



Joint modeling in presence of informative censoring on the retrospective time scale with application to palliative care research

Quran Wu¹, Michael Daniels², Areej El-Jawahri³, Marie Bakitas⁴,
Zhigang Li^{1,*}

¹Department of Biostatistics, 2004 Mowry Rd, University of Florida, Gainesville, FL, 32610, USA

²Department of Statistics, 102 Griffin-Floyd Hall, University of Florida, Gainesville, FL, 32611, USA

³Department of Oncology, Massachusetts General Hospital, 55 Fruit St, Boston, MA, 02114, USA

⁴School of Nursing, University of Alabama at Birmingham, 1720 2nd Avenue South, Birmingham, AL, 35294, USA

*To whom correspondence should be addressed: Zhigang Li Email: zhigang.li@ufl.edu

SUMMARY

Joint modeling of longitudinal data such as quality of life data and survival data is important for palliative care researchers to draw efficient inferences because it can account for the associations between those two types of data. Modeling quality of life on a retrospective from death time scale is useful for investigators to interpret the analysis results of palliative care studies which have relatively short life expectancies. However, informative censoring remains a complex challenge for modeling quality of life on the retrospective time scale although it has been addressed for joint models on the prospective time scale. To fill this gap, we develop a novel joint modeling approach that can address the challenge by allowing informative censoring events to be dependent on patients' quality of life and survival through a random effect. There are two sub-models in our approach: a linear mixed effect model for the longitudinal quality of life and a competing-risk model for the death time and dropout time that share the same random effect as the longitudinal model. Our approach can provide unbiased estimates for parameters of interest by appropriately modeling the informative censoring time. Model performance is assessed with a simulation study and compared with existing approaches. A real-world study is presented to illustrate the application of the new approach.

KEYWORDS: Frailty model; Joint modeling; Longitudinal data analysis; Palliative care; Semi-parametric.

1. INTRODUCTION

Palliative care specialty focuses on improving quality of life (QOL) for patients and is appropriate at any age and at any stage in a serious illness; it can be provided along with curative treatment (Kelley and Morrison, 2015). Palliative care research has led to a paradigm shift in clinical practice. This includes early integration of palliative care for seriously ill older adults, such as advanced cancer patients (Temel and others, 2010; Parikh and others, 2013; Bakitas and others, 2009) and a 3-fold increase in palliative care programs at hospitals with 50+ beds across the nation since 2000 (Dumanovsky and others, 2016). With the rapid growth, it is critical that high-quality palliative care research studies are appropriately performed to generate results that can inform clinical and nursing

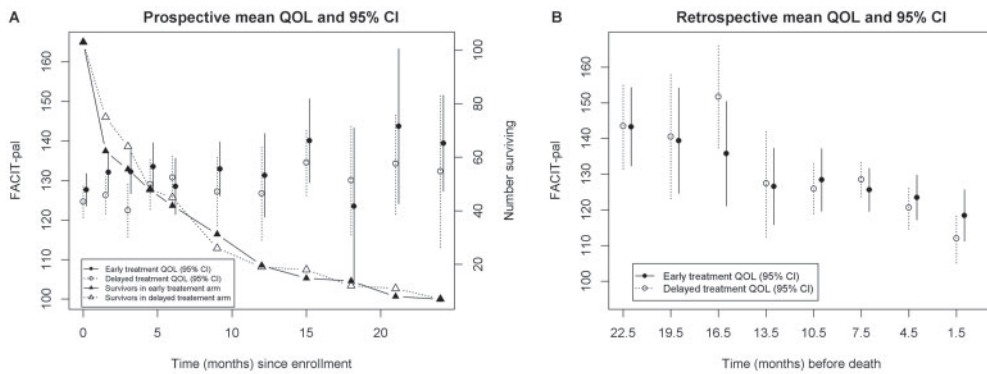


Figure 1. This figure is extracted from [Li and others \(2017\)](#) where the means and 95% CI of the QOL score (FACIT-pal) were presented on the prospective time scale (left panel) and on the retrospective time scale (right panel) for a two-arm randomized trial. Number of patients surviving at each time point was also presented in the left panel.

