		$I = 100$				$I = 500$			
t	**VCFs**	**Bias**	**SD**	**SE**	**SD_{SE}**	**Bias**	**SD**	**SE**	**SD_{SE}**
0	$\beta_0(t)$	−.057	.110	.046	.004	−.012	.051	.022	.001
	$\beta_1(t)$	.026	.141	.146	.014	.024	.069	.068	.003
	$\beta_2(t)$	−.020	.140	.144	.012	−.027	.067	.068	.002
	$\theta_1(t)$	.032	.144	.148	.018	.016	.074	.068	.003
	$\theta_2(t)$	−.029	.138	.148	.017	−.021	.069	.068	.004
2.5	$\beta_0(t)$	−.005	.107	.030	.003	.002	.049	.015	.001
	$\beta_1(t)$	.011	.073	.087	.007	−.001	.040	.046	.002
	$\beta_2(t)$	−.007	.076	.088	.007	−.004	.040	.046	.002
	$\theta_1(t)$	.083	.087	.093	.011	.039	.045	.046	.002
	$\theta_2(t)$	−.082	.091	.091	.011	−.043	.045	.046	.002
5	$\beta_0(t)$	−.081	.127	.094	.008	−.077	.062	.046	.002
	$\beta_1(t)$	−.004	.254	.304	.031	−.007	.124	.147	.006
	$\beta_2(t)$	−.012	.265	.304	.032	.007	.140	.148	.007
	$\theta_1(t)$	.014	.258	.303	.039	< .001	.135	.146	.008
	$\theta_2(t)$	−.028	.266	.303	.037	−.017	.135	.147	.008
	Variances								
	σ_b^2	−.014	.042	.039	.004	−.009	.019	.017	.001
	σ_γ^2	−.042	.140	.151	.027	−.017	.063	.068	.005

Table 3

The average bias for varying coefficient functions associated with subject- ($\beta(t)$) and facility-level covariates ($\theta(t)$). Also reported, is the bias for variance components σ_b^2 and σ_γ^2 under misspecified random effects distributions based on 200 Monte Carlo runs with $I = 100$ facilities. Scenarios (I), (II) and (III) correspond to misspecified subject-level, facility-level and both subject- and facility-level random effects, respectively. The last column shows results under no violation with normally distributed random effects.

Scenario	**(I)**			**(II)**			**(III)**			**(IV)**
λ	**.5**	**2**	**4**	**.5**	**2**	**4**	**.5**	**2**	**4**	
$\hat{\beta}_0(t)$	.068	.061	.041	.029	.026	.028	.104	.058	.034	.031
$\hat{\beta}_1(t)$	.035	.035	.035	.034	.033	.034	.033	.034	.032	.032
$\hat{\beta}_2(t)$	.034	.036	.032	.033	.031	.033	.034	.039	.036	.032
$\hat{\theta}_1(t)$	.058	.062	.065	.033	.062	.065	.061	.054	.064	.064
$\hat{\theta}_2(t)$	.060	.054	.070	.072	.062	.069	.056	.062	.071	.051
$\hat{\sigma}_b^2$	.007	−.100	−.084	.010	.010	.008	−.065	−.120	−.104	.019
$\hat{\sigma}_\gamma^2$	.024	.008	.015	.076	.001	.025	.024	−.027	−.004	.034

Table 4

The bias, true standard errors (SD), sample average (SE) and sample standard deviation (SD_{SE}) of the estimated standard errors at three time points (beginning, middle and end of follow-up) for varying coefficient functions (VCFs) associated with subject-level covariates $\beta(t)$ and facility-level covariates $\theta(t)$, as well as the variance components σ_b^2 and σ_γ^2. Both results from MVCM and MME-VCM are reported based on 200 *Monte Carlo runs with* $I = 100$ *facilities.*

