Published in final edited form as:

Stat Biosci. 2017 December ; 9(2): 504–524. doi:10.1007/s12561-016-9186-4.

Joint Modeling of Repeated Measures and Competing Failure Events In a Study of Chronic Kidney Disease

Wei Yang, Dawei Xie, Qiang Pan, Harold I. Feldman, and Wensheng Guo

Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine

Abstract

We are motivated by the Chronic Renal Insufficiency Cohort (CRIC) study to identify risk factors for renal progression in patients with chronic kidney diseases. The CRIC study collects two types of renal outcomes: glomerular filtration rate (GFR) estimated annually and end stage renal disease (ESRD). A related outcome of interest is death which is a competing event for ESRD. A joint modeling approach is proposed to model a longitudinal outcome and two competing survival outcomes. We assume multivariate normality on the joint distribution of the longitudinal and survival outcomes. Specifically, a mixed effects model is fit on the longitudinal outcome and a linear model is fit on each survival outcome. The three models are linked together by having the random terms of the mixed effects model as covariates in the survival models. EM algorithm is used to estimate the model parameters and the non-parametric bootstrap is used for variance estimation. A simulation study is designed to compare the proposed method with an approach that models the outcomes sequentially in two steps. We fit the proposed model to the CRIC data and show that the protein-to-creatinine ratio is strongly predictive of both estimated GFR and ESRD but not death.

Keywords

Joint modeling; Multivariate normality; Informative dropout; Competing risk; Chronic kidney disease

1 Introduction

In chronic kidney disease (CKD) research, it is often seen that both longitudinal and survival outcomes are studied [9,17,41]. The most popular longitudinal renal outcome is the estimated glomerular filtration rate (eGFR), which is a measure of kidney function. In addition, clinical hard endpoints including end-stage renal disease (ESRD) and death events are also collected. Analyses of these different outcomes are typically done separately depending on the study hypotheses. The issues and challenges of modeling individual longitudinal and survival outcomes have been discussed extensively in literature, for example [6,13,20,18]. Specific to the CKD research, the value of eGFR is not clinically meaningful after a patient has developed ESRD and obviously cannot be collected after a patient's death. This creates the so-called informative dropout for eGFR outcome data because patients who have either died or reached ESRD may have faster eGFR decline over time and therefore have fewer eGFR measures compared to those who are event-free. Similar arguments can be made in the failure time analysis when death is a competing event

for ESRD. Patients who have died may have experienced a higher risk of ESRD if they had not died than those who are still in the study. Due to the correlated nature of these different outcomes, modeling a single outcome without the consideration of the others may produce biased estimates of the parameter of interest.

Joint modeling methods have been proposed to handle informative dropout by modeling the censoring process explicitly. There has been extensive research about joint modeling methods in literature [3,11,14,21,22,24,28,30,32,34,36-38,43]. A typical joint modeling approach includes a version of the mixed effects model for the longitudinal outcome and a survival model for the failure time outcome. The random components specified in the mixed effects model are included as predictors for the survival outcome. The informative dropout is addressed by explicitly modeling the correlation between the underlying longitudinal profile, i.e., the random components in the mixed effects model and the cause of dropout, i.e., the survival outcome. The most popular choice for the survival model is the Cox proportional hazards model. One exception is to use a linear model for the survival outcome proposed by DeGruttola and Tu [3] and independently by Schluchter [27]. In this approach, multivariate normality is assumed for the joint distribution of the longitudinal and log transformed survival outcomes. Vonesh and coauthors [37] generalized the method by considering continuous parametric and semi-parametric models for the survival outcome. In their formulation, the longitudinal data can be modeled using a generalized non-linear mixed effects model and the survival outcome can be modeled using a family of accelerated failure-time models with different distributional assumptions. The class of models include a number of proportional hazards models (e.g. when the survival outcome distribution is Weibull) and log-normal model when the survival outcomes follows a log-normal distribution. Maximum likelihood with Laplace approximation was proposed for parameter estimate and they illustrated the method by modeling the longitudinal GFR data and time to dropout from the MDRD study [17].

Standard competing risk analysis is typically done in the context of estimating the cause specific hazard. Cox proportional hazard models can be used to explore associations adjusted for covariates [23]. The cumulative incidence, which is the probability of failing due to a particular cause in the presence of competing causes, can then be estimated by combining the cause specific hazard of both events. Fine and Gray [10] proposed a proportional sub-distribution hazard model to directly model the association between covariates and the cumulative incidence function of event of interest. The cumulative incidence for a subject with specific covariate level can then be estimated based on the single model. The joint modeling method has also been extended to the case where there are competing failure time events [2,5,8,15,40]. Cause specific sub-distribution hazard models[10] are used to model the competing survival outcomes in all these proposed methods. For a detailed review of joint modeling methods, please see [25,35,42].

The Chronic Renal Insufficiency Cohort (CRIC) study is a multi-center observational study of individuals with chronic kidney disease [9]. One objective of the study is to identify risk factors for kidney disease progression. There are two popular choices to define the outcome of interest. One is the eGFR change over time. The other uses clinical hard endpoints like ESRD. In this paper, our goal is to model the repeated eGFR measures and ESRD together

by taking into account the competing risk by death. Our approach of handling informative dropout and the competing risk is to model the repeated eGFR, ESRD and death outcomes jointly. Following DeGruttola and Tu [3], we assume the joint distribution of the repeated eGFR measures and the log transformed time to ESRD and death follows a multivariate normal distribution. Compared to other joint modeling methods, our models for death and ESRD are essentially the accelerated failure time models which are readily interpreted.

The remainder of this manuscript is organized as follows. Section 2 presents the model for longitudinal and competing survival outcomes and estimation procedure. Section 3 describes a simulation study comparing the proposed joint modeling approach to a naive two-step approach. Section 4 presents the results from the model for CRIC data. We finish the manuscript with a few discussions in Section 5.

2 Methods

2.1 Notation

Without loss of generality, we consider a scenario in which there are two competing failure time events and a single longitudinal outcome. Assume the data for each individual $i = 1, 2, \dots, N$ is $(\mathbf{y}_i; x_{i,1}, x_{i,2}, \tau_i, \delta_i)$. For each individual i , $\mathbf{y}_i = \{y_{ij}, j = 1, \dots, m_i\}$ denotes the repeated longitudinal measures in which m_i is the number of repeated measures. $x_{i,1}$ denotes the time to the first event and $x_{i,2}$ denotes the time to the second event. τ_i is minimum of the two survival event time and censoring time, i.e., $\tau_i = \min(x_{i,1}, x_{i,2}, c_i)$, where c_i is the time of censoring. $\delta_i = 0, 1, 2$ denotes the event that is observed with $\delta_i = 0$ indicating the censoring event. In the competing event setting, at most one failure event is observed.

2.2 Model

A standard mixed effects model can be used to model the repeated longitudinal measures. For each individual $i, i = 1, \dots, N$, the model is

$$\mathbf{y}_i = T_i \alpha + Z_i b_i + \varepsilon_i \quad (1)$$

In the model, T_i is the design matrix for the fixed terms and is $m_i \times p$, where p is the number of fixed covariates. Z_i is the design matrix for the random terms and is $m_i \times k$, where k is the number of random terms. α is a $p \times 1$ vector, representing the coefficients for the fixed effect terms and b_i is $k \times 1$ vector representing the random effects. We further assume that both b_i and ε_i follow multivariate normal distributions, i.e., $b_i \sim N(0, D)$ and $\varepsilon_i \sim N(0, \sigma^2 I_i)$, where D is a $k \times k$ covariance matrix and I_i is a $m_i \times m_i$ identity matrix.

For the two competing failure time events, we fit separately a linear regression model on either the failure time itself or some types of one-to-one transformation, e.g., natural logarithm transformation. For individual $i, i = 1, \dots, N$ and failure time event $j, j = 1, 2$, we have

$$x_{i,j} = W_{i,j}^T \xi_j + \lambda_j^T b_i + r_{i,j} \quad (2)$$

Note that x_{ij} denotes either the failure time itself or its transformation. In model (2), W_{ij} denotes the fixed terms in the model j and is $q \times 1$. ξ_j denotes the coefficients for the fixed terms in model j and is also $q \times 1$. λ_j denotes the coefficients for the random terms b_j in model j and is $k \times 1$. r_{ij} denotes the residual of model j and follows a normal distribution $N(0, s_j^2)$.