practice for palliative medicine to maximize its potential value throughout the health care system ([Kelley and Morrison, 2015](#)). However, despite the progress, there is still a lack of statistical analysis methods to appropriately analyze the longitudinal QOL data and survival data which are commonly seen in palliative care studies. One of the challenges is how to handle dropouts in palliative care studies when the data are analyzed with a joint model and QOL is modeled on a retrospective time scale from death ([Chan and Wang, 2010](#); [Li and others, 2013, 2017](#); [Kong and others, 2018](#)). Dropping out is often associated with QOL and survival. Such dropouts are called informative dropouts and lead to informative censoring. Missing QOL due to informative censoring is expected to be systematically different than uncensored QOL; noninformative censoring does not systematically mask low or high QOL. Both types of censoring can be observed in a typical palliative care study and should be treated differently in a statistical model because they have systematically different impacts on QOL data. To address this analysis issue in palliative care research, we propose a semiparametric model that jointly models longitudinal QOL data, death time and informative censoring time in the terminal trend model framework ([Kurland and others, 2009](#); [Li and others, 2017](#)) to make efficient and valid inference.

It is well known in the statistical literature that joint modeling of longitudinal data and survival data can improve efficiency and reduce bias ([Hogan and Laird, 1997](#); [Tsiatis and Davidian, 2004](#); [Rizopoulos and Lesaffre, 2014](#); [Dempsey and McCullagh, 2018](#); [Elashoff and others, 2017](#)). Joint models can be roughly classified based on whether the longitudinal data is modeled on the retrospective time scale from death, or prospectively from study entry, involving death as an event ([Wulfsohn and Tsiatis, 1997](#); [Hogan and Laird, 1997](#); [Henderson and others, 2000](#); [Li and others, 2010](#); [Su and Hogan, 2010](#)). The estimated longitudinal trajectories from the latter class do not have direct interpretation relative to death time ([Kurland and Heagerty, 2005](#); [Li and others, 2017](#)). Another major issue, as [Li and others \(2017\)](#) pointed out (see Figure 1), is that longitudinal trajectories on the prospective time scale could be misleading for making inference on terminal trend. As shown in Figure 1, the longitudinal trajectories on the left panel (on prospective time scale) were clearly not correct patterns for terminal trends in the two treatment arms of a randomized trial, whereas the right panel (on retrospective scale) shows decreasing terminal trends as expected. Joint models on the prospective time scale could also extrapolate the longitudinal trajectories beyond death as pointed out by [Kurland and Heagerty \(2005\)](#) who proposed a partly conditional model that can address the risk of extrapolation beyond death by modeling the longitudinal data conditional on subjects being alive. Joint models on the retrospective time scale from death ([Liu and others, 2007](#); [Chan and Wang, 2010](#); [Li and others, 2013](#);

Chan and Wang, 2017; Li and others, 2017; Kong and others, 2018) often have a different target of inference (i.e., the terminal trend relative to death) than modeling the longitudinal data on the prospective time scale. Terminal trend models (TTM) (Kurland and others, 2009; Li and others, 2013, 2017) have been proposed for jointly analyzing longitudinal QOL data and survival data in palliative care studies and have advantages for estimating QOL on the retrospective time scale and interpreting the results for studies that have relatively short life expectancies. Under the assumption of noninformative censoring, there have been developments such as Chan and Wang (2017) which excluded the longitudinal data of subjects whose death times are censored from the analysis which results in efficiency loss and others (Liu and others, 2007; Li and others, 2013, 2017; Kong and others, 2018) that include all longitudinal data in the analysis. Although informative censoring have been extensively addressed in the literature for modeling the longitudinal data on the prospective time scale (Daniels and Hogan, 2008; Elashoff and others, 2017), informative censoring remains an unresolved issue for modeling longitudinal data on the retrospective time scale because the time points of the longitudinal measures are unknown on the retrospective time scale when death time is censored and the censoring time is associated with the longitudinal data and death time.

To fill this gap, we develop a new semiparametric model under the TTM framework for analyzing data in palliative care studies or other studies with relatively short life expectancy. The joint model has two submodels: a semiparametric mixed effect submodel for the longitudinal QOL data and a competing-risk survival submodel with piecewise hazards for the death time and dropout time. Dropout and death can be modeled as competing risks because the event of dropout prevents the event of death to occur in the study (Putter and others, 2007). The mixed effect submodel uses splines to model the longitudinal trajectory, making it flexible enough to approximate linear and nonlinear curves. The piecewise hazards for the death time and dropout time are also flexible to account for variation of hazard over time in palliative care studies. Both submodels share a common random effect that is used to handle the association between dropout time, QOL, and death time.

