CSCanada
Progress in Applied Mathematics
Vol. 4, No. 2, 2012, pp. [182–195]
DOI: 10.3968/j.pam.1925252820120402.S853

ISSN 1925-251X [Print]
ISSN 1925-2528 [Online]
www.cscanada.net
www.cscanada.org

Joint Modeling of All-Cause Mortality and Longitudinally Measured Serum Albumin

Abdus Sattar[a],*, Sanjoy K. Sinha[b], Christos Argyropoulos[c] and Mark Unruh[d]

[a] Department of Epidemiology & Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH, USA.
[b] School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada.
[c] Renal and Electrolyte Division, University of Pittsburgh, PA, USA.
[d] Nephrology Division, University of New Mexico, NM, USA.

* Corresponding author.
Address: Department of Epidemiology & Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH 44106-4945, USA; E-Mail: sattar@case.edu

Received: August 26, 2012/ Accepted: October 12, 2012/ Published: October 31, 2012

Abstract: In clinical studies, longitudinal and survival data are often obtained simultaneously from the same individual. Linear mixed effects models are widely used for analyzing longitudinal continuous outcome data, while survival models are used for analyzing time-to-event data. It is a common practice to analyze these longitudinal and time-to-event data separately. However, when multivariate outcomes are obtained from a given individual, they can be correlated by nature, and one can attain considerable gain in efficiency by jointly analyzing the outcomes. An objective of this study is to analyze such multivariate data by jointly modeling longitudinally measured continuous outcomes and time-to-event data. In this joint modeling, we formulate a joint likelihood function for both outcomes and use the maximum likelihood method to estimate the parameters in the two sub-models (longitudinal and survival models). We demonstrate the merits of joint modeling by considering a joint analysis of longitudinally measured serum albumin (biomarker) and time-to-all-cause mortality data obtained from a hemodialysis (HEMO) study. This HEMO study was a large NIH (National Institute of Health) sponsored multicenter clinical trial contrasting the effects of dialysis dose and dialysis membrane permeability in end-stage

renal disease patients receiving hemodialysis. We find that the parameter estimates obtained under joint modeling of HEMO data are more efficient than those obtained under separate modeling of the outcome variables.

Key words: Hemodialysis; Joint model; Likelihood method; Mixed model; Survival model

Sattar, A., Sinha, S.K., Argyropoulos, C., & Unruh, M. (2012). Joint Modeling of All-Cause Mortality and Longitudinally Measured Serum Albumin. *Progress in Applied Mathematics, 4*(2), 182–195. Available from http://www.cscanada.net/index.php/pam/article/view/j.pam.1925252820120402.S853 DOI: 10.3968/j.pam.1925252820120402.S853

1. INTRODUCTION

In many biological, biomedical, clinical, and environmental studies, both longitudinal and survival data are measured from the same individual simultaneously. For analyzing such bivariate outcomes, a common practice is to perform separate analyses of the longitudinal and survival data. If the focus of a study is to estimate the effects of an intervention or treatment on the longitudinal and survival outcomes together, then the separate analyses of longitudinal and survival data may provide biased estimators of the effect size (Ibrahim, Chu & Chen, 2010). In this study, our objective is to perform a joint analysis of longitudinal continuous outcomes and time-to-event data. We consider a linear mixed effects model for the longitudinal outcome variable and a parametric survival model for the time-to-event data. Both models are linked through shared parameters. We estimate the model parameters and draw inferences based on the method of maximum likelihood.

Joint modeling (JM) of longitudinal and survival data is a powerful method that takes into account the association between longitudinal measurements and time-to-event data. This joint modeling framework is used to make simultaneous inference on the model parameters for a better assessment of treatment effects (Ibrahim *et al.*, 2010). In the joint modeling of longitudinal measurements and survival data, the focus may be on the change in longitudinal response or on the hazard functions or on both. Depending on the focus of the study, different inferential methods can be invoked. For example, when a goal is to understand the change in longitudinal response in relation to the treatment/intervention, all other parameters in the JM can be treated as nuisance parameters (Lang Wu, 2010). In this scenario, a pseudo-likelihood method can be utilized to make inference on the parameters of interest (Pawitan, 2001). When the focus of the study is on both process (longitudinal and survival), a joint likelihood can be developed for all observed data. The method of maximum likelihood (ML) can be used to estimate the model parameters by maximizing the joint likelihood function simultaneously. The ML estimators obtained from the joint likelihood function are efficient and asymptotically normally distributed under some regularity conditions. Specifically, when JM is used for analyzing clinical trial data, it provides: 1) efficient estimators of the treatment effects on the time-to-event, 2) efficient estimators of treatment effects on the longitudinal biomarker, and 3) improved estimators (with reduced bias and smaller standard errors) of the overall treatment effects (Chen, Ibrahim & Chu, 2011). Consequently, JM is a preferred modeling approach over the individual Cox regression model for time-to-event data. Chen *et al.* (2011) showed that JM has great implications

in designing a clinical trial. Higher efficiency of parameter estimators from a JM framework leads to higher power and lower sample size requirements to detect a desired effect size. Thus these authors advocate the inclusion of longitudinal data into the design of a study for yielding lower sample sizes with higher power compared with that of a standard design based on survival data alone.