Finally we assume b_i , e_i and r_{ij} , $j = 1, 2$ are independent of each other. The parameters in the models are $\theta = \{\alpha, \xi_1, \xi_2, \lambda_1, \lambda_2, \sigma^2, s_1^2, s_2^2, D\}$.

2.3 Estimation

In a study with two competing failure time events, there are three types of individuals in the observed data: 1) those who have an observed failure event one; 2) those who have an observed failure event two and 3) those who are censored without observing either event. We use N_1 , N_2 and N_3 to denote the number of individuals for each type respectively and that the total study sample size is $N = N_1 + N_2 + N_3$. The log likelihood for the observed data is:

$$\begin{aligned} \log L &= \sum_{i=1}^{N_1} \log \left[\int_{b_i} \varphi(y_i | b_i, \alpha, \sigma^2) \varphi(x_{i,1} | b_i, \xi_1, s_1^2) \left\{ 1 - \int_{-\infty}^{x_{i,1}} \varphi(x_{i,2} | b_i, \xi_2, s_2^2) \right\} \varphi(b_i | D) db_i \right] \\ &+ \sum_{i=1}^{N_2} \log \left[\int_{b_i} \varphi(y_i | b_i, \alpha, \sigma^2) \varphi(x_{i,2} | b_i, \xi_2, s_2^2) \left\{ 1 - \int_{-\infty}^{x_{i,2}} \varphi(x_{i,1} | b_i, \xi_1, s_1^2) \right\} \varphi(b_i | D) db_i \right] \\ &+ \sum_{i=1}^{N_3} \log \left[\int_{b_i} \varphi(y_i | b_i, \alpha, \sigma^2) \left\{ 1 - \int_{-\infty}^{c_i} \varphi(x_{i,1} | b_i, \xi_1, s_1^2) \right\} \left\{ 1 - \int_{-\infty}^{c_i} \varphi(x_{i,2} | b_i, \xi_2, s_2^2) \right\} \varphi(b_i | D) db_i \right] \end{aligned}$$

where $\phi(\cdot)$ denotes the density function of a normal distribution.

It is challenging to maximize the observed data likelihood directly as it involves the integration over the random components b_i . Instead, this problem can be solved by treating it as a missing data problem and the EM algorithm [4] can be used to maximize the likelihood. Assuming all data are observed including the random components b_i , the log likelihood for

the full data is $\log L = \sum_{i=1}^N \log L_i$ where

$$\begin{aligned} \log L_i &= \log \{ \varphi(y_i | b_i, \alpha, \sigma^2) \varphi(x_{i,1} | b_i, \xi_1, s_1^2) \varphi(x_{i,2} | b_i, \xi_2, s_2^2) \varphi(b_i | D) \} \\ &= -\frac{m_i}{2} \log(2\pi\sigma^2) - \frac{1}{2} (y_i - T_i\alpha - Z_i b_i)^T (\sigma^2 I)^{-1} (y_i - T_i\alpha - Z_i b_i) \\ &\quad - \frac{1}{2} \log(2\pi s_1^2) - \frac{1}{2} (x_{i,1} - W_{i,j}^T \xi_1 - \lambda_1^T b_i)^2 (s_1^2)^{-1} \\ &\quad - \frac{1}{2} \log(2\pi s_2^2) - \frac{1}{2} (x_{i,2} - W_{i,j}^T \xi_2 - \lambda_2^T b_i)^2 (s_2^2)^{-1} \\ &\quad - \frac{k}{2} \log(2\pi |D|) - \frac{1}{2} b_i^T D^{-1} b_i \end{aligned}$$

EM algorithm includes two steps to maximize the likelihood. The E-step is to calculate the expectation of the log likelihood and the M-step is to maximize the expected log likelihood. If the full data is observed, the maximum likelihood estimate (MLE) of the parameters are:

$$\begin{aligned}
& -\hat{\sigma}^2 = \left(\sum_{i=1}^N m_i \right)^{-1} \left(\sum_{i=1}^N \varepsilon_i^T \varepsilon_i \right) \\
& -\hat{s}_j^2 = N^{-1} \left(\sum_{i=1}^N r_{i,j}^2 \right), j=1, 2 \\
& -\hat{D} = N^{-1} \left(\sum_{i=1}^N b_i b_i^T \right) \\
& -\hat{\alpha} = \left(\sum_{i=1}^N T_i^T T_i \right)^{-1} \left\{ \sum_{i=1}^N T_i^T (y_i - Z_i b_i) \right\} \\
& -\hat{\xi}_j = \left(\sum_{i=1}^N W_{i,j} W_{i,j}^T \right)^{-1} \left\{ \sum_{i=1}^N W_{i,j} (x_{i,j} - \lambda_j^T b_i) \right\}, j=1, 2 \\
& -\hat{\lambda}_j = \left(\sum_{i=1}^N b_i b_i^T \right)^{-1} \left\{ \sum_{i=1}^N (x_{i,j} - W_{i,j}^T \xi_j) b_i \right\}, j=1, 2
\end{aligned}$$

When there are missing data, the missing quantities in the above formulas can be replaced by their expectations. So in the M-step, the MLE solutions are:

$$\begin{aligned}
& -\hat{\sigma}^2 = \left(\sum_{i=1}^N m_i \right)^{-1} \sum_{i=1}^N E \left(\varepsilon_i^T \varepsilon_i | Y_{obs}, \hat{\theta} \right) \\
& -\hat{s}_j^2 = N^{-1} \sum_{i=1}^N E \left(r_{i,j}^2 | Y_{obs}, \hat{\theta} \right), j=1, 2 \\
& -\hat{D} = N^{-1} \sum_{i=1}^N E \left(b_i b_i^T | Y_{obs}, \hat{\theta} \right) \\
& -\hat{\alpha} = \left(\sum_{i=1}^N T_i^T T_i \right)^{-1} \sum_{i=1}^N T_i^T \left\{ y_i - Z_i E \left(b_i | Y_{obs}, \hat{\theta} \right) \right\} \\
& -\hat{\xi}_j = \left(\sum_{i=1}^N W_{i,j} W_{i,j}^T \right)^{-1} \sum_{i=1}^N W_{i,j} \left\{ E \left(x_{i,j} | Y_{obs}, \hat{\theta} \right) - \hat{\lambda}_j^T E \left(b_i | Y_{obs}, \hat{\theta} \right) \right\}, j=1, 2 \\
& -\hat{\lambda}_j = \left\{ \sum_{i=1}^N E \left(b_i b_i^T | Y_{obs}, \hat{\theta} \right) \right\}^{-1} \sum_{i=1}^N \left\{ E \left(x_{i,j} b_i | Y_{obs}, \hat{\theta} \right) - W_{i,j}^T \hat{\xi}_j E \left(b_i | Y_{obs}, \hat{\theta} \right) \right\}, j=1, 2
\end{aligned}$$

where $\hat{\theta} = \{\hat{\alpha}, \hat{\xi}_1, \hat{\xi}_2, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\sigma}^2, \hat{s}_1^2, \hat{s}_2^2, \hat{D}\}$ is the current parameter estimate and $Y_{obs} = \{\mathbf{y}_i, \tau_i = \min(x_{i,1}, x_{i,2}, c_i), \delta_i, i = 1, \dots, N\}$.

