Title
Cardiovascular event risk dynamics over time in older patients on dialysis: A generalized multiple-index varying coefficient model approach

Journal
Biometrics, 70(3)

ISSN
0006-341X

Authors
Estes, Jason P
Nguyen, Danh V
Dalrymple, Lorien S
et al.

Publication Date
2014-09-01

DOI
10.1111/biom.12176

Peer reviewed

HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2015 July 31.

Published in final edited form as:

Biometrics. 2014 September ; 70(3): 751–761. doi:10.1111/biom.12176.

Cardiovascular Event Risk Dynamics Over Time in Older Patients on Dialysis: A Generalized Multiple-Index Varying Coefficient Model Approach

Jason P. Estes[1], **Danh V. Nguyen**[2,3], **Lorien S. Dalrymple**[4], **Yi Mu**[5], and **Damla Şentürk**[1,*]

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

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

[3]Institute for Clinical and Translational Science, University of California, Irvine, California 92687, U.S.A

[4]Division of Nephrology, Department of Medicine, University of California, Sacramento, California 95817, U.S.A

[5]Graduate Group in Epidemiology, University of California, Davis, California 95616, U.S.A

Abstract

Among patients on dialysis, cardiovascular disease and infection are leading causes of hospitalization and death. Although recent studies have found that the risk of cardiovascular events is higher after an infection-related hospitalization, studies have not fully elucidated how the risk of cardiovascular events changes over time for patients on dialysis. In this work, we characterize the dynamics of cardiovascular event risk trajectories for patients on dialysis while conditioning on survival status via multiple time indices: (1) time since the start of dialysis, (2) time since the pivotal initial infection-related hospitalization and (3) the patient's age at the start of dialysis. This is achieved by using a new class of generalized multiple-index varying coefficient (GM-IVC) models. The proposed GM-IVC models utilize a multiplicative structure and one-dimensional varying coefficient functions along each time and age index to capture the cardiovascular risk dynamics before and after the initial infection-related hospitalization among the dynamic cohort of survivors. We develop a two-step estimation procedure for the GM-IVC models based on local maximum likelihood. We report new insights on the dynamics of cardiovascular events risk using the United States Renal Data System database, which collects data on nearly all patients with end-stage renal disease in the U.S. Finally, simulation studies assess the performance of the proposed estimation procedures.

Keywords

Cardiovascular outcomes; End stage renal disease; Generalized linear models; Infection; Time-varying effects; United States Renal Data System

*dsenturk@ucla.edu.

Supplemental Materials

The provided R code and Web Appendices in Sections 3, 4, 5 and 6 are available with this paper at the Biometrics website on Wiley Online Library.

1 Introduction

As of 2010, end-stage renal disease affected more than 570,000 adults in the United States. Of these, more than 400,000 were on dialysis, a life-sustaining treatment (United States Renal Data System Annual Data Report [USRDS ADR], 2012). End-stage renal disease is associated with premature death, and cardiovascular disease is the leading cause of death in this population (USRDS ADR, 2012). An area of particular interest is whether infection contributes to the high risk of cardiovascular disease observed in this population, as infections are relatively common in patients on dialysis (Dalrymple et al., 2010; USRDS ADR, 2012). Previous studies have used an interval Poisson model, a Cox proportional hazards model and case-series analysis to support the notion that infection may contribute to a higher risk of cardiovascular disease in both the general population (Smeeth et al., 2004) and in the dialysis population (Foley et al., 2004; Ishani et al., 2005; Dalrymple et al., 2011; Mohammed et al., 2012). However, to date, studies have not fully elucidated how the risk (probability) of cardiovascular events changes over time for patients on dialysis and furthermore how the risk trajectory depends on individual characteristics.

Our primary objective is to understand how the risk of cardiovascular events dynamically evolves over time, and, in particular, how the changes depend simultaneously on multiple key time indices of: (a) time since the start of dialysis (vintage), (b) time since the initial infection-related hospitalization during dialysis and (c) baseline age at dialysis. While the dynamic cardiovascular risk trajectories as a function of the multiple time indices are of main interest, it is also important to characterize the effects of baseline covariates, which may potentially depend on baseline age. Baseline covariates of interest include demographic characteristics (sex, race), comorbidities (diabetes, coronary heart disease, congestive heart failure, peripheral vascular disease), body mass index (BMI) and estimated glomerular filtration rate (eGFR).

In addition to modeling risk trajectories over multiple indices, an important methodological challenge in the analysis of longitudinal data from USRDS is follow-up truncated by death. This is particularly relevant to the dialysis population because nationally, the annual mortality in the dialysis population is 20–25% (USRDS ADR, 2012). For the analysis of infection and cardiovascular risk in the dialysis population (USRDS data), the predominant dropout is due to death and it is certainly related to cardiovascular events (outcome). A cardiovascular event is defined as myocardial infarction, unstable angina, stroke, or transient ischemic attack; for a more detailed description, see Web Appendix B. When dropout is due to death, analysis demands careful consideration of the relevant target of inference. Kurland and Heagerty (2005) and Kurland et al. (2009) have considered truncation by death in longitudinal studies of geriatric populations, including studies examining disability or cognitive function outcomes and have proposed a ‘partly conditional’ target of inference where the analysis is conditional on being alive. Authors argue that an unconditional target of inference as is commonly used in drop-out or missing data literatures may not be a meaningful target when the missing data is due primarily to truncation by death, since it concerns a population where there are no deaths. Instead, more relevant scientific questions

can be addressed through a partly conditional model for the dynamic cohort of survivors. Some overall questions of particular clinical relevance for our study include:

- **a.** What is the cardiovascular risk trajectory during the course of dialysis for the dynamic cohort of survivors and how does it depend on baseline age or other baseline covariates?
- **b.** What is the cardiovascular risk at, for instance, 2 years after dialysis for patients who survive at least 2 years on dialysis without an infection?
- **c.** What is the cardiovascular risk for patients who acquired an infection at 2 years after dialysis and who survive 2 or more years?

To address the aforementioned modeling objectives, we propose generalized multiple-index varying coefficient (GM-IVC) models for generalized outcome data that (a) accommodate several time indices, (b) utilize one dimensional varying coefficient functions along each time index to facilitate ease of interpretability similar to standard varying coefficient models, (c) allow for multiple cross-sectional and longitudinal covariates and (d) target a partly conditional inference, conditional on survival status. It is known that modeling time-varying effects with multiple indices generally is unreasonably difficult because of the curse of dimensionality. To address the curse of dimensionality, we utilize a multiplicative structure for the multiple-index varying effects that is able to capture several time-dynamic cardiovascular risk trajectories. As detailed in Section 2.1, the proposed GM-IVC models are adaptive to the time period before and after the pivotal initial infection-related hospitalization. That is, the cardiovascular risk is modeled as a function of vintage for patients who never experience the pivotal infection and for patients who do experience the pivotal infection before their initial infection-related hospitalization. The GM-IVC models then shift to the time period after the initial infection-related hospitalization to estimate the cardiovascular risk as a function of time since the initial infection. Furthermore, since the cardiovascular risk as a function of these two time indices (vintage and time since infection) is associated with baseline age at dialysis, the models allow for baseline age as a third index.

We note that the literature on the standard varying coefficient models (Cleveland et al., 1991; Hastie and Tibshirani, 1993), generalized varying coefficient models (Cai et al., 2000; Zhang et al., 2004; Qu and Li, 2006), and their adaptations for analyzing longitudinal data (e.g., see Hoover et al., 1998; Wu et al., 2000; Chiang et al., 2001; Fan et al., 2000; 2003; Huang et al., 2002; Huang et al., 2004; and references therein) are limited to a single time index for the varying coefficient functions due to the curse of dimensionality. When understanding how the response trajectory changes with respect to each index is not of interest, a dimension reduction approach where a linear combination of several indices serve as a one-dimensional index of the varying coefficient model was proposed by Fan et al. (2003). Although these approaches are very useful in their respective areas of applications, they are not directly applicable to our objective of modeling/understanding cardiovascular risk over multiple indices.