Our motivation for the joint modeling of longitudinal and survival data arises from a hemodialysis (HEMO) study. The HEMO study was a large multi-center clinical trial of 1846 patients with end-stage renal disease where 15 clinical centers and 72 associated dialysis units participated. Eligible patients were randomly assigned in equal proportions to a 2×2 design of high/low dose and dialysis with a high/low flux dialyzer. The primary outcome was death from any cause and one of the secondary outcomes was serum albumin level. Details of the study design, patient population, intervention, follow-up, and outcomes can be found in Greene *et al.* (2000). The study findings on all-cause mortality analysis was published in Eknoyan *et al.* (2002) using the Cox regression model. In this analysis, a set of covariates were considered including interventions (Kt/V and Flux) and baseline albumin. Most of the covariates in this model were time-stationary or only the baseline measurements were used. However, there are some covariates or biomarkers in the study that are measured longitudinally.

The standard Cox regression model assumes that the hazard rate depends on time-stationary covariates. When there is a time-dependent exogenous covariate, then the extended Cox model with time-varying covariates can be used to analyze time-to-event data. When there is a time-dependent endogenous covariate (e.g., biomarker, clinical parameters, etc.) that is associated with the risk of an event, the extended Cox regression model cannot be used to estimate the risk of hazard. The values of the endogenous covariate at any occasion cannot be precisely predicted since they are governed by some random mechanism. The preceding and/or subsequent values of the endogenous covariate confound the relationship between current values of the endogenous covariate and risk of hazard. This can lead to biased estimators of regression coefficients. Even if we get consistent estimators of regression parameters, these parameters may not be given a causal interpretation (Fitzmaurice, Laird & Ware, 2004). A good discussion on exogenous and endogenous covariates and their use in Cox regression models can be found in Section 3.4 in Rizopoulos (Rizopoulos, 2012).

At this point, we would like to emphasize that the serum albumin measured as a secondary outcome variable cannot be considered as an exogenous time-dependent covariate and hence cannot be used in the extended Cox regression model to estimate the hazard risk. Instead, we can jointly model longitudinally measured serum albumin biomarker and time-to-event data (Li, Hu & Greene, 2009). We emphasize that the longitudinal biomarker albumin that we consider for modeling was a secondary outcome variable in the HEMO study, so that albumin values were collected prospectively on a pre-determined schedule and thus any missing data can be assumed to be missing at random. Hence our focus in this joint modeling is on both survival and longitudinal outcomes together with a goal of understanding the association between the two outcome processes and the effects of interventions on the outcomes. We consider a parametric approach for JM of longitudinal and survival data. Our proposed JM approach differs from Li *et al.*'s (2009) JM framework in that these authors considered a semiparametric approach for the joint analysis of longitudinal albumin biomarker and all-cause mortality data, whereas we have

considered a full likelihood approach for analyzing the data. Our proposed method and numerical results offer an extension of the method and results of Li *et al.* We consider an extended set of risk factors to adjust the effects of intervention on the outcomes, informed by the covariates used in the original and secondary analyses of the HEMO dataset.

Joint modeling has received increased attention in the last two decades, and has been described in a number of reviews. A good (nontechnical) review of joint modeling can be found in Ibrahim *et al.* (2010) and Wu *et al.* (2012), and an excellent book-length discussions on JM can be found in Rizopoulos (Rizopoulos, 2012). Tsiatis and Davidian (Tsiatis, 2004) provided another succinct technical review of earlier work on joint modeling of longitudinal and survival data. A recent work on joint modeling of censored longitudinal and survival data can be found in Pike and Weissfeld (Pike & Weissfeld, 2012). Wu *et al.* (2012) discussed that the JM arises from three different scenarios, i) survival models with measurement errors or missing data in time-varying covariates, ii) longitudinal models with informative dropouts, and iii) a longitudinal response and a time-to-event data that are associated via a latent process. A typical joint modeling framework assumes a linear mixed effects model for longitudinal data and a Cox regression model or an accelerated failure time model for survival data, where the two models for the two processes are connected via random effects. A parametric accelerated failure time model is an attractive alternative to the semi-parametric Cox proportional hazards model when the proportionality assumption is difficult to meet in studying the relationship between the survival time and time-varying covariates. In addition, when time-varying covariates have complications due to intermittent measurements at different time points for different subjects, measurement errors or missing for terminal events, JM of longitudinal and survival data can alleviate these problems (Tseng, Hsieh & Wang, 2005). Another common approach for joint inference of several models with shared unobserved variables is to use a two-stage method (Lang Wu, 2010). This is closely related to the regression calibration method in the measurement error literature and is computationally easy to implement. However, this naive two-stage method may lead to biased estimators of model parameters. Also, the standard errors of the estimators could be under-estimated by this naïve approach.