So in the E-step, we need to calculate these quantities: $E(b_i | Y_{obs}, \hat{\theta})$, $E(b_i b_i^T | Y_{obs}, \hat{\theta})$, $E(\varepsilon_i^T \varepsilon_i | Y_{obs}, \hat{\theta})$, $E(r_{i,j}^2 | Y_{obs}, \hat{\theta})$, $j = 1, 2$, $E(x_{i,j} | Y_{obs}, \hat{\theta})$, $E(x_{i,j} b_i | Y_{obs}, \hat{\theta})$, $j = 1, 2$, which can be done analytically using the properties of multivariate normal distribution. Because of competing risk issue, at most one of the two survival time, $x_{i,1}$ and $x_{i,2}$ is observed. The calculation of these quantities requires the estimation of the first two moments of the joint distributions of $x_{i,1}$ and $x_{i,2}$. The appendix describes details of these calculations. Initial values of the parameters used in the EM algorithm are estimated from the models that are fit on each outcome separately. The analytical formula for the variance calculation is very complicated [3]. Instead, we propose to use the non-parametric bootstrap [7] for the variance estimation.

3 Simulation Study

We implemented a simulation to investigate the finite sample properties of the model. We considered two covariates in the model: one continuous variable (age) and one binary variable (sex). The data were simulated by first simulating the two covariates. Age followed a normal distribution $N(60, 5)$ and race followed a Bernoulli distribution with 0.5 probability of being male. Each individual had a maximum of 12 eGFR measures. At each follow-up time $t = 1, \dots, 12$, eGFR was simulated as follows:

$$eGFR_{it} = \alpha_1 + \alpha_2 * age_i + \alpha_3 * sex_i + \alpha_4 * t + \alpha_5 * age_i * t + \alpha_6 * sex_i * t + b_{1,i} + b_{2,i} * t + \varepsilon_{i,t}$$

where $a = [100, -2, -4, -6, -0.5, -1]$, b_1 and b_2 followed a bivariate normal distribution

with mean of $(0, 0)^T$ and variance of $\begin{pmatrix} 100 & 5 \\ 5 & 4 \end{pmatrix}$ and $\varepsilon_{it} \sim N(0, 10)$. The next step was to simulate the time to ESRD and death respectively. The time to ESRD followed the model:

$$time_{i,ESRD} = \exp(\xi_{11} + \xi_{12} * age_i + \xi_{13} * sex_i + \lambda_{11} b_{1,i} + \lambda_{12} b_{2,i} + r_{i,1})$$

in which $\xi_1 = [3.1, -0.2, -0.6]$, $\lambda_1 = [0.1, 0.5]$ and $r_{i,1} \sim N(0, 0.6)$. Death time followed a similar model:

$$time_{i,death} = \exp(\xi_{21} + \xi_{22} * age_i + \xi_{23} * sex_i + \lambda_{21} b_{1,i} + \lambda_{22} b_{2,i} + r_{i,2})$$

in which $\xi_2 = [2.7, -0.15, -0.2]$, $\lambda_2 = [0.3, 0.3]$ and $r_{i,2} \sim N(0, 0.7)$. The observed time τ_i was taken as the minimum of $time_{i,ESRD}$, $time_{i,death}$ and c_i where c_i followed a uniform distribution with a minimum of 3 and a maximum of 12. Any $eGFR_{it}$ value that was after time τ_i was discarded. The sample size of each simulation was 1,000 and we did 100 iterations.

We fit the proposed joint models and compared the performance with a two-step approach as was mentioned in [27] and is often used in clinical literature. In the two-step approach, a random intercept and slope model was fit on the eGFR data first to have an estimate of individual specific intercept and slope. In the second step, a censored regression model was fit on the log transformed time to ESRD and death separately. The subject specific intercept and slope were included as covariates in both survival models in addition to age and sex.

The two methods were compared with respect to bias, variance and mean square error (MSE). Table 1 summarizes the difference in bias and variance estimates between the two methods. All parameters estimated using the joint modeling approach were unbiased. In the two-step approach, large bias was observed in the coefficients for the random intercept and slope terms, e.g., λ_{11} and λ_{21} in both survival models. There was also large bias in the nuisance parameters including the residual variance terms in both event time models and the variance covariance matrix of the random terms (data not shown). Figure 1 shows the comparison of MSE estimation. The joint modeling approach generally had smaller MSE than the two-step approach.

4 Analysis of the CRIC Data

The CRIC study is a multi-center prospective cohort study of individuals with chronic kidney disease. One objective of the study is to identify risk factors for renal disease progression. The study enrolled participants who had chronic kidney disease and follows them through annual clinic visit during which the blood and urine samples are collected. GFR is estimated using an estimating equation that includes the serum creatinine and

cystatin C concentration measured from the blood samples [1]. In addition to the annually estimated GFR, the study also records the time that a patient develops end-stage renal disease (ESRD), as indicated by the onset of chronic dialysis therapy or kidney transplantation. Date of death is also recorded. After a patient develops ESRD, the estimated GFR is no longer relevant as it is no longer a truly reflection of kidney function. We limited our analysis to the subgroup of patients who had diabetes at study entry. Table 2 shows the baseline characteristics of the study population in this analysis. The mean age of the study participants at baseline was 59.5 years. 56% were male and 44% were of black race. Other covariates of interest include high-sensitivity troponin T (TnT), N-terminal pro-B-type natriuretic peptide (BNP), systolic blood pressure (SBP) and the protein-to-creatinine ratio (PCR), all measured at baseline. The average follow-up time was 5.6 years. During the follow-up, the study observed 573 ESRD events (event rate: 57 per 1000 person years follow-up) and 314 death events (event rate: 31 per 1000 person years follow-up).

We fit a random intercept and slope model for the repeated eGFR measures. For each individual i , we have

$$\begin{aligned} eGFR_i &= \alpha_0 + \alpha_1 * age_i + \alpha_2 * female_i + \alpha_3 * black_i + \alpha_4 * hispanic_i \\ &\quad + \alpha_5 * BNP_i + \alpha_6 * TnT_i + \alpha_7 * SBP_i + \alpha_8 * PCR_i + b_{1,i} \\ &\quad + time * (\alpha_9 + \alpha_{10} * age_i + \alpha_{11} * female_i + \alpha_{12} * black_i + \alpha_{13} * hispanic_i \\ &\quad + \alpha_{14} * BNP_i + \alpha_{15} * TnT_i + \alpha_{16} * SBP_i + \alpha_{17} * PCR_i + b_{2,i}) \\ &\quad + \varepsilon_i \end{aligned}$$

where $b_{1,i}$ is the individual specific intercept term and $b_{2,i}$ is the individual specific slope term. Due to the skewness, we used natural logarithm transformed values of BNP, TnT and PCR in the model.

Death is a competing risk for ESRD, i.e., death prevents the occurrence of ESRD. But ESRD is not a competing event for death. The risk factor profile for death among ESRD patients is quite different from those who do not develop ESRD. Since we were interested in the risk factors for death before developing ESRD, we excluded the death event that occurred after an ESRD event and treated ESRD as a competing risk for death in this analysis. Linear regression models were fit on the natural logarithm transformation of the time to ESRD and death separately. The models are

$$\begin{aligned} \log(time_{ESRD,i}) &= \xi_{10} + \xi_{11} * age_i + \xi_{12} * female_i + \xi_{13} * black_i + \xi_{14} * hispanic_i \\ &\quad + \xi_{15} * BNP_i + \xi_{16} * TnT_i + \xi_{17} * SBP_i + \xi_{18} * PCR_i \\ &\quad + \lambda_{11} * b_{1,i} + \lambda_{12} * b_{2,i} + r_{1,i} \\ \log(time_{death,i}) &= \xi_{20} + \xi_{21} * age_i + \xi_{22} * female_i + \xi_{23} * black_i + \xi_{24} * hispanic_i \\ &\quad + \xi_{25} * BNP_i + \xi_{26} * TnT_i + \xi_{27} * SBP_i + \xi_{28} * PCR_i \\ &\quad + \lambda_{21} * b_{1,i} + \lambda_{22} * b_{2,i} + r_{2,i} \end{aligned}$$

The three models share the common parameters $b_{1,i}$ and $b_{2,i}$. Standard errors of the parameters were estimated using a non-parametric bootstrap method. The results from the modeling of repeated eGFR are summarized in Table 3. In the model for eGFR, the

coefficients for the intercept can be interpreted as the association between the covariates and the baseline eGFR value. The coefficients for the slope term can be interpreted as the association between the covariates and the change in eGFR over time, which is the relationship of primary interest. Age and sex were not associated with eGFR decline over time. The black race and Hispanic ethnicity groups had faster rates of eGFR decline, indicated by the negative parameter estimates for the slope terms. Higher values of baseline BNP, SBP and PCR were all associated with faster rates of eGFR decline. PCR had the strongest association with eGFR slope. A unit increase in the log transformed PCR value was associated with 2.76 units smaller GFR slope. The covariance between the random intercept and slope terms was positive, suggesting that higher baseline eGFR was associated with slower eGFR decline over time, i.e., less negative or larger positive slope estimates.