We will use a real-world study, the ENABLE III study (Bakitas and others, 2015), to illustrate our proposed method. The ENABLE III study was a two-arm randomized clinical trial comparing early palliative care with delayed palliative care for patients with advanced cancer to evaluate and test the benefit of introducing early palliative care versus delayed palliative care. The primary endpoint was QOL assessed by the 46-item Functional Assessment of Chronic Illness Therapy-Palliative Care (FACIT-pal), the higher the better. The treatment arm received palliative care immediately after enrollment (early treatment arm) and the control arm received standard care for the first 12 weeks and then switched to palliative care (delayed treatment arm). There were 207 patients enrolled in the study between October, 2009 and March, 2013 with 103 patients randomized to the early treatment arm and 104 patients randomized to the delayed treatment arm. QOL data were collected at baseline (i.e., enrollment) and at 6, 12, 18, and 24 weeks and every 12 weeks thereafter post enrollment until death, dropout or study completion. Survival data were collected as well. Given the strong association between QOL and survival for patients with short life expectancies, it is important to jointly model the QOL and survival data to estimate and test the treatment effect on QOL because better QOL in an arm could simply be due to longer survival in that arm.

This article is organized as follows. Models and notation are introduced in Section 2 with details on the longitudinal model and hazard models for the death time and the dropout time. Section 3 describes the calculation of the full log-likelihood function and the estimation procedure. A simulation study to assess the model performance is given in Section 4. A real study application is presented in Section 5 followed by a discussion in Section 6.

2. MODEL AND NOTATION

In this article, we focus on two-arm palliative care trials. The method can be naturally extended to trials with more than two arms or studies with binary independent variables. Let Y_i , D_i , W_i , and C_i denote the vector of longitudinal QOL measurements, death time, dropout time, and administrative

censoring time (e.g., study completion) respectively for the i th subject. We will treat death and dropout as competing risks in our model. The administrative censoring time C_i can be calculated as the duration between the study completion time and the enrollment time. Let $T_i = \min(D_i, W_i, C_i)$ which is the observed event time and Δ_i take value of 0, 1, or 2, with $\Delta_i = 0$ indicating a censored event (i.e., survived beyond the end of the study), $\Delta_i = 1$ indicating a death and $\Delta_i = 2$ indicating a dropout. Let A_i denote the binary treatment assignment in a two-arm palliative care trial and X_i denote a $P \times 1$ vector of potential confounders or covariates including baseline quality of life. We specify the following longitudinal model for QOL, constructed on a retrospective time scale starting from the death time:

$$Y_i(t^*) = \beta_\mu(t^*) + A_i\beta_A(t^*) + X_i^T\psi_X + Z(t^*)U_i + \epsilon_i(t^*), \tag{2.1}$$

where $\beta_\mu(t^*)$ denotes the mean trajectory in the control group and the $\beta_A(t^*)$ denotes the time-varying treatment effect at time t^* counting backward from the time of death. Both of these functions are unspecified and assumed to be continuous. The time-varying coefficients allow for a flexible time-varying treatment effect during the terminal phase of life and testing the treatment effect at any specific time point prior to death. Here, ψ_X is the P -dimensional vector of coefficients of X_i . U_i is a scalar random effect following the standard normal distribution $N(0, 1)$. The function $Z(t^*)$ is an unspecified and continuous function of t^* . For identifiability of the model, the function $Z(t^*)$ is constrained to be positive. Typical assumptions for mixed effects models are retained here. The residual error $\epsilon_i(t^*)$ is assumed to be normally distributed with mean $E[\epsilon_i(t^*)] = 0$ and $\text{var}[\epsilon_i(t^*)] = \tau^2$, where $\epsilon_i(t^*)$ and $\epsilon_i(s^*)$ are independent for any $t^* \neq s^*$; The random effects U_i and residual error $\epsilon_i(t^*)$ are assumed independent.