The article is organized as follows, in section "Materials and Methods", we have developed the joint modeling framework for longitudinal and survival data. In section "Illustration: HEMO study data analysis", we have presented the HEMO study results. We have offered a discussion and conclusion in the final section.

2. MATERIALS AND METHODS

We consider joint modelling of the longitudinal data and survival data when the survival times are right-censored. We use a latent variable to link the survival model for time-to-event data to the longitudinal model for biomarker measurements.

Longitudinal model

Suppose in an experiment with N individuals, the ith individual has n_i repeated biomarker measurements y_{ij} at times $t_{i1} < ... < t_{ij} < ... < t_{in_i}$. We consider describing the longitudinal outcomes y_{ij} as a function of the vectors of covariates

x_{ij} and z_{ij} using the linear mixed effects model

$$y_{ij} = x'_{ij}\beta + z'_{ij}v_i + \varepsilon_{ij} \tag{1}$$

for $i = 1, ..., N$, $j = 1, ..., n_i$, where x_{ij} is the jth row of the design matrix X_i for fixed effects and z_{ij} is the jth row of the design matrix Z_i for random effects; β is the vector of regression parameters; the vector of random effects v_i is assumed to be independently and normally distributed with mean vector zero and covarinace matrix $G(\theta)$, depending on the vector of variance components θ; the random error term ε_{ij} is assumed to be independently and normally distributed with mean zero and variance σ^2. Further, v_i and ε_{ij} are assumed to be independent.

We can rewrite model (1) in the matrix form (Laird & Ware, 1982)

$$y_i = X_i\beta + Z_iv_i + \varepsilon_i \tag{2}$$

where y_i and ε_i are the vectors of repeated responses and random error terms for the ith individual. In this setting, we can show that the response vector y_i is multivariate normally distributed with the mean vector

$$E[y_i] = \mu_i = X_i\beta$$

and the covariance matrix

$$Cov(y_i) = V_i = Z_iGZ'_i + \sigma^2 I_{n_i}.$$

The individual maximum likelihood estimators of the regression parameters β and variance components (θ, σ^2) may be obtained by maximizing the marginal log-likelihood function

$$\log L_1(\beta, \theta, \sigma^2) = \sum_{i=1}^{N} \left[-\frac{n_i}{2} \log(2\pi) - \frac{1}{2} \log |V_i| - \frac{1}{2}(y_i - X_i\beta)' V_i^{-1} (y_i - X_i\beta) \right].$$

Survival model

Let T_i denote the event time for the ith individual and δ_i denote the censoring information ($\delta_i = 0$ if T_i is right-censored; $\delta_i = 1$, otherwise). We assume that the censoring is noninformative. Let $\chi_i = (\chi_{i1}, ..., \chi_{ip})'$ denote the vector of explanatory variables measured at baseline for the ith individual. Suppose the hazard rate for the ith individual at time t follows the accelerated failure time model (Klein & Moeschberger, 2003)

$$h_i(t) = h_0(t/\exp(\eta_i)) \exp(-\eta_i) \tag{3}$$

where $\eta_i = \chi'_i\alpha + v'_i\varphi$; the latent vector v_i is used to link the survival model (3) to the longitudinal model (1) for the biomarker measurements, y_{ij}, which may be subject to measurement errors; $h_0(t)$ is a baseline hazard function at time t, depending on unknown parameters τ, and α is a p-dimensional vector of unknown regression coefficients.

Given the data $\{(t_i, \delta_i); i = 1, ..., N\}$, the conditional likelihood of the model parameters may be expressed in the form

$$L_2(\tau, \alpha, \varphi) = \prod_{i=1}^{N} \{h_i(t_i)\}^{\delta_i} S_i(t_i) \tag{4}$$

where $S_i(t)$ is the survivor function for the ith individual at time t. We consider a general Weibull accelerated failure time model for the survival data for which the baseline hazard function at time t is defined by

$$h_0(t) = \lambda\gamma t^{\gamma-1} \tag{5}$$

where λ and γ are the scale and shape parameters of the Weibull distribution, respectively. In this setting, we have $\tau = (\lambda, \gamma)'$. Then under the accelerated failure time model (3), the hazard of death at time t for the ith individual is

$$h_i(t) = [\exp(-\eta_i)]^{\gamma}\lambda\gamma t^{\gamma-1} \tag{6}$$

The corresponding survivor function at time t for the ith individual is

$$S_i(t) = \exp\{-(\exp(-\eta_i))^{\gamma}\lambda t^{\gamma}\} \tag{7}$$

Here our goal is to estimate the model parameters β, θ, σ^2, τ, α and φ by maximizing the joint likelihood for the biomarker measurements and time-to-event data.