We also note that although the proposed GM-IVC models are motivated by our goal to better understand the dynamics of cardiovascular risk over several time and age indices for patients on dialysis with the initial infection-related hospitalization as the pivotal exposure,

the models are sufficiently general for a variety of other potential applications. In many longitudinal investigations, a pivotal exposure of interest marks the shift to a new 'time' index for modeling the response trajectory. The remainder of this paper is organized as follows. We introduce our proposed generalized multiple-index varying coefficient model along with model interpretation and assumptions in Section 2. Section 3 outlines the proposed estimation algorithm based on local maximum likelihood. In Section 4, we examine the aforementioned cardiovascular risk trajectories in older patients on dialysis with data from the USRDS. Section 5 contains simulation studies to demonstrate the efficacy of the proposed estimation method, followed by concluding remarks in Section 6.

2 Proposed Generalized Multiple-Index Varying Coefficient Model

2.1 Model Specification

Let a_i denote the age of the ith patient at the initiation of dialysis and S_i denote the survival time of the ith patient. While t_i will be used to denote overall follow-up times after initiation of dialysis, t_{0i} and t_{1i} will specifically track follow-up times before and after the potential infection-related hospitalization, respectively. Hence for patients who had a pivotal initial infection-related hospitalization at time Z_i, we note that $t_i = t_{0i}\, \mathbb{I}\{t_i < Z_i\} + (Z_i + t_{1i})\, \mathbb{I}\{t_i \geq Z_i\}$, where $\mathbb{I}\{A\}$ denotes the indicator function for event A. For patients who do not experience a pivotal infection during follow up time, $Z_i = 0$ and $t_i = t_{0i}$. To examine the changes in cardiovascular event probability (risk) while conditioning on survival status over these time indices, we model the binary indicator of having a cardiovascular event within a three month follow-up interval. We consider a binary outcome instead of a count outcome in our modeling, since having more than one cardiovascular event in a three month interval is very rare; it is less than 0.1% in our data. The goal is to model the expected outcome, denoted

$$\mu_i \equiv \mu_i(a_i, t_i, t_{0i}, t_{1i}) = E\{Y_i(a_i, t_i, t_{0i}, t_{1i}) | Z_i, X_i, \mathbb{I}_{\mathrm{P}_i}(t_i), S_i > t_i\},$$

where $Y_i(a_i, t_i, t_{0i}, t_{1i})$ is the indicator of a cardiovascular event for subject i in a three month time interval centered around a fixed value of t_{0i} or t_{1i}; Z_i is the vintage till first infection-related hospitalization given that the ith patient has at least one infection-related hospitalization ($Z_i = 0$ for patients who do not experience an infection-related hospitalization); $\mathbb{I}_{\mathrm{P}_i}(t_i)$ denotes a subject-specific time-varying indicator of infection-related hospitalization *prior* to time t_i (i.e. equals 1 for $Z_i > t_i$ and zero otherwise); X_i is a vector of $p - 1$ additional baseline covariates. A link (transformation) function, denoted $g(\mu_i)$, connects the conditional expected outcome (cardiovascular event risk) to the time-varying effects and age-varying effects corresponding to the multiple time indices and the covariates. More precisely, our proposed generalized multiple-index varying coefficient model has the form:

$$g(\mu_i) = \gamma_0(a_i)\alpha_0(t_{0i})\{1 - \mathbb{I}_{\mathrm{P}_i}(t_i)\} + \gamma_1(a_i)\alpha_1(t_{1i})\mathbb{I}_{\mathrm{P}_i}(t_i) + \beta_1(a_i)Z_i\mathbb{I}_{\mathrm{P}_i}(t_i) + \sum_{r=2}^{p}\beta_r(a_i)X_{ri}, \quad (1)$$

where the term $\gamma_0(a_i)\alpha_0(t_{0i})$ jointly captures vintage- and age-varying effects; the term $\gamma_1(a_i)\alpha_1(t_{1i})$ captures the age- and time-varying effects since the initial infection-related

hospitalization; the age-varying coefficient functions $\beta_r(a_i)$, $r = 1, \ldots, p$, correspond to vintage prior to the first infection-related hospitalization and baseline covariates. For formality, the supports for the varying coefficient functions in (1) are: $t_{0i} \in [0, T_{0i}]$, $t_{1i} \in [0, T_{1i}]$, $T_{0i} \leq T$, $T_{1i} \leq T$ and $a_i \in [A_0, A_1]$, where T is the maximum follow-up duration along each time axis. In our application, we model the cardiovascular event risk during a maximum follow-up period along each time axis with $T = 5$ years, both after the initiation of dialysis and after the initial infection-related hospitalization. We estimate the age-varying effects for $a_i \in [65, 90] = [A_0, A_1]$. The target population is older patients on dialysis since the cardiovascular event probability is expected to be higher in this cohort.

Note that for our application, the outcome $Y_i(a_i, t_i, t_{0i}, t_{1i})$ is binary so that $\mu_i = \Pr\{Y_i(a_i, t_i, t_{0i}, t_{1i}) = 1 | Z_i, X_i, \mathbb{I}_{P_i}(t_i), S_i > t_i\}$ and we use the logit link function, $\text{logit}(\mu_i) = \log\{\mu_i/(1 - \mu_i)\}$. Finally, we note that a classical generalized varying coefficient model with a single time index is a special case of our proposed GM-IVC model (1). More specifically, when $\alpha_0(t_{0i})$ and $\alpha_1(t_{1i})$ are constant functions, model (1) reduces to a standard generalized baseline age-varying coefficient model with cross-sectional covariates (X_{ri}'s) and longitudinal covariates (namely, $\mathbb{I}_{P_i}(t_i)$ and $Z_i\,\mathbb{I}_{P_i}(t_i)$),

$$g(\mu_i) = \gamma_0^*(a_i) + \{\gamma_1^*(a_i) - \gamma_0^*(a_i)\}\mathbb{I}_{P_i}(t_i) + \beta_1(a_i) Z_i \mathbb{I}_{P_i}(t_i) + \sum_{r=2}^{p} \beta_r(a_i) X_{ri}.$$

Other simplifications such as parametric forms can be considered for the varying coefficient functions. The proposed model is given for the most general setting for potentially complex features of the varying coefficient functions in diverse applications allowing for nonparametric forms along each time index.

2.2 Model Interpretation and Assumption

The proposed GM-IVC model (1) adapts to the follow-up time periods of patients before and after a potential infection-related hospitalization in order to model changes in the cardiovascular event risk over several time and age indices. This aspect is illustrated in Figure 1. With respect to vintage, model (1) reduces to

$$g(\mu_i) = \gamma_0(a_i)\alpha_0(t_{0i}) + \sum_{r=2}^{p} \beta_r(a_i) X_{ri}. \quad (2)$$

Note that for modeling infection-free vintage, times from subjects who had no infection-related hospitalization during their entire follow-up and times prior to the initial infection-related hospitalization for patients with at least one infection-related hospitalization contribute to model (2). Hence, before a potential infection-related hospitalization, the risk of a cardiovascular event is modeled as a function of baseline age a_i (at the start of dialysis), vintage t_{0i} and baseline covariates, whose effects are allowed to vary with baseline age. On the other hand, for those subjects with at least one infection-related hospitalization, after their initial infection-related hospitalization, model (1) shifts to

$$g(\mu_i)=\gamma_1(a_i)\alpha_1(t_{1i})+\beta_1(a_i)Z_i+\sum_{r=2}^{p}\beta_r(a_i)X_{ri}. \quad (3)$$

Therefore, after the initial infection-related hospitalization, we model the cardiovascular event risk primarily as a function of baseline age a_i and time since the initial infection-related hospitalization t_{1i} (along with baseline covariates). Thus, the infection-related hospitalization introduces an additional time index, namely time since the initial infection-related hospitalization. Note that model (3) also accounts for vintage till the initial infection-related hospitalization (Z_i). The time varying indicator, namely $\mathbb{I}_{[t_i]}(t_i)$, in model (1) allows the switch between models (2) and (3) determined by the time of the initial infection-related hospitalization during a patient's course of dialysis. This flexibility will allow us to study the longitudinal effects of a pivotal initial infection-related hospitalization on the cardiovascular event risk and also compare these effects with the longitudinal effects along the time since dialysis index. In this respect the proposed model does have similarities with a change point varying coefficient model with subject specific change points at the potential initial infection-related hospitalizations. However the main innovation of the proposed GM-IVC model remains in that it can accommodate multiple time indices which is also novel in the change point models for survival or longitudinal data.