Table 4 summarizes the results from the models for ESRD and death. Female sex, black race, Hispanic ethnicity and larger values of BNP, TnT, SBP and PCR were all associated with shorter time to ESRD, i.e., increased risk of ESRD. Higher values of random intercept and slope terms were associated with longer time to ESRD, i.e., decreased risk of ESRD. One unit increase in eGFR slope (i.e., less steep eGFR slope) was associated with 33% longer time to ESRD. In the model for death, older age and higher values of BNP and TnT were associated with shorter time to death. The random intercept and slope terms were not predictive of death time.

In a sensitivity analysis, we modeled the log transformed eGFR (data were not shown) and the interpretation of the results were similar. However, the residual variance for the ESRD outcome is substantially larger, suggesting a model with eGFR in the original scale fit the data better. We also did the model checking by visually examining the residuals. Q-Q plots of the residuals and the two random terms were shown in Figure 2, suggesting no violation of the normal distribution assumption. For censored survival outcomes, we imputed the event time based on the assumed model and calculated the residuals as the difference between the imputed and predicted values [12,19]. We also plotted the residuals versus the predicted values in all three models in Figure 3, which suggested the models fit the data well.

5 Discussion

We proposed a joint model for a longitudinal outcome and competing survival outcomes. We fit a mixed effects model for the repeated longitudinal measures and a regression model for each survival outcome. The correlation among the different outcomes was accounted for by having the random effects terms of the mixed effects model in the model for survival outcomes. Many of the limitations in modeling the outcomes separately arise from the correlations between them. The joint modeling method tackles this problem by explicitly modeling these correlations. The joint model also allows simultaneous estimation of the association of baseline covariates and different outcomes. As demonstrated in the simulation, a joint modeling method provides unbiased parameter estimates and has smaller MSE compared to the a two-step modeling approach.

The model parameters were estimated via the EM algorithm. Using the properties of multivariate normality, closed form solution can be derived in the M-step within each iteration and the computation is relatively straightforward compared to other joint modeling methods in which Cox proportional hazards models are used for the survival outcomes. Compared to the point estimate, it is fairly complicated to derive the analytic formula for the variance estimation. Therefore, a non-parametric bootstrap method was used for the variance estimation in this work.

One key assumption of the model is the independence of residuals from the different models. There are two important implications in the context of the CRIC example. First, the assumption assumes that the deviation of eGFR from the assumed underlying trajectory is uncorrelated with either the ESRD or death time. This assumption may not hold in our example as physicians often rely on the observed eGFR instead of the true underlying value to determine whether a patient should go to dialysis or have a kidney transplantation. It may become more plausible if we model the iothalamate GFR, a direct measure of kidney function which is often unavailable to physicians, as was done in [27,37]. Second, it also assumes that the times to ESRD and death are independent by conditioning on the underlying eGFR trajectory and baseline covariates. This assumption may also not be plausible because those who developed ESRD had an increased risk of death and those who had died would have an increased risk of ESRD if they were alive, unless the only cause of ESRD and death other than baseline covariates is the underlying GFR trajectory. However, this assumption is important for model identification. It has been shown that the observed data contains no information about the dependence structure of the competing event times [33] and so the model is not identified without this assumption. One possible way to make the assumption more plausible is to include all the time-varying predictors of ESRD and death in the model. So that we can assume the two survival times are independent conditioning on all time-varying predictors. Another possible solution is to make a composite of the two survival events so that the survival outcome is time to either ESRD or death. In this case, the model reduces to a standard shared parameter model in which there is only a single survival outcome. The CRIC data example is actually an example of semi-competing risks. Death is a competing event for ESRD but ESRD does not prevent the occurrence of death. Since we were interested in the risk factor profile for death without ESRD, we ignored the death information after ESRD. The additional death information after ESRD may potentially help for model identification which deserves future research.

In studies of chronic kidney disease, analyses of ESRD as the outcome are often under-powered because of the low incidence of ESRD. There is a desire to use eGFR slope as a surrogate for ESRD [16,26,29] in various settings. The analysis in this work provides evidence that eGFR slope may not be a good surrogate for ESRD. For example, in the mixed effects model for the repeated eGFR outcome, we observed a strong association between SBP and eGFR slope. In the model for ESRD, eGFR slope was also highly predictive of ESRD. If eGFR is a good surrogate for ESRD, the association between SBP and ESRD should be largely attenuated after accounting for eGFR slope in the model. However, there was still a significant association between SBP and ESRD, suggesting a link between SBP and ESRD through a pathway not involving eGFR slope. Another example is that TnT was not associated with eGFR slope but had a strong association with ESRD. One reviewer

suggested this may also be explained by the model mis-specification. For example, assuming a scenario in which the association of TnT with eGFR slope (and so with ESRD) is stronger among faster progressors than among slow progressors, the association between TnT and eGFR slope will tend to be underestimated given the greater emphasis on the slow progressors in the longitudinal model. For similar reason, its association with ESRD will be overestimated. One solution is to include an interaction term for TnT with random eGFR slope as the additional shared parameter to take into account the heterogeneous association.

One limitation of the CRIC data is related to the clinical definition of ESRD. In CKD, the initiation of dialysis or kidney transplantation is often triggered when the eGFR value crosses some threshold. It is reported by the USRDS that the average eGFR value at the time of initiation of hemodialysis therapy is 10.9 mL/min/1.73 m² in the US population within the same age group as CRIC participants. It is possible the correlation between the time to ESRD and the eGFR trajectories is not fully captured by the shared parameters, e.g., random intercept and slope terms, in our model.

The validity of our proposed method builds on the multivariate normality assumption. However, it is possible that some of the outcomes may not follow the normal distribution. Some types of transformations may normalize these distributions. For example, we implemented a natural logarithm transformation of the two survival times in the CRIC data analysis. We also did the log transformation on the eGFR values and the results were similar. The plots of residuals suggested the normality assumption was satisfied in this particular example. However, it is still not clear how sensitive the model is with respect to the violations of the normality assumption, which should be a topic of future research.

Acknowledgments

Wei Yang, Dawei Xie, Qiang Pan and Harold I. Feldman were supported by U01DK060990. Wensheng Guo was supported by R01GM104470. The authors thank the CRIC Study patients, study coordinators, and investigators for their efforts. Other grant Support: By the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health (cooperative agreements U01DK060984, U01DK061022, U01DK061021, U01DK061028, U01DK060980, U01DK060963, and U01DK060902) and in part by the following institutional Clinical Translational Science Awards from the National Center for Advancing Translational Sciences and other National Institutes of Health grants: University of Pennsylvania (UL1TR000003, K01DK092353, and K24DK002651), Johns Hopkins University (UL1TR-000424), University of Maryland General Clinical Research Center (M01RR-16500), Clinical and Translational Science Collaborative of Cleveland (UL1TR000439), Michigan Institute for Clinical and Health Research (UL1TR000433), University of Illinois at Chicago (UL1RR029879), Tulane University Translational Research in Hypertension and Renal Biology (P30GM103337), and Kaiser Permanente National Institutes of Health/National Center for Research Resources University of California, San Francisco Clinical and Translational Science Institute (UL1RR-024131).