Joint model

For the given the data $\{(y_{ij}, t_i, \delta_i);\ i = 1, ..., N;\ j = 1, ..., n_i\}$, the joint likelihood of the survival model (3) and the longitudinal model (1) can be defined as (Fitzmaurice, Davidian, Verbeke & Molenberghs, 2009)

$$L(\beta, \theta, \sigma^2, \tau, \alpha, \varphi) = \prod_{i=1}^{N} \int_{-\infty}^{\infty} \{h_i(t_i)\}^{\delta_i} S_i(t_i) f(y_i|v_i, \beta, \sigma^2) g(v_i|\theta) dv_i \tag{8}$$

where $f(y_i|v_i, \beta, \sigma^2)$ is the conditional density of the response vector y_i for the given random effects v_i and $g(v_i|\theta)$ is the density function of v_i, which is assumed to be multivariate normal. Note that when $\varphi = 0$, the joint analysis would be equivalent to the individual analyses of the survival model (3) and the longitudinal model (1). But for a non-zero φ, one can attain gain in efficiency in the ML estimators from the joint analysis of the data. We investigate this further in the analysis of hemodialysis (HEMO) data discussed in the next section.

3. ILLUSTRATION: HEMO STUDY DATA ANALYSIS

The HEMO study was a fifteen-center randomized clinical trial of the effects of HD dose and membrane flux on mortality in chronic hemodialysis. Study participants were randomized to either standard or high dose dialysis (measured with an index (single pool Kt/V, spKt/V) derived by compartmental modeling of urea kinetics during dialysis) and to either High Flux (HF) or Low Flux (LF) membranes. Dialysis dose quantification utilizing the spKt/V index, assumes that the total body urea is distributed in a single compartment of volume C, so that its dialytic removal can be described as the product of dialyzer clearance (K), dialysis treatment time (t) divided by V (Daugirdas, 1993). HF membranes allow enhanced removal of larger (high molecular weight) toxins, so that fluxness can be quantified by the in vivo attained beta-2 microglobulin clearance of < 10 ml/min or > 20 ml/min, respectively. Patients were randomized to standard (spKt/V 1.25) or high (spKt/V 1.65) and either HF or LF dialyzers in a 2×2 factorial design; randomization was stratified

by clinical center, age (older vs. younger than 55 years old) and diabetic status. Enrollment in the HEMO Study began in March 1995 and ended in October 2000, while follow-up ended in December 31st 2001.

Baseline demographics (age, sex, race, albumin levels, comorbidity scores, diabetic status, dialysis access, residual urine volume) and study related information (Kt/V and flux assignment, clinical center duration of follow up and events), causes of death and event times were ascertained from the HEMO analytic files distributed by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). In the HEMO study, *Comorbidity* was measured using the Index of Coexistent Disease (ICED) (Miskulin *et al.*, 2001), a composite scoring system based on 19 medical (Index of Disease Severity, IDS) and 11 physical impairment categories (Index of Physical Impairment, IPI). Based on the peak IDS and IPI scores, an ICED level is assigned on a 4-point scale (0-3) with a higher score reflecting greater severity. Laboratory assessments of the HEMO study participants were performed monthly and serum albumin (a marker of nutritional status and an important predictor of mortality among dialysis patients) was prospectively collected during the entire study duration. In this study, albumin was measured by nephelometry, a technique that yields more reproducible and less noisy results than the methods commonly used in clinical laboratories (Bromocresol Green or Bromocresol Purple) (Carfray *et al.*, 2000). As a biomarker and a predictor of mortality (Kovesdy & Kalantar-Zadeh, 2012), serum albumin integrates a variety of factors that directly or indirectly affect health status, including nutrition (malnourished patients have lower albumin due to limited intake), inflammation (serum albumin synthesis in the liver declines during inflammation) and the general hypercatabolic state (which increases albumin degradation) that many patients with chronic wasting diseases and conditions exhibit. Serum albumin may also be affected by the quality of dialysis, so that a priori one would like to examine the hypothesis that higher dialysis dose or enhanced removal of larger uremic toxins may affect outcomes through a change in the serum albumin level.

In this analysis, the outcome variables are time-to-death from any cause and longitudinally measured serum albumin. The data consist of 1695 patients with 1 to 16 visits over 6.55 years. The longitudinal serum albumin was measured biannually. We consider the albumin data that were coincidental with the semi-annual exams. Since we are interested in the time scales that are of relevance to the phenomenon studied (dialysis patients have an approximate annual mortality rate of 20%), so the duration of six months assessment time is considered relevant here. The time associated with the longitudinal serum albumin measurements is computed from time of randomization to the biannual exam date when blood was drawn for serum albumin and other laboratory measurements (including creatinine, a muscle breakdown product that reflects lean mass and protein intake in dialysis patients). According to Li *et al.* (2009), there is a variation in albumin levels depending on the hemodialysis schedules. The HEMO study patients received dialysis three times a week following either a Monday-Wednesday-Friday or a Tuesday-Thursday-Saturday schedule. Because of the variation in total body water, it is anticipated that the albumin concentration would be lower on Monday or Tuesday than on other days. Therefore, we created a binary indicator variable for Mon/Tue schedule to adjust for the effects of days on the longitudinally measured serum albumin.

We have rescaled a few of the covariates for stable parameter estimation and better interpretation (Hogan, Lin & Herman, 2004). Age is expressed in decades.