Model (1) addresses the curse of dimensionality from accommodating multiple time indices via the multiplicative forms $\gamma_0(a)\alpha_0(t_0)$ and $\gamma_1(a)\alpha_1(t_1)$. Hence, effects along the two time indices and age are modeled through one-dimensional coefficient functions, rather than bivariate varying coefficient functions (e.g., $h_0(a, t_0)$ and $h_1(a, t_1)$); this leads to easier interpretation and more straight forward comparisons along different time indices. The proposed multiplicative forms in model (1) are not identifiable without restrictions, hence we assume the following identifiability conditions:

$$\int_0^T \alpha_0^2(t_0)dt_0=1, \quad \int_0^T \alpha_1^2(t_1)dt_1=1, \quad \alpha_0(0)>0 \quad \text{and} \quad \alpha_1(0)>0. \quad (4)$$

These identifiability conditions imply that the estimated effects along the time indices, t_0 and t_1, are normalized and that the estimated coefficients, $\alpha_0(\cdot)$ and $\alpha_1(\cdot)$, carry the shapes of the regression effects, while the magnitude and the sign of the effects are reflected through the coefficient functions, $\gamma_0(a)$ and $\gamma_1(a)$. We note that the assumed multiplicative forms along with the proposed identifiability conditions imply that the cardiovascular risk probabilities as a function of vintage and time since the initial infection-related hospitalization, for patients initiating dialysis at different ages, share a common shape captured by $\alpha_0(t_0)$ and $\alpha_1(t_1)$, respectively. Also, the magnitude of these trajectories are allowed to change as functions of baseline age at dialysis. We will illustrate in the analysis of the USRDS data in Section 4 that the plausibility of the assumed multiplicative forms can be easily assessed graphically during the implementation of the proposed estimation algorithm.

3 Two-Step Estimation Via Local Maximum Likelihood

For estimation in the GM-IVC model, we propose a two-step estimation algorithm that utilizes an extension of the local maximum likelihood estimator of Cai et al. (2000) to longitudinal data. In the first step, we target $\alpha_0(t_0)$ and $\alpha_1(t_1)$ based on the observation that for fixed a, the proposed model reduces to a generalized varying coefficient model in t_0 and t_1, both indexing the longitudinal follow-up of each subject. We bin patients according to baseline age a and obtain stratified estimates of the varying coefficient functions within each bin. This is equivalent to estimating slices of the two dimensional surfaces $h_0(a, t_0) = \gamma_0(a)\alpha_0(t_0)$ and $h_1(a, t_1) = \gamma_1(a)\alpha_1(t_1)$ at fixed a values. Hence we estimate features of the two-dimensional surfaces by estimation in one dimension. Since the stratified estimates share a common shape according to our identifiability conditions (4), we combine and normalize the stratified estimates to obtain our final estimators for $\alpha_0(t_0)$ and $\alpha_1(t_1)$. The second step of the proposed estimation algorithm utilizes the observation that for known $\alpha_0(t_0)$ and $\alpha_1(t_1)$, the proposed model reduces to a baseline age-varying coefficient model in a with longitudinal and cross-sectional covariates. Thus, using the estimated $\alpha_0(t_0)$ and $\alpha_1(t_1)$ from the first step, we estimate $\gamma_0(a)$, $\gamma_1(a)$ and $\{\beta_r(a); r = 1, \ldots, p\}$ of the baseline age-varying coefficient model in the second step of the estimation algorithm.

3.1 Step I: Estimation of $\alpha_0(t_0)$ and $\alpha_1(t_1)$

We begin by binning the subjects according to their baseline age a_i. In our application to the USRDS data, we use two year intervals. Denote by $\{\mathcal{A}_j; j = 1, \ldots, J\}$ the disjoint sets of patient indices that partition the cohort. Next, in each age bin $\mathcal{A}_j$, we partition each patient's follow-up period into disjoint three month intervals both after the start of dialysis and after the initial infection-related hospitalization if the patient has at least one infection-related hospitalization. For time since dialysis, patients are followed up to their initial infection-related hospitalization or to the end of follow-up (for patients with no infection-related hospitalization). For time since the initial infection-related hospitalization, patients are followed to the end of their follow-up.

In our application we consider five year maximum follow-up periods, $T = 5$ in model (1), both after the start of dialysis and after the initial infection-related hospitalization, since the median follow-up in the entire cohort is approximately 2 years. Define t_{0ik} and $t_{1ik'}$ to be the midpoints of the kth and k'th three month time intervals since dialysis start and time since the initial infection-related hospitalization intervals, respectively. Also, let $i(j)$ denote the ith patient in the age bin $\mathcal{A}_j$. We define the binary response variable $Y_{0,ijk} \equiv Y_{i(j)}(a_i, t_{i(j)} = t_{0i(j)} = t_{0i(j)k}) = 1$, if the ith patient in baseline age bin $\mathcal{A}_j$ had at least one cardiovascular event in the kth three month interval after the start of dialysis. Similarly, $Y_{1,ijk'} \equiv Y_{i(j)}(a_i, t_{i(j)} = Z_{i(j)} + t_{1i(j)}, t_{1i(j)} = t_{1i(j)k'}) = 1$ if the ith patient in age bin $\mathcal{A}_j$ had at least one cardiovascular event in the k'th three month interval after the initial infection-related hospitalization. In addition, we denote by $X_{ri(j)}$ and $Z_{i(j)}$ the value of the rth baseline covariate and the vintage until the initial infection-related hospitalization of patient i in age bin $\mathcal{A}_j$, respectively.

We note that for a fixed age a, for patients within the age stratum/bin $\mathcal{A}_j$, the proposed GM-IVC model (1) reduces to the following varying coefficient model in the longitudinal follow-up time ($t_{i(j)}$, $t_{0i(j)}$ and $t_{1i(j)}$ all tracking longitudinal time),

$$g(\mu_{i(j)})=\alpha_{0j}(t_{0i(j)})\{1-\mathbb{I}_{P_{i(j)}}(t_{i(j)})\}+\alpha_{1j}(t_{1i(j)})\mathbb{I}_{P_{i(j)}}(t_{i(j)})+b_{1j}Z_{i(j)}\mathbb{I}_{P_{i(j)}}(t_{i(j)})+\sum_{r=2}^{p}b_{rj}X_{ri(j)}, \quad (5)$$

where $\mu_{i(j)} \equiv E\{Y_{i(j)}(a_{i(j)}, t_{i(j)}, t_{0i(j)}, t_{1i(j)})|Z_{i(j)}, X_{ri(j)}, \mathbb{I}_{P_{i(j)}}(t_{i(j)}), S_{i(j)} > t_{i(j)}\}$, $g(\cdot)$ is a known link function, $\alpha_{0j}(t_{0i(j)}) \equiv \gamma_0(a_{i(j)})\, \alpha_0(t_{0i(j)})$, $\alpha_{1j}(t_{1i(j)}) \equiv \gamma_1(a_{i(j)})\, \alpha_1(t_{1i(j)})$, $b_{1j} \equiv \beta_1(a_{i(j)})$ and $\{b_{rj} \equiv \beta_r(a_{i(j)}); r = 2, \ldots, p\}$. Model (5) is fitted via local maximum likelihood. For details of the fitting procedure including the adopted Newton-Raphson algorithm, see Web Appendix A.