Appendix

A Covariance matrix for the joint distribution

According to models (1) and (2), we have

$$Cov \begin{pmatrix} b \\ \varepsilon \\ r_1 \\ r_2 \\ y \\ x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} D & 0 & 0 & 0 & DZ^T & D\lambda_1 & D\lambda_2 \\ 0 & \sigma^2 I & 0 & 0 & \sigma^2 I & 0 & 0 \\ 0 & 0 & s_1^2 & 0 & 0 & s_1^2 & 0 \\ 0 & 0 & 0 & s_2^2 & 0 & 0 & s_2^2 \\ ZD & \sigma^2 I & 0 & 0 & ZDZ^T + \sigma^2 I & ZD\lambda_1 & ZD\lambda_2 \\ \lambda_1^T D & 0 & s_1^2 & 0 & \lambda_1^T DZ^T & \lambda_1^T D\lambda_1 + s_1^2 & \lambda_1^T D\lambda_2 \\ \lambda_2^T D & 0 & 0 & s_2^2 & \lambda_2^T DZ^T & \lambda_2^T D\lambda_1 & \lambda_2^T D\lambda_2 + s_2^2 \end{pmatrix}$$

We suppress the subscript i for clarity.

B Expectation of the sufficient statistics

We first derive the expectations assuming both x_1 and x_2 are observed. We will then revise the formula accounting for the scenario when either or both x_1 or x_2 are censored. We use Y_{comp} to denote the full data including $\mathbf{y}$, x_1 and x_2 .

According to models (1) and (2) and the covariance structure in the appendix, we have

$$E(b|Y_{comp}, \hat{\theta}) = C_{12}C_{22}^{-1} \begin{pmatrix} y - T\hat{\alpha} \\ x_1 - W^T\hat{\xi}_1 \\ x_2 - W^T\hat{\xi}_2 \end{pmatrix}$$

$$Var(b|Y_{comp}, \hat{\theta}) = \hat{D} - C_{12}C_{22}^{-1}C_{12}^T$$

where

$$C_{12} = (\hat{D}Z^T \hat{D}\hat{\lambda}_1 \hat{D}\hat{\lambda}_2)$$

and

$$C_{22} = \begin{pmatrix} Z\hat{D}Z^T + \hat{\sigma}^2 I & Z\hat{D}\hat{\lambda}_1 & Z\hat{D}\hat{\lambda}_2 \\ \hat{\lambda}_1^T \hat{D}Z^T & \hat{\lambda}_1^T \hat{D}\hat{\lambda}_1 + \hat{s}_1^2 & \hat{\lambda}_1^T \hat{D}\hat{\lambda}_2 \\ \hat{\lambda}_2^T \hat{D}Z^T & \hat{\lambda}_2^T \hat{D}\hat{\lambda}_1 & \hat{\lambda}_2^T \hat{D}\hat{\lambda}_2 + \hat{s}_2^2 \end{pmatrix}$$

$E(bb^T|Y_{comp}, \hat{\theta})$ can then be calculated as

$$E(bb^T|Y_{comp}, \hat{\theta}) = E(b|Y_{comp}, \hat{\theta}) E^T(b|Y_{comp}, \hat{\theta}) + Var(b|Y_{comp}, \hat{\theta})$$

In the case when either or both x_1 and x_2 are censored, $E(b|Y_{obs}, \hat{\theta})$ and $E(bb^T|Y_{obs}, \hat{\theta})$ can be calculated as follows. We have

$$E(b|Y_{obs}, \hat{\theta}) = E\{E(b|Y_{comp}, \hat{\theta})|Y_{obs}, \hat{\theta}\} = E(\hat{b}|Y_{obs}, \hat{\theta})$$

and

$$\begin{aligned}
 & E(bb^T|Y_{obs}, \hat{\theta}) \\
 &= E\left\{E\left(bb^T|Y_{comp}, \hat{\theta}\right)|Y_{obs}, \hat{\theta}\right\} \\
 &= E\left\{E\left(b|Y_{comp}, \hat{\theta}\right)E^T\left(b|Y_{comp}, \hat{\theta}\right)+Var\left(b|Y_{comp}, \hat{\theta}\right)|Y_{obs}, \hat{\theta}\right\} \\
 &= E\left(\hat{b}\hat{b}^T|Y_{obs}, \hat{\theta}\right)+E\left\{Var\left(b|Y_{comp}, \hat{\theta}\right)|Y_{obs}, \hat{\theta}\right\} \\
 &= E\left(\hat{b}\hat{b}^T|Y_{obs}, \hat{\theta}\right)+Var\left(b|Y_{comp}, \hat{\theta}\right)
 \end{aligned}$$

where $\hat{b} = E(b|Y_{comp}, \hat{\theta})$. The last step follows because $Var(b|Y_{comp}, \hat{\theta})$ is not a function of Y_{comp} .

So the next step is to calculate $E(\hat{b}\hat{b}^T|Y_{obs}, \hat{\theta})$. Since

$$\hat{b}\hat{b}^T = C_{12}C_{22}^{-1} \begin{pmatrix} y - T\hat{\alpha} \\ x_1 - W^T\hat{\xi}_1 \\ x_2 - W^T\hat{\xi}_2 \end{pmatrix} \begin{pmatrix} y - T\hat{\alpha} \\ x_1 - W^T\hat{\xi}_1 \\ x_2 - W^T\hat{\xi}_2 \end{pmatrix}^T C_{22}^{-1}C_{21}$$

we need to calculate $E(x_1|Y_{obs}, \hat{\theta})$, $E(x_2|Y_{obs}, \hat{\theta})$, $E(x_1^2|Y_{obs}, \hat{\theta})$, $E(x_2^2|Y_{obs}, \hat{\theta})$ and $E(x_1x_2|Y_{obs}, \hat{\theta})$. See next section for the details of how to calculate these expectations under different scenarios. The calculation of $E(\varepsilon\varepsilon^T|Y_{obs}, \hat{\theta})$, $E(r_j^2|Y_{obs}, \hat{\theta})$, $j = 1, 2$ can be done in a similar fashion.

To calculate $E(x_jb|Y_{obs}, \hat{\theta})$, $j = 1, 2$, we have

$$\begin{aligned}
 & E(x_jb|Y_{obs}, \hat{\theta}) \\
 &= E\left\{E(x_jb|Y_{comp}, \hat{\theta})|Y_{obs}, \hat{\theta}\right\} \\
 &= E\left\{x_jC_{12}C_{22}^{-1} \begin{pmatrix} y - T\hat{\alpha} \\ x_1 - W^T\hat{\xi}_1 \\ x_2 - W^T\hat{\xi}_2 \end{pmatrix} |Y_{obs}, \hat{\theta}\right\}
 \end{aligned}$$

Again see next section for the calculation of $E(x_1^2|Y_{obs}, \hat{\theta})$, $E(x_2^2|Y_{obs}, \hat{\theta})$ and $E(x_1x_2|Y_{obs}, \hat{\theta})$.

To summarize, here are the formulas to calculate these quantities.

$$\begin{aligned}
 & -E(b_i|Y_{obs}, \hat{\theta}) = C_{12}C_{22}^{-1}E(O_i) \\
 & -E(b_ib_i^T|Y_{obs}, \hat{\theta}) = C_{12}C_{22}^{-1}E(O_iO_i^T)C_{22}^{-1}C_{12}^T + \hat{D} - C_{12}C_{22}^{-1}C_{12}^T \\
 & -E(\varepsilon_i\varepsilon_i^T|Y_{obs}, \hat{\theta}) = F_{12}C_{22}^{-1}E(O_iO_i^T)C_{22}^{-1}F_{12}^T + \hat{\sigma}^2I - F_{12}C_{22}^{-1}F_{12}^T \\
 & -E(r_{i,1}^2|Y_{obs}, \hat{\theta}) = G_{12}C_{22}^{-1}E(O_iO_i^T)C_{22}^{-1}G_{12}^T + \hat{s}_1^2 - G_{12}C_{22}^{-1}G_{12}^T \\
 & -E(r_{i,2}^2|Y_{obs}, \hat{\theta}) = H_{12}C_{22}^{-1}E(O_iO_i^T)C_{22}^{-1}H_{12}^T + \hat{s}_2^2 - H_{12}C_{22}^{-1}H_{12}^T \\
 & -E(x_{i,1}b_i|Y_{obs}, \hat{\theta}) = C_{12}C_{22}^{-1}E(x_{i,1}O_i) \\
 & -E(x_{i,2}b_i|Y_{obs}, \hat{\theta}) = C_{12}C_{22}^{-1}E(x_{i,2}O_i)
 \end{aligned}$$