Serum creatinine and diastolic blood pressure are rescaled, dividing by their standard deviations. We used a dichotomized smoking variable (smoked vs. never smoked) as covariate in the regression model. The Akaike Information Criterion (AIC) is used as a guideline in selecting covariates for the regression model. A smaller AIC value generally indicates a better model. Although some of the covariates were neither statistically significant nor provided a lower AIC value, we nevertheless kept them in the regression model for their clinical relevance and face validity, such as treatments and the Mon/Tue indicator variable. Inclusion of some interaction terms in the model caused non-convergence of the model fit or warning messages that the Hessian matrix was not positive definite, and hence it was dropped from the submodel.

For analyzing the longitudinal serum albumin and time-to-death data, we considered using the following specific joint model. For the serum albumin data, we considered the linear mixed effects model

$$\begin{aligned}&\text{Albumin}_{ij} = \beta_0 + \beta_1\,\text{Creatinine} + \beta_2\,\text{Age} + \beta_3\,\text{Black Race} + \beta_4\,\text{Diabetes}\\&+\beta_5\,\text{Smoking} + \beta_6\,\text{BMI} + \beta_7\text{Yrs of Dialysis} + \beta_8\text{High Kt/V} + \beta_9\text{High Flux}\\&+\beta_{10}\text{Mon/Tue} + \beta_{11}\ \text{Time}_{ij} + \beta_{12}\ \text{Diabetes} \times \text{Time}_{ij} + v_{0i} + v_{1i}\text{Time}_{ij} + \varepsilon_{ij},\end{aligned} \tag{9}$$

where the random error terms ε_{ij} are assumed to be identically and independently distributed as $N(0, \sigma^2)$; the vectors of random effects $v_i = (v_{0i}, v_{1i})'$ are assumed to be independent multivariate normal with mean vector 0 and covariance matrix

$$G(\theta) = G(\sigma_{v11}, \sigma_{v12}, \sigma_{v22}) = \begin{pmatrix} \sigma^2_{v11} & \sigma_{v12} \\ \sigma_{v12} & \sigma^2_{v22} \end{pmatrix}$$

Further, ε_{ij} and v_i are assumed to be mutually independent. For all-cause mortality we used the accelerated failure time model (3) with

$$\begin{aligned}&\eta_i = \alpha_1\text{High Kt/V} + \alpha_2\text{High Flux} + \alpha_3\ \text{Age} + \alpha_4\,\text{Black Race} + \alpha_5\,\text{Diabetes}\\&+\alpha_6\ \text{Yrs of Dialysis} + \alpha_7\,\text{ICED score}+\psi_0 v_{0i} + \psi_1 v_{1i}\end{aligned} \tag{10}$$

where $\varphi = (\varphi_0, \varphi_1)'$ is a vector of "shared" parameters.

The likelihood function corresponding to this JM is defined in equation (8). The quasi-Newton method in the SAS NLMIXED program is used to optimize the log-likelihood function and to obtain the maximum likelihood estimates of the model parameters and their standard errors (SAS Institute Inc.). This program fits nonlinear mixed effects models and integrates over the random effects by using an adaptive Gaussian quadrature method. We used 5 quadrature points and set the relative convergence criterion 1E-9.

The baseline characteristics of all patients are presented in Table 1. In the HEMO study, 871 (47.18%) patients died during the study period. Patients who died were likely to be older, male, white, and with a higher burden of comorbid disease, and were more likely to have tunneled dialysis catheters compared to surviving patients. Patients who died were in dialysis longer and more likely to be past smokers. On the other hand, patients who died had lower residual urine output, diastolic blood pressure, calcium, phosphorus, serum total cholesterol, albumin, and creatinine. Separate analysis on all-cause mortality is presented in Table 2. This analysis is based on the accelerated failure time model (3), without the latent variable in the model. After adjusting for a number of covariates this result indicate that the higher Kt/V and Flux do not have any beneficial effects on patient survival. All