REMARK 2.1 (Flexible random effect) Notice that the term $Z(t^*)U_i$ in model (2.1) has a normal distribution with mean of 0 and variance of $Z(t^*)^2$. So model (2.1) is equivalent to the model below where $U_i^*(t^*)$ is a random effect with the normal distribution $N(0, Z(t^*)^2)$ and its variance $Z(t^*)^2$ is a nonparametric function of t^* . Therefore, the model is flexible to account for a variety of different random effects with the nonparametric variance for the random effect $U_i^*(t^*)$.

$$Y_i(t^*) = \beta_\mu(t^*) + A_i\beta_A(t^*) + X_i^T\psi_X + U_i^*(t^*) + \epsilon_i(t^*),$$

REMARK 2.2 (Connection with prospective models) Let t denote time on the prospective time scale counting from enrollment. Model (2.1) can be rewritten on the prospective time scale as follows:

$$Y_i(t) = \beta_\mu(D_i - t) + A_i\beta_A(D_i - t) + X_i^T\psi_X + Z(D_i - t)U_i + \epsilon_i(t).$$

It is straightforward to see that QOL, $Y_i(t)$, depends on the death time through $D_i - t$ which measures how soon the subject will die after time t . Later we will see that QOL also depends on the death time through the random effect U_i . Let n_i denote the total number of follow-up longitudinal measurements, t_{ij} denote the measurement time of the j th measure on the prospective time scale and t_{ij}^* denote the measurement time of the j th measure on the retrospective time scale for the i th subject; Note $t_{ij}^* = D_i - t_{ij}$. When D_i is censored, t_{ij}^* cannot be observed which is a challenge for parameter estimation under informative censoring, such as dropouts, when modeling the longitudinal data on the retrospective time scale.

Let δ take values in $\{1, 2\}$ with $\delta=1$ indicating the event of death and $\delta=2$ indicating the event of dropout. We use a piecewise-exponential competing-risk model for the time to event data (T_i, Δ_i) . Dropout and death can be modeled as competing risks because the event of dropout prevents

the event of death to occur in the study (Putter and others, 2007). The hazard function for the competing risk model is given by:

$$\lambda_i^\delta(t) = \exp\left(\alpha_0^\delta(t) + A_i\alpha_A^\delta(t) + \alpha_U^\delta U_i\right), \text{ for } \delta = 1, 2,$$

where $\lambda_i^\delta(t)$ is the hazard rate for type δ event at time t given treatment status A_i and frailty U_i . Here, we assume $(\alpha_0^1(t), \alpha_0^2(t))$ and $(\alpha_A^1(t), \alpha_A^2(t))$ are piecewise-exponential hazard; and $\alpha_U = (\alpha_U^1, \alpha_U^2)$ are the coefficients of the frailty U_i for the two competing risks. Without loss of generality, we assume the piecewise-constant hazards, $\alpha_0^1(t)$ and $\alpha_A^1(t)$, have the same J breakpoints denoted by $\{t_1^\alpha, \dots, t_J^\alpha\}$, $0 < t_1^\alpha < \dots < t_J^\alpha < \infty$. We also assume the piecewise-constant hazards of dropout, $\alpha_0^2(t)$ and $\alpha_A^2(t)$, have the same J' breakpoints denoted by $\{t_1^\gamma, \dots, t_{J'}^\gamma\}$, $0 < t_1^\gamma < \dots < t_{J'}^\gamma < \infty$. We can write the piecewise-constant hazards as follows:

$$\alpha_0^1(t) = \sum_{j=1}^{J+1} \eta_0^j \mathcal{B}_j(t), \alpha_A^1(t) = \sum_{j=1}^{J+1} \eta_A^j \mathcal{B}_j(t), \alpha_0^2(t) = \sum_{j'=1}^{J'+1} \gamma_0^{j'} \mathcal{B}_{j'}^\gamma(t), \alpha_A^2(t) = \sum_{j'=1}^{J'+1} \gamma_A^{j'} \mathcal{B}_{j'}^\gamma(t),$$

where $\mathcal{B}_j(t) = 1_{(t_{j-1}^\alpha \leq t < t_j^\alpha)}$ and $\mathcal{B}_{j'}^\gamma(t) = 1_{(t_{j'-1}^\gamma \leq t < t_{j'}^\gamma)}$ are binary indicator functions, $t_0^\alpha = t_0^\gamma = 0$, $t_{J+1}^\alpha = t_{J'+1}^\gamma = \infty$ and the coefficients: $\{\eta_0^j, \eta_A^j, \gamma_0^{j'}, \gamma_A^{j'}\}$, $j = 1, \dots, J + 1, j' = 1, \dots, J' + 1$ of the indicator functions determine the constant hazard on each piece. Note that this model shares the same random effect, U_i with the longitudinal model (2.1). We can add covariates, X_i , in the model as follows:

$$\lambda_i^\delta(t) = \exp\left(\alpha_0^\delta(t) + A_i\alpha_A^\delta(t) + \alpha_U^\delta U_i + X_i^T \alpha_X^\delta\right), \text{ for } \delta = 1, 2, \tag{2.2}$$

where $\alpha_X = (\alpha_X^1, \alpha_X^2)$ are the parameters associated with X_i for the two competing risks, respectively. Notice that the above model can be also written in the form of a Cox model:

$$\lambda_i^\delta(t) = \lambda_0^\delta(t) \exp\left(A_i\alpha_A^\delta(t) + \alpha_U^\delta U_i + X_i^T \alpha_X^\delta\right), \text{ for } \delta = 1, 2,$$

where $\lambda_0^1(t)$ and $\lambda_0^2(t)$ are the baseline hazard functions for the two competing risks, respectively, in the Cox model, and $\lambda_0^1(t) = \exp(\alpha_0^1(t))$ and $\lambda_0^2(t) = \exp(\alpha_0^2(t))$.

Taken together, (2.1) and (2.2) form our joint model of longitudinal QOL, death time, and dropout time.

3. PARAMETER ESTIMATION WITH REGRESSION SPLINES

Maximum likelihood estimation (MLE) will be used to estimate the parameters. We use regression splines to handle the nonparametric functions $\beta_\mu(t^*), \beta_A(t^*)$, and $Z(t^*)$. Without loss of generality, we use the same knots for the regression splines of the three functions and the knots will be placed at equally spaced quantiles of the observed times of the longitudinal measurements. In particular, $\beta_\mu(t^*), \beta_A(t^*)$, and $Z(t^*)$ will be approximated by linear splines with K knots:

$$\beta_\mu(t^*) = \beta_0 + \sum_{k=1}^K \beta_k B_k(t^*), \beta_A(t^*) = \beta_0^A + \sum_{k=1}^K \beta_k^A B_k(t^*), Z(t^*) = \zeta_0 + \sum_{k=1}^K \zeta_k B_k(t^*), \tag{3.3}$$

where $B_k(t^*) = (t^* - t^{k-1})1_{(t^{k-1} \leq t^* < t^k)} + (t^k - t^{k-1})1_{(t^k \leq t^*)}$ are the linear spline basis functions, $t^0 = 0$ and $0 < t^1 < \dots < t^K < \infty$ are the knots which will be placed at the equally spaced sample

quantiles of the data. Notice that the curves become flat after the last knot. Equation (2.1) can be rewritten as:

$$Y_i(t^*) = \mathbb{X}_i(t^*)^T \beta + Z(t^*)U_i + \epsilon_i(t^*), \tag{3.4}$$

where $\mathbb{X}_i(t^*)^T = (1, B_1(t^*), \dots, B_K(t^*), 1, A_1B_1(t^*), \dots, A_KB_K(t^*), X_i^T)$ is a $(2K + P)$ -dimensional vector of variables, $\beta = (\beta_0, \beta_1, \dots, \beta_K, \beta_0^A, \beta_1^A, \dots, \beta_K^A, \psi_X^T)^T$. As defined previously, n_i denotes the number of measurements of the longitudinal outcome for the i th subject. Since $\epsilon_i(t^*)$ follows a normal distribution, conditional on U_i , the outcome variable vector $Y_i = (Y_i(t_{i1}^*), \dots, Y_i(t_{in_i}^*))^T$ has a n_i -dimensional multivariate normal distribution $N(\mu_i, V_i)$ with mean vector μ_i having its j th element equal to $\mathbb{X}_i(t_{ij}^*)^T \beta + Z(t_{ij}^*)U_i$ and covariance matrix $V_i = \tau^2 I_{n_i}$, where I_{n_i} is the $n_i \times n_i$ identity matrix. Here, t_{ij}^* denotes the measurement time of the j th measure on the retrospective time scale for the i th subject. Using equations (2.2)–(3.4), the log-likelihood function can be constructed to obtain MLE, given prespecified knots, for all the parameters denoted by $\theta = (\beta, \zeta, \Sigma, \tau, \{\eta_0^j, \eta_A^j, j = 1, \dots, J\}, \{\gamma_0^j, \gamma_A^j, j = 1, \dots, J'\}, \alpha_U, \alpha_X)$, where $\zeta = (\zeta_0, \zeta_1, \dots, \zeta_K)$. We will use equally spaced quantiles as candidates for knots and the optimal number of knots will be determined by using AIC (Akaike, 1974).