where

$$\begin{aligned}
 -C_{12} &= \begin{pmatrix} \hat{D}Z^T & \hat{D}\hat{\lambda}_1 & \hat{D}\hat{\lambda}_2 \end{pmatrix} \\
 -F_{12} &= \begin{pmatrix} \hat{\sigma}^2 I & \mathbf{0} & \mathbf{0} \end{pmatrix} \\
 -G_{12} &= \begin{pmatrix} \mathbf{0} & \hat{s}_1^2 & 0 \end{pmatrix} \\
 -H_{12} &= \begin{pmatrix} \mathbf{0} & 0 & \hat{s}_2^2 \end{pmatrix} \\
 -C_{22} &= \begin{pmatrix} Z\hat{D}Z^T + \hat{\sigma}^2 I & Z\hat{D}\hat{\lambda}_1 & Z\hat{D}\hat{\lambda}_2 \\ \hat{\lambda}_1^T \hat{D}Z^T & \hat{\lambda}_1^T \hat{D}\hat{\lambda}_1 + \hat{s}_1^2 & \hat{\lambda}_1^T \hat{D}\hat{\lambda}_2 \\ \hat{\lambda}_2^T \hat{D}Z^T & \hat{\lambda}_2^T \hat{D}\hat{\lambda}_1 & \hat{\lambda}_2^T \hat{D}\hat{\lambda}_2 + \hat{s}_2^2 \end{pmatrix} \\
 -O_i &= \begin{pmatrix} y_i - T\hat{\alpha} \\ x_{i,1} - W^T \hat{\xi}_1 \\ x_{i,2} - W^T \hat{\xi}_2 \end{pmatrix}
 \end{aligned}$$

Note that $E(\varepsilon_i^T \varepsilon_i | Y_{obs}, \hat{\theta})$ can be calculated by summing over the diagonal elements of $E(\varepsilon_i \varepsilon_i^T | Y_{obs}, \hat{\theta})$, i.e., $E(\varepsilon_i^T \varepsilon_i | Y_{obs}, \hat{\theta}) = \text{trace} \left\{ E(\varepsilon_i \varepsilon_i^T | Y_{obs}, \hat{\theta}) \right\}$.

C Calculation of the first two moments of $\mathbf{x}$ given Y_{obs}

There are three types of individuals in the observed data: 1) those whose x_1 are observed but not x_2 , i.e., $x_1 < c$, $x_2 > c$; 2) those whose x_2 are observed but not x_1 , i.e., $x_1 > c$, $x_2 < c$; and 3) neither x_1 nor x_2 is observed, i.e., $x_1 > c$, $x_2 > c$. We now derive $E(x_1 | Y_{obs}, \hat{\theta})$, $E(x_2 | Y_{obs}, \hat{\theta})$, $E(x_1^2 | Y_{obs}, \hat{\theta})$, $E(x_2^2 | Y_{obs}, \hat{\theta})$ and $E(x_1 x_2 | Y_{obs}, \hat{\theta})$ for each of the three scenarios.

C.1 Scenario I: x_2 alone is censored

For the first scenario in which x_1 is observed and x_2 is censored, we have $Y_{obs} = (\mathbf{y}, x_1, x_2 > c)$ and

$$\begin{aligned}
 -E(x_1 | Y_{obs}, \hat{\theta}) &= x_1 \\
 -E(x_1^2 | Y_{obs}, \hat{\theta}) &= x_1^2 \\
 -E(x_2 | Y_{obs}, \hat{\theta}) &= E(x_2 | \mathbf{y}, x_1, x_2 > c, \hat{\theta}) \\
 -E(x_1 x_2 | Y_{obs}, \hat{\theta}) &= x_1 E(x_2 | \mathbf{y}, x_1, x_2 > c, \hat{\theta}) \\
 -E(x_2^2 | Y_{obs}, \hat{\theta}) &= E^2(x_2 | \mathbf{y}, x_1, x_2 > c, \hat{\theta}) + \text{Var}(x_2 | \mathbf{y}, x_1, x_2 > c, \hat{\theta})
 \end{aligned}$$

Because

$$E \begin{pmatrix} x_2 \\ x_1 \\ y \end{pmatrix} = \begin{pmatrix} W^T \xi_2 \\ W^T \xi_1 \\ T\alpha \end{pmatrix}$$

and

$$Cov \begin{pmatrix} x_2 \\ x_1 \\ y \end{pmatrix} = \begin{pmatrix} \lambda_2^T D \lambda_2 + s_2^2 & \lambda_2^T D \lambda_1 & \lambda_2^T D Z^T \\ \lambda_1^T D \lambda_2 & \lambda_1^T D \lambda_1 + s_1^2 & \lambda_1^T D Z^T \\ Z D \lambda_2 & Z D \lambda_1 & Z D Z^T + \sigma^2 I \end{pmatrix}$$

we have

$$E(x_2|x_1, \mathbf{y}, \hat{\theta}) = W^T \xi_2 + K_{12} K_{22}^{-1} \begin{pmatrix} x_1 - W^T \hat{\xi}_1 \\ y - T \hat{\alpha} \end{pmatrix}$$

and

$$Var(x_2|x_1, y, \hat{\theta}) = \hat{\lambda}_2^T D \hat{\lambda}_2 + \hat{s}_2^2 - K_{12} K_{22}^{-1} K_{12}^T$$

where $K_{12} = (\hat{\lambda}_2^T D \hat{\lambda}_1, \hat{\lambda}_2^T D Z^T)$ and $K_{22} = \begin{pmatrix} \hat{\lambda}_1^T \hat{D} \hat{\lambda}_1 + \hat{s}_1^2 & \hat{\lambda}_1^T \hat{D} Z^T \\ Z \hat{D} \hat{\lambda}_1 & Z \hat{D} Z^T + \hat{\sigma}^2 I \end{pmatrix}$.

So the calculation of $E(x_2|\mathbf{y}, x_1, x_2 > c, \hat{\theta})$ and $Var(x_2|\mathbf{y}, x_1, x_2 > c, \hat{\theta})$ can be done based on the truncated normal distribution property. In R, this can be done easily using the *truncnorm* package [31].

C.2 Scenario II: x_1 alone is censored

Similarly, for the second scenario in which x_2 is observed and x_1 is censored, we have $Y_{obs} = (\mathbf{y}, x_1 > c, x_2)$ and

$$\begin{aligned} & - E(x_2|Y_{obs}, \hat{\theta}) = x_2 \\ & - E(x_2^2|Y_{obs}, \hat{\theta}) = x_2^2 \\ & - E(x_1|Y_{obs}, \hat{\theta}) = E(x_1|\mathbf{y}, x_1 > c, x_2, \hat{\theta}) \\ & - E(x_1 x_2|Y_{obs}, \hat{\theta}) = x_2 E(x_1|\mathbf{y}, x_1 > c, x_2, \hat{\theta}) \\ & - E(x_1^2|Y_{obs}, \hat{\theta}) = E^2(x_1|\mathbf{y}, x_1 > c, x_2, \hat{\theta}) + Var(x_1|\mathbf{y}, x_1 > c, x_2, \hat{\theta}) \end{aligned}$$

To calculate $E(x_1|\mathbf{y}, x_2, x_1 < c, \hat{\theta})$ and $Var(x_1|\mathbf{y}, x_2, x_1 < c, \hat{\theta})$, we have

$$E(x_1|x_2, \mathbf{y}, \hat{\theta}) = W^T \hat{\xi}_1 + P_{12} P_{22}^{-1} \begin{pmatrix} x_2 - W^T \hat{\xi}_2 \\ y - T \hat{\alpha} \end{pmatrix}$$

and

$$Var(x_1|x_2, \mathbf{y}, \hat{\theta}) = \hat{\lambda}_1^T D \hat{\lambda}_1 + \hat{s}_1^2 - P_{12} P_{22}^{-1} P_{12}^T$$

where $P_{12} = (\hat{\lambda}_1^T D \hat{\lambda}_2, \hat{\lambda}_1^T D Z^T)$ and $P_{22} = \begin{pmatrix} \hat{\lambda}_2^T \hat{D} \hat{\lambda}_2 + \hat{s}_2^2 & \hat{\lambda}_2^T \hat{D} Z^T \\ Z \hat{D} \hat{\lambda}_2 & Z \hat{D} Z^T + \hat{\sigma}^2 I \end{pmatrix}$.