$\sigma_\gamma^2 = 1$		MVCM				MME-VCM			
t	**VCFs**	**Bias**	**SD**	**SE**	**SD_{SE}**	**Bias**	**SD**	**SE**	**SD_{SE}**
0	$\beta_0(t)$	−.066	.108	.042	.001	−.067	.097	.046	.004
	$\beta_1(t)$	.002	.138	.135	.005	.004	.133	.144	.013
	$\beta_2(t)$	−.023	.135	.134	.005	−.002	.129	.144	.012
	$\theta_1(t)$	.032	.232	.133	.008	.021	.135	.146	.018
	$\theta_2(t)$	−.001	.235	.134	.008	−.003	.135	.146	.017
2.5	$\beta_0(t)$	−.014	.106	.026	.001	−.013	.094	.030	.003
	$\beta_1(t)$	−.004	.092	.084	.003	.002	.079	.086	.007
	$\beta_2(t)$	.004	.099	.084	.003	−.001	.083	.087	.008
	$\theta_1(t)$	.097	.206	.084	.005	.084	.095	.091	.011
	$\theta_2(t)$	−.073	.207	.084	.005	−.072	.090	.091	.011
5	$\beta_0(t)$	−.111	.125	.088	.004	−.101	.118	.093	.008
	$\beta_1(t)$	−.014	.263	.283	.019	.001	.257	.301	.032
	$\beta_2(t)$	−.008	.249	.283	.017	−.002	.238	.304	.029
	$\theta_1(t)$	.014	.319	.280	.019	.002	.257	.303	.037
	$\theta_2(t)$	−.001	.350	.280	.019	.002	.255	.303	.035
	Variances								
	σ_b^2	.942	.165	.065	.005	−.019	.037	.038	.003
	σ_γ^2					−.034	.141	.154	.027
$\sigma_\gamma^2 = .25$									
t	**VCFs**	**Bias**	**SD**	**SE**	**SD_{SE}**	**Bias**	**SD**	**SE**	**SD_{SE}**
0	$\beta_0(t)$	−.055	.062	.042	.001	−.042	.060	.046	.004
	$\beta_1(t)$	.030	.131	.134	.004	.020	.129	.144	.012
	$\beta_2(t)$	−.020	.135	.134	.004	−.014	.133	.143	.012
	$\theta_1(t)$	.024	.156	.133	.007	.045	.138	.145	.015
	$\theta_2(t)$	−.009	.154	.132	.008	−.030	.135	.145	.016
2.5	$\beta_0(t)$	−.007	.054	.025	.001	.003	.052	.029	.002
	$\beta_1(t)$	.013	.079	.080	.002	.012	.076	.085	.007
	$\beta_2(t)$	−.012	.076	.080	.002	−.015	.071	.085	.006
	$\theta_1(t)$	.075	.134	.080	.004	.091	.095	.088	.010

	$\theta_2(t)$	−.075	.121	.080	.004	−.089	.088	.089	.010
5	$\beta_0(t)$	−.086	.089	.085	.004	−.078	.088	.092	.009
	$\beta_1(t)$	−.026	.243	.274	.016	−.017	.242	.292	.027
	$\beta_2(t)$	−.018	.235	.276	.015	−.032	.236	.296	.027
	$\theta_1(t)$	.027	.311	.272	.020	.043	.282	.295	.037
	$\theta_2(t)$	−.017	.283	.270	.020	−.033	.270	.296	.037
	Variances								
	σ_b^2	.242	.063	.044	.002	−.001	.039	.039	.003
	σ_γ^2					.006	.043	.045	.008

Table 5

Percentiles of the mean squared deviation error (MSDE) for the time-varying coefficient estimates of effects of subject-level covariates $\beta(t)$ and facility-level covariates $\theta(t)$ from MVCM and MME-VCM based on 200 *Monte Carlo runs with* $I = 100$ *facilities.*

$\sigma_\gamma^2 = 1$	MVCM			MME-VCM		
MSDE	**25%**	**50%**	**75%**	**25%**	**50%**	**75%**
$\hat{\beta}_0(t)$	.003	.006	.013	.002	.004	.011
$\hat{\beta}_1(t)$	.014	.022	.032	.012	.018	.029
$\hat{\beta}_2(t)$	.013	.020	.037	.011	.018	.029
$\hat{\theta}_1(t)$	.022	.065	.146	.013	.025	.044
$\hat{\theta}_2(t)$	.028	.059	.148	.014	.025	.041
$\sigma_\gamma^2 = .25$						
MSDE	**25%**	**50%**	**75%**	**25%**	**50%**	**75%**
$\hat{\beta}_0(t)$	.001	.002	.004	.001	.002	.004
$\hat{\beta}_1(t)$	.012	.019	.029	.011	.019	.028
$\hat{\beta}_2(t)$	.011	.019	.028	.010	.018	.028
$\hat{\theta}_1(t)$	.016	.031	.067	.015	.029	.050
$\hat{\theta}_2(t)$	.013	.032	.066	.013	.026	.048