3.1. The log-likelihood function

Let $(y_i, t_i, \delta_i, x_i)$ denote the observed data for subject i , where y_i is the vector of observed longitudinal outcomes $(y_i(t_{i1}^*), \dots, y_i(t_{in_i}^*))^T$, t_i is the observed event time or censoring time and δ_i is the observed value of Δ_i . Due to the random effect U_i , we need to integrate over the distribution of U_i to calculate the log-likelihood function. To derive the log-likelihood function, we divide the subjects into six groups based on the value of δ_i and whether they have any longitudinal QOL measurements. Details of the derivation can be found in the [Supplementary material](#) available at *Biostatistics* online. For each of the six groups, we show the log-likelihood contribution of the i th subject if they are in that group. Group 1 consists of subjects who had death times observed (i.e., $\delta_i = 1$) and had longitudinal QOL measurements. The log-likelihood contribution of subject i is given by:

$$\ell_1^i(\theta) = \log \left(\int_{-\infty}^{\infty} \exp(\ell_1^{i1}(\theta|u)) du \right) - 0.5(n_i + 1) \log(2\pi) + \log(1 - F_{C_i}(t_i)), \tag{3.5}$$

where

$$\begin{aligned} \ell_1^{i1}(\theta|u) = & -n_i \log(\tau) - \frac{0.5(y_i - \mu_i)^T (y_i - \mu_i)}{\tau^2} - H_i(t_i, u) \\ & + \alpha_{1U}^1 u + \alpha_0^1(t_i) + A_i \alpha_A^1(t_i) + x_i^T \alpha_X^1 - \Lambda_i(t_i, u) - 0.5u^2, \end{aligned}$$

$H_i(t, u) = \int_0^t \lambda_i^2(v|U_i = u) dv$, $\Lambda_i(t, u) = \int_0^t \lambda_i^1(v|U_i = u) dv$, and $F_{C_i}(t)$ is the cumulative distribution function (CDF) of C_i .

Group 2 consists of subjects who dropped out ($\delta_i = 2$) and had longitudinal QOL measurements before dropout. The log-likelihood contribution of subject i is given by:

$$\ell_2^i(\theta) = \log \left(\int_{-\infty}^{\infty} \exp(\ell_2^{i1}(\theta|u)) du \right) - 0.5(n_i + 1) \log(2\pi) + \log(1 - F_{C_i}(t_i)), \tag{3.6}$$

where

$$\begin{aligned} \ell_2^{i1}(\theta|u) &= -n_i \log(\tau) \\ &+ \log \left(\int_{t_i}^{\infty} \exp \left(-\frac{0.5(y_i - \mu_i)^T (y_i - \mu_i)}{\tau^2} \right) \lambda_i^1(s|U_i = u) \right. \\ &\quad \left. \exp \left(\Lambda_i(t_i, u) - \Lambda_i(s, u) \right) ds \right) \\ &- \Lambda_i(t_i, u) + \alpha_U^2 u + \alpha_0^2(t_i) + A_i \alpha_A^2(t_i) + x_i^T \alpha_X^2 - H_i(t_i, u) - 0.5u^2. \end{aligned}$$

Group 3 consists of subjects who survived beyond the end of the study ($\delta_i = 0$) and had longitudinal QOL measurements. The log-likelihood contribution of subject i is given by:

$$\ell_3^i(\theta) = \log \left(\int_{-\infty}^{\infty} \exp(\ell_3^{i1}(\theta|u)) du \right) - 0.5(n_i + 1) \log(2\pi) + \log(f_{C_i}(t_i)), \quad (3.7)$$

where

$$\begin{aligned} \ell_3^{i1}(\theta|u) &= -n_i \log(\tau) \\ &+ \log \left(\int_{t_i}^{\infty} \exp \left(-\frac{0.5(y_i - \mu_i)^T (y_i - \mu_i)}{\tau^2} \right) \lambda_i^1(s|U_i = u) \right. \\ &\quad \left. \exp \left(\Lambda_i(t_i, u) - \Lambda_i(s, u) \right) ds \right) \\ &- \Lambda_i(t_i, u) - H_i(t_i, u) - 0.5u^2, \end{aligned}$$

and $f_{C_i}(t)$ is the density function of C_i .