Table 1
Baseline Characteristics of 1846 HEMO Dialysis Patients, USA, 1995-2000

Factors	All patients ($N = 1846$)	Patients survived ($N = 975$)	Patients died ($N = 871$)	p-value[a]
Age	57.62(14.04)	53.20(14.1)	62.57(11.89)	<0.01
Female sex(%)	56.23	56.82	55.57	0.59
Black race(%)	62.62	64.72	60.28	0.05
BMI	25.46(5.28)	25.84(5.21)	25.04(5.33)	<0.01
Years of dialysis[‡]	3.75[0.94,4.68]	3.69[0.90,4.29]	3.82[1.02,5.01]	0.04
Residual urine output(%)[†]	12.02	14.29	9.53	<0.01
High Kt/V(%)	49.84	50.15	49.48	0.77
High flux(%)	49.89	50.46	49.25	0.60
Access(%)				
Permanent catheter	5.80	4.51	7.23	0.01
AVF/AVG/other	94.20	95.49	92.77	
Co-morbidity ICED score(%)				
0-1	35.59	47.08	22.73	
2	31.26	28.10	34.79	<0.01
3	33.15	24.82	42.48	
Blood pressure				
Systolic	151.02(25.64)	151.27(24.41)	150.74(26.98)	0.65
Diastolic	81.28(15.24)	83.25(14.71)	79.04(15.52)	<0.01
Smoking(%)				
Never	50.24	53.90	46.14	
Past	32.39	28.23	37.05	<0.01
Current	17.36	17.86	16.80	
Calcium(mg/dl)	9.34(0.99)	9.38(1.03)	9.29(0.96)	0.05
Phosphorus(mg/dl)	5.85(1.83)	5.94(1.86)	5.74(1.81)	0.02
Serum total cholesterol(mg/dl)	171.39(40.01)	171.42(40.13)	171.37(39.89)	0.98
Serum albumin (g/dl)	3.63(0.36)	3.92(0.35)	3.79(0.33)	<0.01
Serum creatinine (g/dl)	10.26(2.90)	10.99(2.98)	9.43(2.57)	<0.01

Abbreviations: BMI, body mass index; AVF, arteriovenous fistula; AVG, Arteriovenous grafts; ICED, Index of Coexistent Disease.

[a] All P values are two-sided.

[†] If the patient produces ≥200 ml/day of urine, is urea clearance measured from interdialytic urine collection>1.5 ml/min (per 35L of total urea volume). (0=no, either produces<200 ml/day or urea clearance≤1.5 ml/min, 1=yes, 9=unknown, to be determined during Baseline)

[‡] Continuous and skewed variables are summarized in the form of Median[1st Quartile, 3rd Quartile]

Table 2
Separate Analysis of the All-Cause Mortality Data by the Parametric Survival Model

Variable	MLE	Std. err	t value	p-value	95% CI	
					Lower	Upper
High Kt/V	0.0551	0.0523	1.05	0.2927	-0.0475	0.1576
High flux	0.0350	0.0523	0.67	0.5042	-0.0676	0.1375
Age/10	-0.2708	0.0244	-11.1	<.0001	-0.3186	-0.2229
Black race	0.2635	0.0540	4.88	<.0001	0.1577	0.3694
Diabetes	-0.1945	0.0568	-3.42	0.0006	-0.3059	-0.0832
Yrs of dialysis	-0.0138	0.0067	-2.05	0.0400	-0.0269	-0.0006
Comorbidity ICED score	-0.2593	0.0343	-7.57	<.0001	-0.3265	-0.1922
Scale parameter(λ)	0.0054	0.0014	3.95	<.0001	0.0027	0.0081
Shape parameter(γ)	1.3803	0.0412	33.47	<.0001	1.2994	1.4611

Table 3
Separate Analysis of the Longitudinal Albumin Data by the Linear Mixed Model

Variable	MLE	Std. err	t value	p-value	95% CI	
					Lower	Upper
Intercept	3.7871	0.0612	61.88	<.0001	3.6670	3.9071
Creatinine/SD	0.1048	0.0085	12.31	<.0001	0.0881	0.1215
Age/10	-0.0158	0.0057	-2.76	0.0058	-0.0270	-0.0046
Black race	-0.0620	0.0157	-3.94	<.0001	-0.0929	-0.0312
Diabetes	-0.0141	0.0175	-0.81	0.4187	-0.0484	0.0201
Smoking(never vs. smoked)	-0.0071	0.0146	-0.49	0.6264	-0.0358	0.0215
BMI	-0.0036	0.0014	-2.51	0.012	-0.0064	-0.0008
Yrs of dialysis	-0.0082	0.0018	-4.7	<.0001	-0.0117	-0.0048
High Kt/V	-0.0063	0.0145	-0.43	0.6662	-0.0348	0.0222
High flux	0.0004	0.0145	0.03	0.9769	-0.0281	0.0289
Mon/Tue	-0.0040	0.0061	-0.66	0.5116	-0.0160	0.0080
Time	-0.0239	0.0040	-5.89	<.0001	-0.0318	-0.0159
Diabetes×time	-0.0168	0.0063	-2.69	0.0073	-0.0291	-0.0045
Residual SD(σ)	0.2585	0.0022	117.71	<.0001	0.2542	0.2628
SD of random intercept(σ_{v11})	0.2634	0.0069	38.07	<.0001	0.2499	0.2770
Covariance(σ_{v12})	-0.0016	0.0010	-1.58	0.1144	-0.0037	0.0004
SD of random slope(σ_{v22})	-0.0595	0.0038	-15.63	<.0001	-0.0669	-0.0520

Table 4
Joint Analysis of Longitudinal Albumin and All-Cause Mortality Data by a Parametric JM