Note that the stratified estimators from different $\mathcal{A}_j$'s target $\alpha_{0j}(t_0) \equiv \gamma_0(a_j)\alpha_0(t_0)$ and $\alpha_{1j}(t_1) \equiv \gamma_1(a_j)\alpha_1(t_1)$, and that they share the same shape as $\alpha_0(t_0)$ and $\alpha_1(t_1)$, respectively. Hence, to arrive at our final estimators for $\alpha_0(t_0)$ and $\alpha_1(t_1)$, we aggregate the stratified estimators coming from different age strata using the identifiability conditions via

$$\hat{\alpha}_0(t_0)=\frac{\sum_j \hat{\alpha}_{0j}(t_0)}{\left[\int_0^5\{\sum_j \hat{\alpha}_{0j}(t_0)\}^2 dt_0\right]^{1/2}}(-1)^{\mathcal{I}_0}, \quad \hat{\alpha}_1(t_1)=\frac{\sum_j \hat{\alpha}_{1j}(t_1)}{\left[\int_0^5\{\sum_j \hat{\alpha}_{1j}(t_1)\}^2 dt_1\right]^{1/2}}(-1)^{\mathcal{I}_1},$$

where $\mathcal{I}_0$ and $\mathcal{I}_1$ denote the indicator functions for $\sum_j \hat{\alpha}_{0j}(0)<0$ and $\sum_j \hat{\alpha}_{1j}(0)<0$.

Note that the number of bins selected does not need to be large as long as there is enough sample size to obtain stable estimates from the fitted generalized varying coefficient models in each age bin. There are a couple of factors that play a role in determining adequate sample size for fitting a generalized varying coefficient model: 1) nature of the response (e.g. continuous or binary), 2) number of predictors in the model, 3) amount of truncation by death. For a more detailed discussion of selection of number of bins, readers are referred to Web Appendix D.

3.2 Step II: Estimation of $\gamma_0(a_i)$, $\gamma_1(a_i)$ and $\beta_r(a_i)$

For estimation of the γ's and β's, we observe that for known $\alpha_0(t_0)$ and $\alpha_1(t_1)$, the proposed GM-IVC model reduces to a varying coefficient model in the single age index a. Therefore, we use the estimators of $\alpha_0(t_0)$ and $\alpha_1(t_1)$ from step I to target the γ's and the β's in the baseline age-varying coefficient model

$$g(\mu_i)=\gamma_0(a_i)\hat{\alpha}_0(t_{0i})\{1-\mathbb{I}_{P_i}(t_i)\}+\gamma_1(a_i)\hat{\alpha}_1(t_{1i})\mathbb{I}_{P_i}(t_i)+\beta_1(a_i)Z_i\mathbb{I}_{P_i}(t_i)+\sum_{r=2}^{p}\beta_r(a_i)X_{ri}, \quad (6)$$

in the second step of the estimation algorithm. We fit model (6) via maximum likelihood; see Web Appendix A for details of the maximization algorithm.

4 Multiple-Index Cardiovascular Event Risk Trajectories in Older Patients on Dialysis

4.1 Description of the Study Cohort

We use data from the United States Renal Data System (USRDS), which collects data on nearly all patients with end-stage renal disease in the U.S. The USRDS is a national database that collects and maintains standard analytic files, including data on inpatient hospitalizations submitted to Medicare, patient demographics, dialysis modality, comorbidities and laboratory measures at the start of dialysis (USRDS, 2011). The defined population of inference in our study are patients aged 65 and older who newly initiated dialysis between January 1, 2000 and December 31, 2007 without a prior history of renal transplant. For detailed descriptions of the study cohort of n = 294, 511 patients and variables used in modeling including baseline covariates and definitions of a cardiovascular event and the infection-related hospitalization, see Web Appendix B.

4.2 Results: Cardiovascular Event Risk Trajectories

We begin our proposed estimation procedure for the GM-IVC model (1) by obtaining the age-stratified α_{0j} and α_{1j} estimates. For this, we bin patients into 2 year baseline age strata, where bins are a little wider at 3 years for strata above age 84, to obtain stable estimates at very high ages, yielding a total of 11 bins. A sensitivity analysis has been run where the total number of bins were selected as 8 and 14; data analysis results were very similar and readers are referred to Web Appendix D for details. The age-stratified estimates ($\hat{\alpha}_{0j}$'s and $\hat{\alpha}_{1j}$'s) are plotted as a function of vintage and time (years) since the pivotal initial infection-related hospitalization in Figure 2(a) and (b), respectively. The plotted age-stratified estimates roughly share a similar increasing pattern, indicating that the multiplicative assumption of model (1) is reasonable in our application.

The final estimated time-varying coefficient functions over the two time indices, namely $\hat{\alpha}_0(t_0)$ and $\hat{\alpha}_1(t_1)$, along with the age-varying coefficient functions $\hat{\gamma}_0(a)$ and $\hat{\gamma}_1(a)$ are displayed in Figure 2(c) and (d), respectively. Also provided along with the cardiovascular event risk trajectories are percentile bootstrap confidence intervals based on 200 bootstrap replications formed by resampling from subject trajectories with replacement. The bandwidths used (h = 1.5 years for $\{\hat{\alpha}_0(t_0), \hat{\alpha}_1(t_1)\}$ in the first step and h = 4 years for $\{\hat{\gamma}_0(a), \hat{\gamma}_1(a), \hat{\beta}_r(a)\}$ in the second step of the local maximum likelihood estimation procedure) were chosen by the 20-fold cross-validation, similar to Cai et al. (2000). To reduce boundary effects, a bandwidth of h = 2.5 years was used at grid points close to 5 years in estimation of $\hat{\alpha}_1(t_1)$. Recall (from Section 2.2) that in assessing the estimated varying coefficient functions, $\hat{\alpha}_0$ and $\hat{\alpha}_1$ do not carry the sign or the magnitude of the regression effects, and they should be compared based on their shapes. Because $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are negative (Figure 2(d)), the general increasing patterns of $\hat{\alpha}_0$ and $\hat{\alpha}_1$ (in Figure 2(c)), with respect to both time indices (t_0 and t_1), imply decreasing cardiovascular event probabilities after the start of dialysis and after the initial infection-related hospitalization for the dynamic cohort of survivors. Consistent with the slight convex pattern of $\hat{\alpha}_1$, around 0, we estimate a faster decrease in cardiovascular event probabilities after the start of dialysis compared to follow-up after the initial infection-related hospitalization. From Figure 2(d), it can be seen

that the estimated age-varying coefficient function $\hat{\gamma_1}(a)$ is greater than $\hat{\gamma_0}(a)$; this implies that cardiovascular event risk is nominally higher across all ages conditional on survival status after the initial infection-related hospitalization compared to after the start of dialysis. Furthermore, both $\hat{\gamma_1}(a)$ and $\hat{\gamma_0}(a)$ are increasing with age *a* and converging together as age *a* approaches 90; this suggests, not surprisingly, that cardiovascular event risk increases with age generally.