C.3 Scenario III: both x_1 and x_2 are censored

When both x_1 and x_2 are censored, i.e, $x_1 > c$ and $x_2 > c$,

$$\begin{aligned}
 & - E(x_1|Y_{obs}, \hat{\theta}) = E(x_1|\mathbf{y}, x_1 > c, x_2 > c, \hat{\theta}) \\
 & - E(x_1^2|Y_{obs}, \hat{\theta}) = E^2(x_1|\mathbf{y}, x_1 > c, x_2 > c, \hat{\theta}) + Var(x_1|\mathbf{y}, x_1 > c, x_2 > c, \hat{\theta}) \\
 & - E(x_2|Y_{obs}, \hat{\theta}) = E(x_2|\mathbf{y}, x_1 > c, x_2 > c, \hat{\theta}) \\
 & - E(x_2^2|Y_{obs}, \hat{\theta}) = E^2(x_2|\mathbf{y}, x_1 > c, x_2 > c, \hat{\theta}) + Var(x_2|\mathbf{y}, x_1 > c, x_2 > c, \hat{\theta}) \\
 & - E(x_1 x_2|Y_{obs}, \hat{\theta}) = Cov(x_1 x_2|\mathbf{y}, x_1 > c, x_2 > c, \hat{\theta}) + E(x_1|\mathbf{y}, x_1 > c, x_2 > c, \hat{\theta}) * E(x_2|\mathbf{y}, x_1 > c, x_2 > c, \hat{\theta})
 \end{aligned}$$

We have

$$E \begin{pmatrix} x_1 \\ | \mathbf{y}, \hat{\theta} \\ x_2 \end{pmatrix} = \begin{pmatrix} W^T \hat{\xi}_1 \\ W^T \hat{\xi}_2 \end{pmatrix} + Q_{12} Q_{22}^{-1} (y - T \hat{\alpha})$$

and

$$Var \begin{pmatrix} x_1 \\ | y, \hat{\theta} \\ x_2 \end{pmatrix} = \begin{pmatrix} \hat{\lambda}_1^T \hat{D} \hat{\lambda}_1 + \hat{s}_1^2 & \hat{\lambda}_1^T \hat{D} \hat{\lambda}_2 \\ \hat{\lambda}_2^T \hat{D} \hat{\lambda}_1 & \hat{\lambda}_2^T \hat{D} \hat{\lambda}_2 + \hat{s}_2^2 \end{pmatrix} - Q_{12} Q_{22}^{-1} Q_{12}^T$$

where $Q_{12} = (\hat{\lambda}_1^T D Z^T, \hat{\lambda}_2^T D Z^T)^T$ and $Q_{22} = Z \hat{D} Z^T + \hat{\sigma}^2 I$. R package *tmvtnorm* [39] can then be used to calculate the mean and covariance of the truncated variables.

References

1. Anderson AH, Yang W, Hsu C-Y, Joffe MM, Leonard MB, Xie D, Chen J, Greene T, Jaar BG, Kao P, et al. Estimating gfr among participants in the chronic renal insufficiency cohort (cric) study. *American Journal of Kidney Diseases*. 2012; 60(2):250–261. [PubMed: 22658574]
2. Andrinopoulou E-R, Rizopoulos D, Takkenberg JJM, Lesaffre E. Joint modeling of two longitudinal outcomes and competing risk data. *Statistics in Medicine*. Aug; 2014 33(18):3167–3178. [PubMed: 24676841]
3. DeGruttola V, Tu XM. Modelling Progression of CD4-Lymphocyte Count and Its Relationship to Survival Time. *Biometrics*. Dec; 1994 50(4):1003–1014. [PubMed: 7786983]
4. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society Series B (methodological)*. 1977:1–38.
5. Deslandes E, Chevret S. Joint modeling of multivariate longitudinal data and the dropout process in a competing risk setting: application to ICU data. *BMC Medical Research Methodology*. Jul.2010 10(1):69. [PubMed: 20670425]
6. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society Series C (Applied Statistics)*. 1994; 43(1):49–93.
7. Efron, B., Tibshirani, RJ. An introduction to the bootstrap. CRC press; 1994.
8. Elashoff RM, Li G, Li N. An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in Medicine*. 2007; 26(14):2813–2835. [PubMed: 17124698]
9. Feldman HI, Appel LJ, Chertow GM, Cifelli D, Cizman B, Daugirdas J, Fink JC, Franklin-Becker ED, Go AS, Hamm LL, et al. The chronic renal insufficiency cohort (cric) study: Design and

- methods. *Journal of the American Society of Nephrology*. 2003; 14(suppl 2):S148–S153. [PubMed: 12819321]
10. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*. 1999; 94(446):496–509.
 11. Guo X, Carlin BP. Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages. *The American Statistician*. Feb; 2004 58(1):16–24.
 12. Hillis SL. Residual plots for the censored data linear regression model. *Statistics in medicine*. 1995; 14(18):2023–2036. [PubMed: 8677402]
 13. Hogan JW, Roy J, Korkontzelou C. Handling drop-out in longitudinal studies. *Statistics in medicine*. 2004; 23(9):1455–1497. [PubMed: 15116353]
 14. Hsieh F, Tseng YK, Wang JL. Joint Modeling of Survival and Longitudinal Data: Likelihood Approach Revisited. *Biometrics*. 2006; 62(4):1037–1043. [PubMed: 17156277]
 15. Huang X, Li G, Elashoff RM, Pan J. A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime Data Analysis*. Jan. 2011 17:80–100. [PubMed: 20549344]
 16. J C, T T, K M, et al. Decline in estimated glomerular filtration rate and subsequent risk of end-stage renal disease and mortality. *JAMA*. 2014; 311(24):2518–2531. [PubMed: 24892770]
 17. Klahr S, Levey AS, Beck GJ, Caggiula AW, Hunsicker L, Kusek JW, Striker G. The effects of dietary protein restriction and blood-pressure control on the progression of chronic renal disease. *New England Journal of Medicine*. 1994; 330(13):877–884. [PubMed: 8114857]
 18. Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? *Statistics in medicine*. 2012; 31(11-12):1089–1097. [PubMed: 21953401]
 19. Law M, Jackson D. Residual plots for linear regression models with censored outcome data: A refined method for visualising residual uncertainty. *Communications in Statistics-Simulation and Computation*, (just-accepted). 2015
 20. Leung K-M, Elashoff RM, Afifi AA. Censoring issues in survival analysis. *Annual review of public health*. 1997; 18(1):83–104.
 21. Li L, Hu B, Greene T. A semiparametric joint model for longitudinal and survival data with application to hemodialysis study. *Biometrics*. 2009; 65(3):737–745. [PubMed: 19173700]
 22. Liu L, Ma JZ, O'Quigley J. Joint analysis of multi-level repeated measures data and survival: an application to the end stage renal disease (esrd) data. *Statistics in Medicine*. 2008; 27(27):5679–5691. [PubMed: 18693300]
 23. Prentice RL, Kalbeisch JD, Peterson AV Jr, Flournoy N, Farewell V, Breslow N. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978:541–554. [PubMed: 373811]
 24. Ratcliffe SJ, Guo W, Ten Have TR. Joint Modeling of Longitudinal and Survival Data via a Common Frailty. *Biometrics*. 2004; 60(4):892–899. [PubMed: 15606409]
 25. Rizopoulos, D. 45th Scientific Meeting of the Italian Statistical Society. Padova: Universitadi Padova; 2010. Joint modelling of longitudinal and time-to-event data: challenges and future directions.
 26. Rosansky SJ, Glasscock RJ. Is a decline in estimated gfr an appropriate surrogate end point for renoprotection drug trials&quest. *Kidney international*. 2014; 85(4):723–727. [PubMed: 24682115]
 27. Schluchter MD. Methods for the analysis of informatively censored longitudinal data. *Statistics in medicine*. 1992; 11(14-15):1861–1870. [PubMed: 1480878]
 28. Song X, Davidian M, Tsiatis AA. A Semiparametric Likelihood Approach to Joint Modeling of Longitudinal and Time-to-Event Data. *Biometrics*. 2002; 58(4):742–753. [PubMed: 12495128]
 29. Stevens LA, Greene T, Levey AS. Surrogate end points for clinical trials of kidney disease progression. *Clinical Journal of the American Society of Nephrology*. 2006; 1(4):874–884. [PubMed: 17699300]
 30. Taylor JMG, Park Y, Ankerst DP, Proust-Lima C, Williams S, Kestin L, Bae K, Pickles T, Sandler H. Real-Time Individual Predictions of Prostate Cancer Recurrence Using Joint Models. *Biometrics*. 2013