Variable	MLE	Std. err	t value	p-value	95% CI Lower	95% CI Upper
			Survival Submodel			
High Kt/V	0.0294	0.0523	0.56	0.574	-0.0731	0.1319
High flux	0.0294	0.0522	0.56	0.5732	-0.0730	0.1318
Age/10	-0.2554	0.0236	-10.83	<.0001	-0.3017	-0.2092
Black race	0.2558	0.0536	4.77	<.0001	0.1506	0.3610
Diabetes	-0.2263	0.0572	-3.96	<.0001	-0.3385	-0.1141
Yrs of dialysis	-0.0094	0.0066	-1.42	0.1552	-0.0225	0.0036
Comorbidity ICED score	-0.2117	0.0328	-6.46	<.0001	-0.2759	-0.1474
Scale parameter(λ)	0.0048	0.0013	3.77	0.0002	0.0023	0.0073
Shape parameter(γ)	1.4823	0.0473	31.34	<.0001	1.3895	1.5751
Shared intercept(φ_0)	0.9939	0.1133	8.77	<.0001	0.7718	1.2161
Shared slope(φ_1)	4.3919	0.7751	5.67	<.0001	2.8716	5.9122
			Longitudinal Submodel			
Intercept	3.8847	0.0611	63.58	<.0001	3.7649	4.0046
Creatinine/SD	0.0925	0.0084	11.01	<.0001	0.0760	0.1090
Age/10	-0.0198	0.0058	-3.41	0.0007	-0.0311	-0.0084
Black race	-0.0534	0.0159	-3.36	0.0008	-0.0846	-0.0222
Diabetes	-0.0151	0.0174	-0.86	0.3875	-0.0492	0.0191
Smoking(never vs. smoked)	0.0051	0.0143	0.36	0.7204	-0.0229	0.0332
BMI	-0.0051	0.0014	-3.61	0.0003	-0.0078	-0.0023
Yrs of dialysis	-0.0080	0.0018	-4.49	<.0001	-0.0115	-0.0045
High Kt/V	-0.0060	0.0147	-0.41	0.6824	-0.0349	0.0228
High flux	-0.0011	0.0147	-0.08	0.9398	-0.0300	0.0277
Mon/Tue	-0.0054	0.0061	-0.89	0.375	-0.0173	0.0065
Time	-0.0340	0.0045	-7.61	<.0001	-0.0428	-0.0253
Diabetes×time	-0.0211	0.0064	-3.31	0.001	-0.0336	-0.0086
Residual SD(σ)	0.2578	0.0022	117.86	<.0001	0.2535	0.2621
SD of random intercept(σ_{v11})	0.2632	0.0069	37.98	<.0001	0.2496	0.2768
Covariance(σ_{v12})	-0.0006	0.0011	-0.54	0.5909	-0.0026	0.0015
SD of random slope(σ_{v22})	0.0624	0.0039	15.81	<.0001	0.0546	0.0701

of the risk factors echo similar inference with the earlier analysis and the findings presented in the literature. Separate analysis of longitudinal serum albumin data is presented in Table 3. In addition to examining the treatment effects on serum albumin, in this analysis we focused on the association between serum creatinine and the longitudinal response variable. Creatinine is a stronger predictor of serum albumin which matches the findings with Dalrymple *et al.* (2012). In addition, age, race, BMI, years of dialysis, diabetes and time interaction are significantly (at the 5% level) associated with serum albumin level. From this analysis, we can say that there is no significant association between treatment allocation to dialysis dose and membrane flux arms and serum albumin.

Joint analysis of all-cause mortality and serum albumin is presented in Table 4. From this joint analysis, it is clear that the shared intercept and shared slope parameter estimates are significant ($p < 0.05$). Therefore our joint analysis of these two outcomes is justified and the separate analysis results are likely to be biased. Although treatment effects are not statistically significant, the parameter estimates from the JM for the treatments are different from the parameter estimates from the separate models. Most of the survival submodel parameter estimates are efficient (the standard errors of the estimates are lower) compared to the separate survival model parameter estimates. On the other hand, some of the longitudinal submodel parameter estimates are efficient, and the remaining estimates' efficiencies are competitive. The efficiency depends on the association among the outcome processes and interrelationships among the covariates considered in the model. In this JM analysis, we could not find that the Mon/Tue variable is a significant covariate ($p = 0.38$). In the separate analysis, years of dialysis is a significant risk factor ($p = 0.04$) whereas in the joint analysis it is not at all a risk factor ($p = 0.16$).

4. DISCUSSION

This work demonstrates that the joint modeling of time-to-death and longitudinally measured serum albumin provides different parameter estimates (and possibly more efficient estimates) than those obtained from the separate models. The two separate models were linked by shared parameters, which appeared to be statistically significant ($p < 0.05$). Although the treatments effects are not statistically significant based on the joint analysis of two outcomes, the parameter estimates are different than those obtained from the separate analysis of time-to-death and serum albumin. From the joint analysis, we conclude that years of dialysis is not a significant risk factor for all-cause mortality.