To compare the cardiovascular event risk trajectories directly, as a function of vintage, time since the initial infection-related hospitalization and patient age at dialysis, Figure 3 provides the cardiovascular event probabilities and their respective bootstrap confidence intervals over both time indices for baseline ages of 65, 78 and 90. The following observations can be made about the cardiovascular risk trajectories from the results in Figure 3.

1. The risk is significantly greater across the five year follow-up time after the pivotal initial infection-related hospitalization compared to the time after the start of dialysis conditional on survival of the patient.
2. The highest risk is near the time of dialysis start and the time of the initial infection-related hospitalization and declines with both time indices.
3. The risks over both time indices increase with increasing baseline age.
4. The effect of baseline age on the risk of cardiovascular event is much more pronounced for time after the start of dialysis compared to the time after the initial infection-related hospitalization. Furthermore, from Figure 2(d), the increasing cardiovascular event probability among the dynamic cohort of survivors after the initial infection-related hospitalization plateaus after baseline age 72.
5. However, the *difference* in risks for time since dialysis and time since the initial infection-related hospitalization declines with increasing baseline age at dialysis.

The later two points are made transparent by Figure 3(d), which overlays the estimated cardiovascular event probabilities across baseline ages. To illustrate the pattern of cardiovascular risk dynamics above, we selected the estimates for white male patients with diabetes and average levels of eGFR and BMI to display in Figure 3.

Figure 4 displays the estimated cardiovascular event risk trajectories for both time indices simultaneously, with the initial infection-related hospitalization occurring at 3, 2 and 1 year(s) after the start of dialysis; similar to Figure 3, the risk trajectories are provided with bootstrap confidence intervals for individuals with baseline ages of 65, 78 and 90 (Figure 4: left, middle and right column, respectively). The increased cardiovascular event probabilities remain elevated and do not decrease to their original levels; for example, even after one year from the initial infection-related hospitalization. This appears to hold independent of when the initial infection related-hospitalization occurred (1, 2 or 3 years after the start of dialysis). Also, consistent with impact of baseline age at dialysis described above, the sustained elevated risks are particularly pronounced for relatively younger patients at the start of dialysis (e.g., age 65 and 78 compared to age 90). For the effects of baseline covariates, see Web Appendix B.

5 Simulation Studies

We carry out simulation studies to examine the efficacy of the proposed estimation procedure to target the true time- and age-varying coefficient functions when there is truncation by death. Similar to the model used for the data analysis of Section 4, we consider the following GM-IVC model

$$g(\mu_i)=\gamma_0(a_i)\alpha_0(t_{0i})\{1-\mathbb{I}_{\mathrm{P}_i}(t_i)\}+\gamma_1(a_i)\alpha_1(t_{1i})\mathbb{I}_{\mathrm{P}_i}(t_i)+\beta_1(a_i)Z_i\mathbb{I}_{\mathrm{P}_i}(t_i)+\beta_2(a_i)X_{2i}+\beta_3(a_i)X_{3i},$$

for $t_{0i} \in [0, T_{0i}]$, $t_{1i} \in [0, T_{1i}]$, $T_{0i} \le T$, $\mathrm{T}_{1i} \le T$, $T = 5$, $a_i \in [65, 90]$ and $g(\cdot)$ the logistic link function. For details of the simulation study design where the response and the survival time are generated jointly, see Web Appendix C.

To study the performance of the proposed estimation procedure for the GM-IVC, we utilize a relative mean squared deviation error (MSDE) defined as

$$\mathrm{MSDE}_{\alpha_0}=\left[\int_0^T\{\alpha_0(t_0)-\hat{\alpha}_0(t_0)\}^2dt_0\right]/\int_0^T\alpha_0^2(t_0)dt_0$$

for the time-vary function $\alpha_0(t_0)$. The MSDEs for the other time- and age-varying coefficient functions, namely $\mathrm{MSDE}_{\alpha 1}$, $\mathrm{MSDE}_{\gamma 0}$, $\mathrm{MSDE}_{\gamma 1}$, $\mathrm{MSDE}_{\beta 1}$, $\mathrm{MSDE}_{\beta 2}$ and $\mathrm{MSDE}_{\beta 3}$, are defined similarly. Table 1 gives the median, first quartile and third quartile of the estimated MSDE values in percentage for the varying coefficient functions over 200 Monte Carlo runs. Results are presented at two sample sizes $n = 3000$ and 5000 with bandwidth chosen by 20-fold cross-validation as described in Cai et al. (2000). Bandwidths used were chosen in a preliminary simulation study yielding $h = (1.5, 1.5)$ for $\hat{\alpha_0}(t_0)$, $\hat{\alpha_1}(t_1)$ and $h = (3.75, 2.5)$ for $\hat{\beta_0}(a)$, $\hat{\beta_1}(a)$, $\hat{\beta_2}(a)$ for $n = 3000$ and $n = 5000$, respectively. Overall, the MSDE values reported are small and get smaller with increased sample size. In addition, Figure 5 displays the estimated median, 5th and 95th percentile varying coefficient functions overlaying the true curves at the sample size $n = 3000$. The estimated functions track the true varying coefficient functions. Overall, the simulation studies illustrate that the proposed estimation procedure for the GM-IVC models is effective in capturing the true time- and age-varying dynamics for data truncated by death.

6 Discussion

Infection and cardiovascular disease remain the leading causes of hospitalization and death in patients on dialysis in the United States (USRDS 2011). Understanding the complex cardiovascular risk trajectories is important for potential strategies to target cardiovascular risk reduction, including implementation of overall infection control or prevention strategies. The results highlight the significant impact of the first infection-related hospitalization on cardiovascular event risk and the dependence of this effect on age at the start of dialysis. An important finding is that the infection-related hospitalization results in *sustained* increases in cardiovascular event risk among the dynamic cohort of survivors; for instance, even one year after the infection-related hospitalization, the cardiovascular event

probability is still higher than the cardiovascular risk at the start of dialysis, a time of high cardiovascular risk with respect to vintage. This pattern of cardiovascular risk dynamics, with respect to vintage (t_0) and time since the pivotal initial infection-related hospitalization (t_1), holds for most older patients starting dialysis, although the difference in cardiovascular event probabilities before and after the infection declines with increasing baseline age at dialysis. The difference in cardiovascular risk converges (equalizes) only for very elderly patients starting dialysis, near 90 years of age.

From a technical perspective, the proposed GM-IVC models add important and necessary flexibility to the current varying coefficient modeling toolkit by the introduction of additional time (and age) indices. We achieved this by employing a sensible multiplicative structure to capture the multiple time- and age-varying effects, and at the same time, avoiding the curse of dimensionality, a known limiting factor in modeling dynamic, varying effects. The multiplicative structure assumption can be assessed in the first step of the proposed estimation algorithm via assessing whether the binned estimators of $\alpha_0(t_0)$ and $\alpha_1(t_1)$ share a common shape. If this does not seem to be a plausible assumption, the more general two dimensional regression surfaces $h_0(a, t_0)$ and $h_1(a, t_1)$ need to be targeted. Also, as illustrated with our USRDS data application, the GM-IVC models provide natural graphical displays of time- and age-varying dynamics that are fairly easy to interpret; thus, retaining a popular feature of standard/classical varying coefficient models. We believe the proposed GM-IVC models are widely applicable since characterizing the outcome trajectories over multiple indices, including time since a pivotal exposure event is often of interest in longitudinal analysis. Finally due to a large percent of truncation by death, we developed a model targeting a partly conditional inference target, conditional on the survival status of the patients. Partly conditional models was originally proposed for generalized linear models (Kurland and Heagerty, 2005); the current work extends them to varying coefficient models incorporating multiple indices. Investigation of the theoretical properties of the proposed estimators is an open problem. We provide R codes for our GM-IVC model in Web Appendix E.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank two referees, the associate editor and the editor for their valuable comments. This publication was made possible by grants R01 DK092232 (DS, DVN, LSD) and K23 DK093584 (LSD) from the National Institute of Diabetes and Digestive and Kidney Diseases, by grant UL1 TR000153 (DVN) and UL1 TR000002 (LSD) from the National Center for Advancing Translational Sciences and by a research grant from Dialysis Clinic Inc. We thank Barbara Grimes, Department of Biostatistics, University of California, San Francisco. 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. This study was approved by the Institutional Review Board of the University of California, Davis.