31. Trautmann H, Steuer D, Mersmann O, Bornkamp B. truncnorm: Truncated normal distribution. R package version 1.0-7. 2014
32. Tseng Y-K, Hsieh F, Wang J-L. Joint modelling of accelerated failure time and longitudinal data. *Biometrika*. 2005; 92(3):587–603.
33. Tsiatis A. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*. 1975; 72(1):20–22.
34. Tsiatis AA, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*. Jun; 2001 88(2):447–458.
35. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*. 2004; 14(3):809–834.
36. Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS. *Journal of the American Statistical Association*. Mar.1995 90(429):27.
37. Vonesh EF, Greene T, Schluchter MD. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine*. 2006; 25(1):143–163. [PubMed: 16025541]
38. Wang Y, Taylor JMG. Jointly Modeling Longitudinal and Event Time Data with Application to Acquired Immunodeficiency Syndrome. *Journal of the American Statistical Association*. Sep; 2001 96(455):895–905.
39. Wilhelm S, M BG. tmvtnorm: Truncated Multivariate Normal and Student t Distribution. R package version 1.4-10. 2015
40. Williamson PR, Kolamunnage-Dona R, Philipson P, Marson AG. Joint modelling of longitudinal and competing risks data. *Statistics in Medicine*. 2008; 27(30):6426–6438. [PubMed: 18825652]
41. Wright JT Jr, Bakris G, Greene T, Agodoa LY, Appel LJ, Charleston J, Cheek D, Douglas-Baltimore JG, Gassman J, Glassock R, et al. Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease: results from the aask trial. *Jama*. 2002; 288(19):2421–2431. [PubMed: 12435255]
42. Wu L, Liu W, Yi GY, Huang Y. Analysis of Longitudinal and Survival Data: Joint Modeling, Inference Methods, and Issues. *Journal of Probability and Statistics*. 2012; 2012:1–17.
43. Yu M, Law NJ, Taylor JM, Sandler HM. Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*. 2004; 14(3):835–862.

Fig. 1. Boxplots of MSE for the two-step and joint modeling methods. Within each sub-panel, the left plot is from the two-step model and the right plot is from the joint model.

Fig. 2. QQ plots of the residuals from (a) eGFR model ε ; (b) ESRD model r_1 ; (c) death model r_2 ; and (d) the random intercept term b_1 ; and (e) random slope term b_2 .

Fig. 3. Residual plots for the models of (a) repeated eGFR measures; (b) time to ESRD; and (c) time to death.

Table 1

Results from the simulation study.

Param	True	Two-step Method				Joint Modeling			
		Mean	SD	Bias	Bias/SD	Mean	SD	Bias	Bias/SD
λ_{11}	0.1	0.12	0.007	0.02	3.36	0.10	0.007	0.00	0.32
λ_{12}	0.5	0.46	0.034	-0.04	-1.27	0.51	0.025	0.01	0.41
λ_{21}	0.3	0.34	0.008	0.04	4.76	0.30	0.009	0.00	0.52
λ_{22}	0.3	0.24	0.042	-0.06	-1.47	0.29	0.035	-0.01	-0.24
a_1	100.0	100.73	3.388	0.73	0.22	101.00	3.429	1.00	0.29
a_2	-2.0	-2.11	0.563	-0.11	-0.19	-2.11	0.571	-0.11	-0.19
a_3	-4.0	-4.10	0.618	-0.10	-0.16	-4.11	0.629	-0.11	-0.17
a_4	-6.0	-5.59	0.969	0.41	0.42	-5.77	1.033	0.23	0.22
a_5	-0.5	-0.45	0.165	0.05	0.31	-0.52	0.174	-0.02	-0.13
a_6	-1.0	-0.79	0.159	0.21	1.34	-0.99	0.167	0.01	0.03
ξ_{11}	3.1	3.66	0.830	0.56	0.68	3.24	0.748	0.14	0.18
ξ_{12}	-0.2	-0.23	0.137	-0.03	-0.25	-0.21	0.125	-0.01	-0.08
ξ_{13}	-0.6	-0.70	0.134	-0.10	-0.75	-0.61	0.125	-0.01	-0.11
ξ_{21}	2.7	3.10	1.163	0.40	0.35	3.04	1.145	0.34	0.30
ξ_{22}	-0.2	-0.18	0.194	-0.03	-0.14	-0.18	0.192	-0.03	-0.17
ξ_{23}	-0.2	-0.21	0.229	-0.01	-0.05	-0.24	0.218	-0.04	-0.20

Table 2

CRIC study population characteristics at baseline. Mean (standard deviation) for continuous variables and percentage for categorical variables.

Variable	Statistics
Age	59.5 (9.7)
Female	44 %
Black	44 %
Hispanic	17 %
Log BNP	5.4 (1.4)
Log TnT	3.0 (0.9)
SBP	133 (23)
Log PCR	0.6 (0.7)

Table 3

Parameter estimates from the analysis of repeated eGFR data in CRIC.

Parameter	eGFR Intercept			eGFR Slope		
	Estimate	SE	t-stat	Estimate	SE	t-stat
Age	-2.33	0.366	-6.3	0.09	0.078	1.2
Female	-6.52	0.680	-9.6	0.13	0.116	1.1
Black	0.46	0.621	0.7	-0.53	0.144	-3.7
Hispanic	-1.59	0.736	-2.2	-0.42	0.213	-2.0
log(BNP)	-2.74	0.227	-12.0	-0.22	0.055	-3.9
log(Trt)	-4.95	0.378	-13.1	-0.14	0.080	-1.7
SBP	0.40	0.138	2.9	-0.19	0.036	-5.3
log(PCR)	-4.08	0.501	-8.2	-2.76	0.146	-18.9
σ^2	31.22	1.085	28.8			
d_{11}	112.9	5.689	19.9			
d_{12}	2.29	0.849	2.7			
d_{22}	4.04	0.312	12.9			

Age is per 10 years old. SBP is per 10 units changes.

Table 4

Parameter estimates from the analysis of ESRD and death time in CRIC.

Parameter	ESRD			Death		
	Estimate	SE	t-stat	Estimate	SE	t-stat
Age	-0.02	0.031	-0.8	-0.19	0.055	-3.4
Female	-0.11	0.055	-1.9	0.04	0.096	0.4
Black	-0.12	0.053	-2.2	-0.10	0.116	-0.9
Hispanic	-0.18	0.070	-2.6	0.02	0.155	0.2
log(BNP)	-0.22	0.020	-10.8	-0.30	0.049	-6.1
log(Trt)	-0.28	0.034	-8.3	-0.36	0.069	-5.3
SBP	-0.02	0.012	-2.0	0.03	0.024	1.4
log(PCR)	-0.73	0.049	-14.8	-0.13	0.092	-1.4
b_1	0.04	0.003	15.4	0.00	0.006	0.3
b_2	0.33	0.017	18.9	0.03	0.072	0.5
s^2	0.05	0.017	2.9	1.70	0.241	7.1

Age is per 10 years old. SBP is per 10 units changes.