These findings extend earlier work demonstrating the benefit of joint analysis of time-to-death and longitudinally measured serum albumin (Li *et al.*, 2009). This finding is important for understanding the association and dependence between the two outcome processes. The JM provides precise estimates of the interventions and various risk factors. Based on this finding, patients may decide to start dialysis or may be evaluated for kidney transplantation. The findings of this study should be interpreted in light of several limitations. The JM considered here is in between time-to-death and serum albumin. We did not consider all endogenous longitudinally measured biomarkers for the JM. In our ongoing research on joint modeling of longitudinal and survival data, we will consider extending this work by including several longitudinal biomarkers in the joint analysis. In conclusion, this work suggests that separate analyses of survival and longitudinal data may provide inef-

ficient estimates of the model parameters, as compared to those obtained from the joint analysis in the case when the two outcome processes are strongly associated.

ACKNOWLEDGEMENT

We thank Robert C. Elston for helpful comments and suggestions that led to improvements in the presentation. Sanjoy Sinha is grateful for the support provided by a grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

[1] Carfray, A., Patel, K., Whitaker, P., Garrick, P., Griffiths, G.J., & Warwick, G.L. (2000). Albumin as an outcome measure in haemodialysis in patients: the effect of variation in assay method. *Nephrology Dialysis Transplantation, 15*(11), 1819–1822.

[2] Chen, L.M., Ibrahim, J.G., & Chu, H. (2011). Sample size and power determination in joint modeling of longitudinal and survival data. *Statistics in Medicine, 30*(18), 2295–2309.

[3] Dalrymple, L.S., Johansen, K.L., Chertow, G.M., Grimes, B., Anand, S., McCulloch, C.E., & Kaysen, G.A. (2012). Longitudinal measures of serum albumin and prealbumin concentrations in incident dialysis patients: the comprehensive dialysis study. *Journal of Renal Nutrition.*

[4] Daugirdas, J.T. (1993). Second generation logarithmic estimates of single-pool variable volume Kt/V: an analysis of error. *Journal of the American Society of Nephrology, 4*(5), 1205–1213.

[5] Eknoyan, G., Beck, G.J., Cheung, A.K., Daugirdas, J.T., Greene, T., Kusek, J.W., Allon, M., ..., Toto, R. (2002). Effect of dialysis dose and membrane flux in maintenance hemodialysis. *New England Journal of Medicine, 347*(25), 2010–2019.

[6] Fitzmaurice, G., Davidian, M., Verbeke, G., & Molenberghs, G. (2009). *Longitudinal data analysis.* Boca Raton, FL: Chapman & Hall/CRC.

[7] Fitzmaurice, G., Laird, N.M., & Ware, J.H. (2004). *Applied longitudinal ana/-lysis.* New Jersey: Wiley.

[8] Greene, T., Beck, G.J., Gassman, J.J., Gotch, F.A., Kusek, J.W., Levey, A.S., Levin, N.W., *et al.* (2000). Design and statistical issues of the hemodialysis (HEMO) study. *Controlled Clinical Trials, 21*(5), 502–525.

[9] Hogan, J.W., Lin, X., & Herman, B. (2004). Mixtures of varying coefficient models for longitudinal data with discrete or continuous nonignorable dropout. *Biometrics, 60*(4), 854–864.

[10] Ibrahim, J.G., Chu, H., & Chen, L.M. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology, 28*(16), 2796–2801.

[11] Klein, J.P., & Moeschberger, M.L. (2003). *Survival Analysis Techniques for censored and truncated data* (2nd ed.). New York: Springer.

[12] Kovesdy, C.P., & Kalantar-Zadeh, K. (2012). Accuracy and limitations of the diagnosis of malnutrition in dialysis patients. *Seminars in Dialysis, 25*(4), 423–427.

[13] Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*(4), 963–974.

[14] Li, L., Hu, B., & Greene, T. (2009). A semiparametric joint model for longitudinal and survival data with application to hemodialysis study. *Biometrics, 65*(3), 737–745.

[15] Miskulin, D.C., Athienites, N.V., Yan, G., Martin, A.A., Ornt, D.B., Kusek, J.W., Meyer, K.B., *et al.* (2001). Comorbidity assessment using the index of coexistent diseases in a multicenter clinical trial. *Kidney International, 60*(4), 1498–1510.

[16] Pawitan, Y. (2001). *In all likelihood: statistical modeling and inference using likelihood*. New York: Oxford University Press Inc.

[17] Pike, F., & Weissfeld, L. (2012). Joint modeling of censored longitudinal and event time data. *Journal of Applied Statistics*, 1–11. doi: 10.1080/02664763.20/-12.725468

[18] Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data with application in R.* New York: CRC press.

[19] SAS Institute Inc. (2008). *SAS® 9.2 enhanced logging facilities.* Cary, NC: SAS Institute Inc.

[20] Tseng, Y. K., Hsieh, P., & Wang, J. L. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika, 92*(3), 587–603.

[21] Tsiatis, A., & Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica, 14*, 809–834.

[22] Wu, L. (2010). *Mixed effects models for complex data.* New York: CRC Press.

[23] Wu, L., Liu, W., Yi, G. Y., & Huang, Y. (2012). Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. *Journal of Probability and Statistics.*