References

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

Chiang CT, Rice JA, Wu CO. Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. Journal of the American Statistical Association. 2001; 96:605–619.

Cleveland, WS.; Grosse, E.; Shyu, WM. Local regression models. In: Chambers, JM.; Hastie, TJ., editors. Statistical Models in S. Wadsworth & Brooks; Pacific Grove: 1991. p. 309-376.

Dalrymple LS, Johansen KL, Chertow GM, Cheng S, Grimes B, Gold EB, Kaysen GA. Infection-related hospitalizations in older patients with end-stage renal disease. American Journal of Kidney Diseases. 2010; 56:522–530. [PubMed: 20619518]

Dalrymple LS, Mohammed SM, Mu Y, Johansen KL, Chertow GM, Grimes B, Kaysen GA, Nguyen DV. The risk of cardiovascular-related events following infection-related hospitalizations in older patients on dialysis. Clinical Journal of the American Society of Nephrology. 2011; 6:1708–1713. [PubMed: 21566109]

Fan J, Yao Q, Cai Z. Adaptive varying-coefficient linear models. Journal of the Royal Statistical Society Series B. 2003; 65:57–80.

Fan J, Zhang JT. Two-step estimation of functional linear model with application to longitudinal data. Journal of the Royal Statistical Society Series B. 2000; 62:303–322.

Foley RN, Guo H, Snyder JJ, Gilbertson DT, Collins AJ. Septicemia in the United States dialysis population, 1991 to 1999. Journal of the American Society of Nephrology. 2004; 15:1038–1045. [PubMed: 15034107]

Hastie T, Tibshirani R. Varying coefficient models. Journal of the Royal Statistical Society Series B. 1993; 55:757–796.

Hoover DR, Rice JA, Wu CO, Yang LP. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika. 1998; 85:809–822.

Huang JZ, Wu CO, Zhou L. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika. 2002; 89:111–128.

Huang JZ, Wu CO, Zhou L. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Statistica Sinica. 2004; 14:763–788.

Ishani A, Collins AJ, Herzog CA, Foley RN. Septicemia, access and cardiovascular disease in dialysis patients: The USRDS Wave 2 study. Kidney International. 2005; 68:311–318. [PubMed: 15954922]

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

Kurland BF, Johnson LL, Egleston BL, Diehr PH. Longitudinal data with follow-up truncated by death: Match the analysis method to research aims. Statistical Science. 2009; 24:211–222. [PubMed: 20119502]

Mohammed SM, Senturk D, Dalrymple LS, Nguyen DV. Measurement error case series models with application to infection-cardiovascular risk in older patients on dialysis. Journal of the American Statistical Association. 2012 in-press.

National Kidney Foundation Inc. Part 5. Evaluation of laboratory measurements for clinical assessment of kidney disease. American Journal of Kidney Diseases. 2002; 39:S76–S110.

Qu A, Li R. Quadratic inference functions for varying coefficient models with longitudinal data. Biometrics. 2006; 62:379–391. [PubMed: 16918902]

Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. New England Journal of Medicine. 2004; 351:2611–2618. [PubMed: 15602021]

United States Renal Data System. Researcher's Guide. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; Bethesda, MD: 2011.

United States Renal Data System. USRDS 2012 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; Bethesda, MD: 2012.

Wu CO, Chiang CT. Kernel smoothing on varying coefficient models with longitudinal dependent variable. Statistica Sinica. 2000; 10:433–456.

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

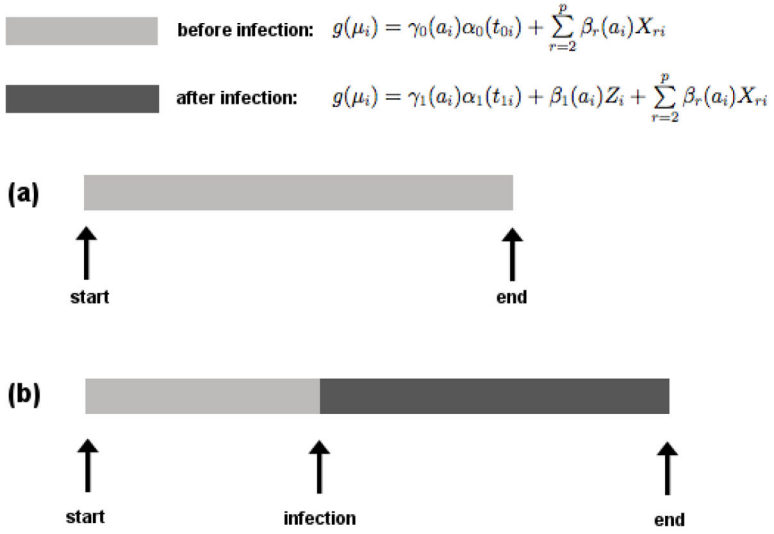

Figure 1.

Example of follow-up data for a subject (a) without and (b) with an infection-related hospitalization along with the proposed models for cardiovascular risk before (light gray) and after (dark gray) the infection-related hospitalization. Note that the model for cardiovascular risk after the initial infection-related hospitalization appropriately accounts for vintage until the infection-related hospitalization (term $\beta_1(a_i)Z_i$). See Section 2 for details.

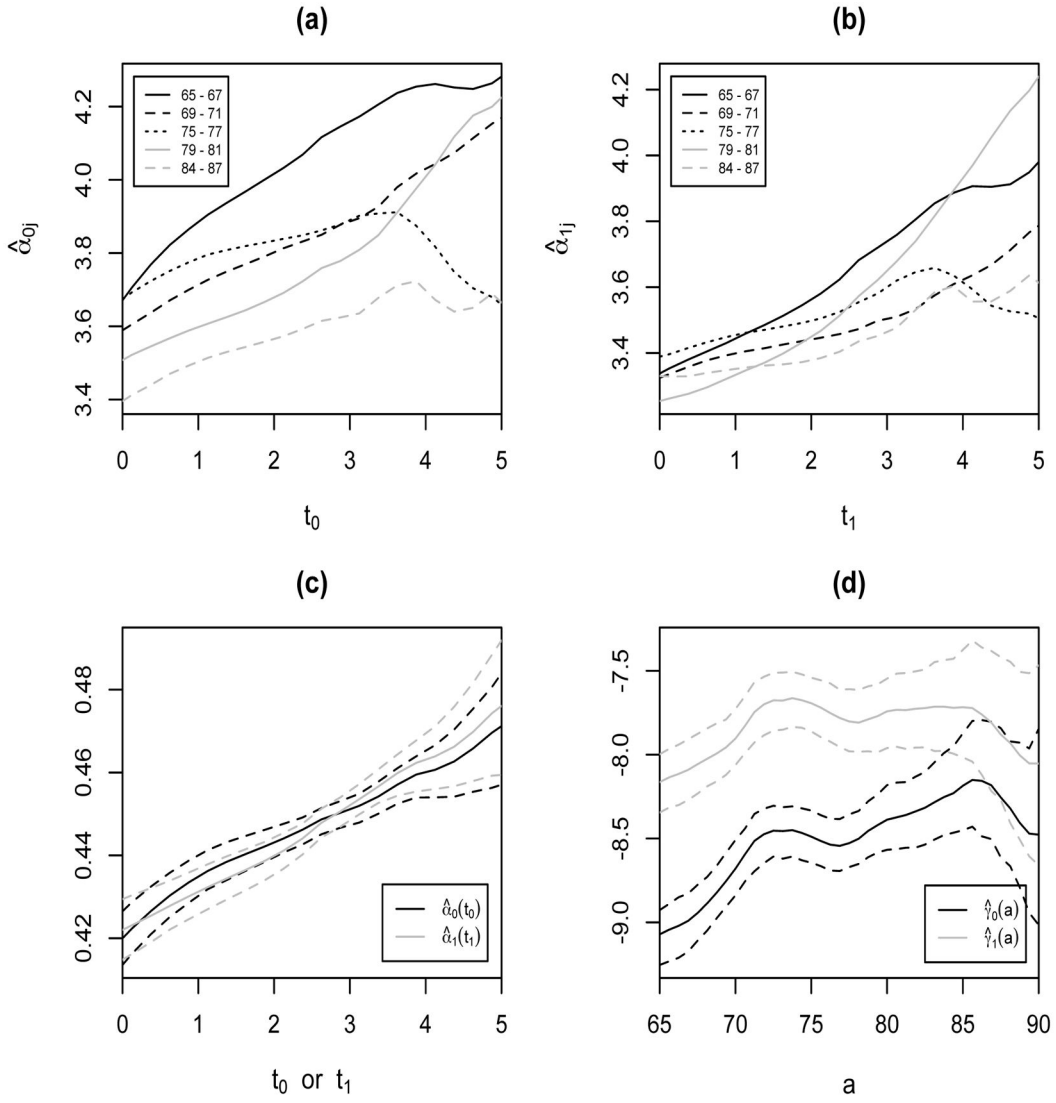

Figure 2.

Baseline age-stratified varying coefficient function estimates after the (a) start of dialysis (vintage t_0) and (b) initial infection-related hospitalization (t_1). Final varying coefficient function estimates as a function of (c) vintage and time since the initial infection-related hospitalization and (d) baseline age at dialysis. 90% bootstrap confidence intervals are given as dashed lines in (c) and (d).

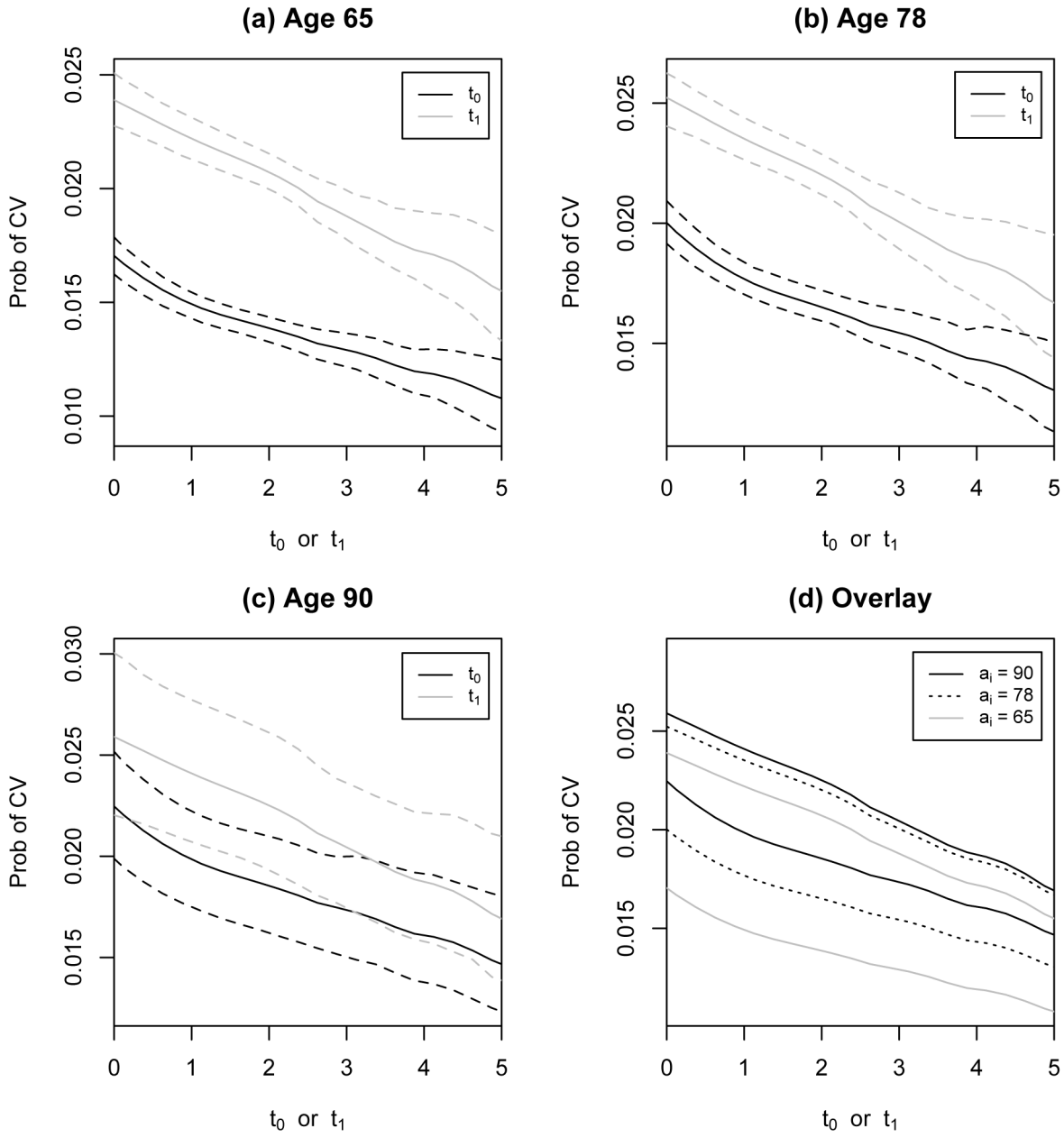

Figure 3.

Estimated probabilities of cardiovascular (CV) events for white male patients with diabetes and with average levels of eGFR and BMI (with the vintage until the first infection-related hospitalization of $Z = 1.4$ years) with baseline ages (a) 65, (b) 78 and (c) 90. Plot (d) overlays/combines the estimated probability trajectories from the three baseline ages. 90% bootstrap confidence intervals are given as dashed lines in (a), (b), and (c).

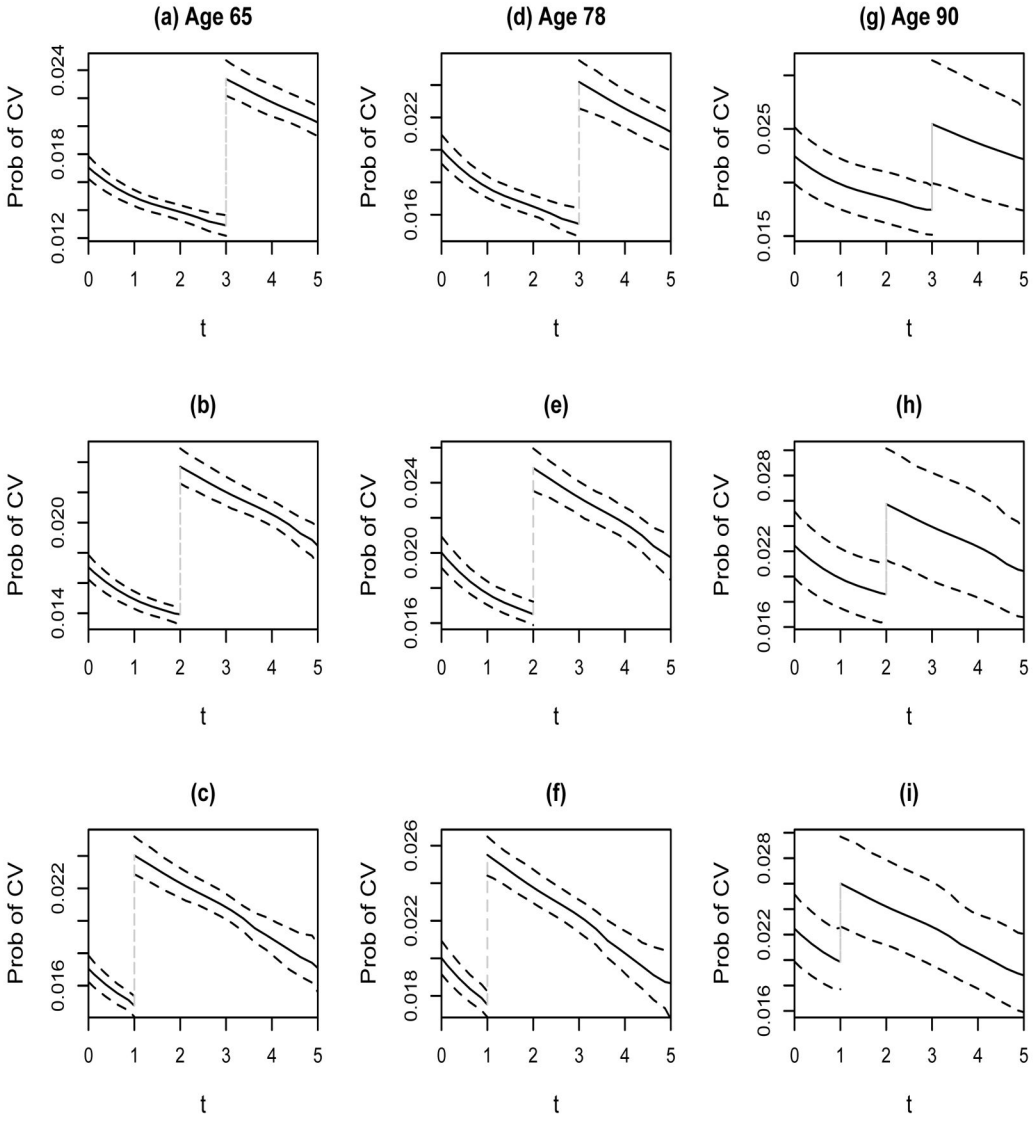

Figure 4.

Estimated probabilities of cardiovascular (CV) events during the course of dialysis for patients experiencing the pivotal initial infection-related hospitalization at 3, 2 and 1 year after the start of dialysis with baseline ages of 65, 78 and 90 (columns left, middle and right, respectively). 90% bootstrap confidence intervals are given dashed.

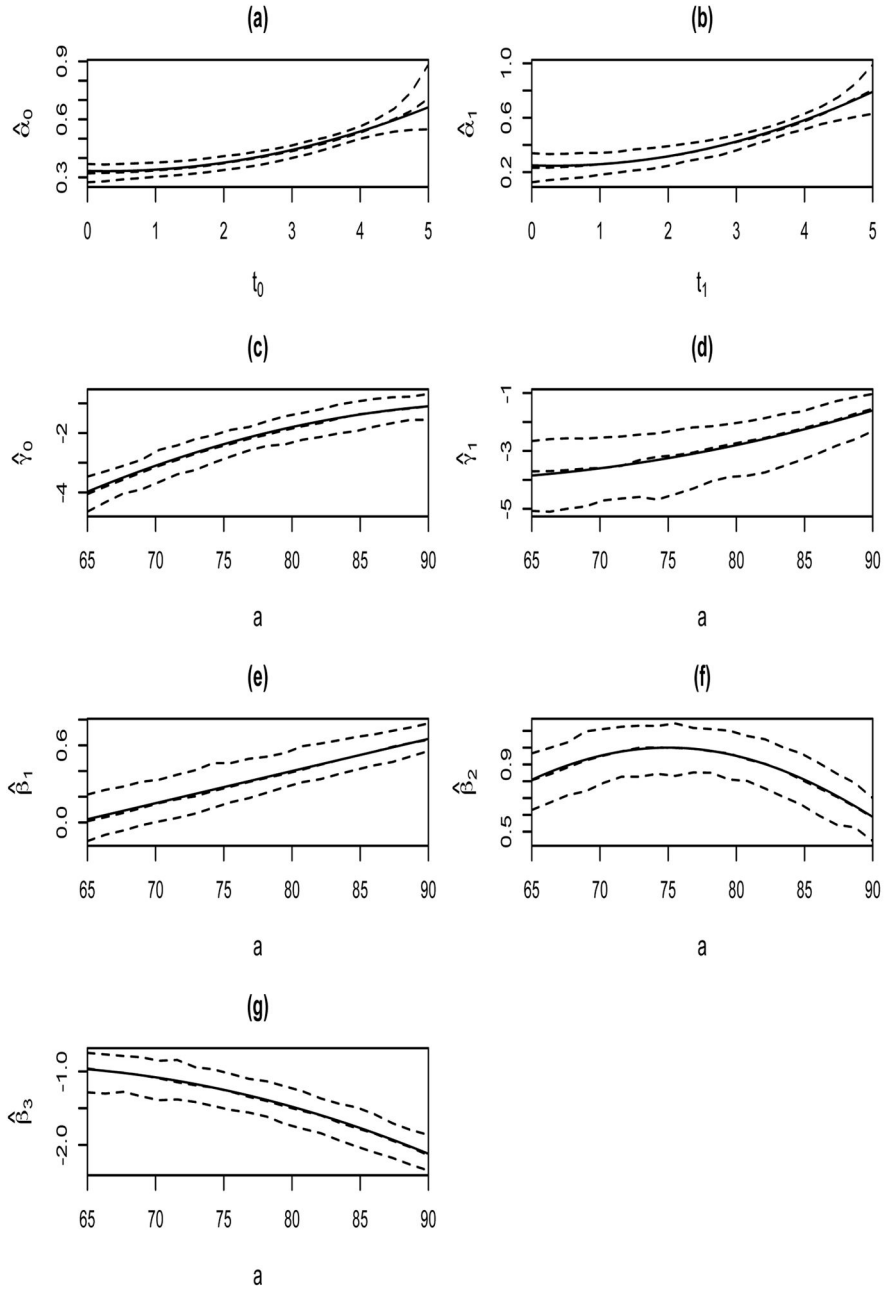

Figure 5.

Simulation results for $n = 3000$. The cross-sectional median curves of the proposed estimates are given along with 5% and 95% cross-sectional percentiles (dotted) overlaying the true varying coefficient functions (solid).

Table 1

Relative mean squared deviation error of the estimated varying coefficient functions from simulation studies. Median and 25th and 75th percentiles of the deviation measures are presented based on 200 Monte Carlo runs.

	$n = 3000$			$n = 5000$		
	Median	**25th Percent**	**75th Percent**	**Median**	**25th Percent**	**75th Percent**
$MSDE_{\alpha 0}$	.0037	.0016	.0068	.0017	.0006	.0030
$MSDE_{\alpha 1}$	.0078	.0038	.0157	.0039	.0022	.0076
$MSDE_{\gamma 0}$	.0127	.0080	.0190	.0109	.0082	.0152
$MSDE_{\gamma 1}$	.0242	.0118	.0545	.0160	.0093	.0338
$MSDE_{\beta 1}$	.0348	.0165	.0735	.0251	.0130	.0481
$MSDE_{\beta 2}$	.0085	.0051	.0140	.0083	.0055	.0111
$MSDE_{\beta 3}$	.0090	.0060	.0134	.0076	.0057	.0105