Group 4 consists of subjects who had death times observed ($\delta_i = 1$) but had no longitudinal QOL measurements. The log-likelihood contribution of subject i is given by:

$$\ell_4^i(\theta) = \log \left(\int_{-\infty}^{\infty} \exp(\ell_4^{i1}(\theta|u)) du \right) - 0.5 \log(2\pi) + \log(1 - F_{C_i}(t_i)), \quad (3.8)$$

where

$$\ell_4^{i1}(\theta|u) = -H_i(t_i, u) + \alpha_U^1 u + \alpha_0^1(t_i) + A_i \alpha_A^1(t_i) + x_i^T \alpha_X^1 - \Lambda_i(t_i, u) - 0.5u^2.$$

Group 5 consists of subjects who dropped out ($\delta_i = 2$) and had no longitudinal QOL measurements before dropout. The log-likelihood contribution of subject i is given by:

$$\ell_5^i(\theta) = \log \left(\int_{-\infty}^{\infty} \exp(\ell_5^{i1}(\theta|u)) du \right) - 0.5 \log(2\pi) + \log(1 - F_{C_i}(t_i)), \quad (3.9)$$

where

$$\ell_5^{i1}(\theta|u) = -\Lambda_i(t_i, u) + \alpha_U^2 u + \alpha_0^2(t_i) + A_i \alpha_A^2(t_i) + x_i^T \alpha_X^2 - H_i(t_i, u) - 0.5u^2.$$

Group 6 consists of subjects who survived beyond the end of the study ($\delta_i = 0$) but had no longitudinal QOL measurements. The log-likelihood contribution of subject i is given by:

$$\ell_6^i(\theta) = \log \left(\int_{-\infty}^{\infty} \exp(\ell_6^{i1}(\theta|u)) du \right) - 0.5 \log(2\pi) + \log(f_{C_i}(t_i)),$$

where

$$\ell_6^{i1}(\theta|u) = -\Lambda_i(t_i, u) - H_i(t_i, u) - 0.5u^2. \quad (3.10)$$

Notice that the terms involving π , $F_{C_i}(t)$, and $f_{C_i}(t)$ can be ignored because they are constants with respect to the parameter vector θ . Thus, we just need to maximize the following objective function to compute the MLE:

$$\ell(\theta) = \sum_{h=1}^6 \sum_{i \in \text{group } h} \log \left(\int_{-\infty}^{\infty} \exp(\ell_h^{i1}(\theta|u)) du \right).$$

Standard errors will be obtained using the observed Fisher information. Details of calculating the integrals with Gauss–Hermite quadrature (Takahasi and Mori, 1973) can be found in the [Supplementary material](#) available at *Biostatistics* online.

3.2. Computational challenge

An alternative way to obtain the MLE is to use an EM algorithm (Rizopoulos, 2012; Li and others, 2022) where the random effect U_i is treated as missing data. For joint models on the prospective time scale, the M-step of an EM algorithm typically has closed-form solutions for the coefficient and variance parameters of the longitudinal submodel and for the cumulative baseline hazard of survival submodel. And it can use one-step Newton–Raphson method to update the coefficient parameters of the survival submodel. All integrations can be calculated with Gauss–Hermite quadrature approximations (Li and others, 2022). The entire process of an EM algorithm for joint models on the prospective time scale is reasonably fast because it does not involve optimization, thanks to the closed-form solutions and one-step Newton–Raphson method in the M-step.

However, for our joint model on the retrospective time scale, due to the integrations with respect to ds in $\ell_2^{i1}(\theta)$ and $\ell_3^{i1}(\theta)$, the M-step of an EM algorithm does not have closed-form solutions for the coefficient parameters in the longitudinal submodel, and thus optimization is needed in each iteration of the M-step which is very computationally demanding. The root of this challenge comes from the unknown time origins of the terminal trends in groups 2 and 3 where death times are censored and thus the integrations with respect to death time is needed to construct the log-likelihood function. Because of the computational challenge for using an EM algorithm, we propose to directly maximize the log-likelihood function to obtain the MLE.

3.3. Choosing the optimal number of knots

The advantage of a regression splines approach is that it can transform a nonparametric or semiparametric model into a parametric model by approximating the nonparametric functions with linear combinations of the spline basis functions once the knots are determined. There are many ways for choosing the optimal number of knots such as AIC, BIC, cross-validation, and general cross validation. In our setting, since we have a log-likelihood function, it is natural to use AIC or BIC for selecting the optimal knots for $\beta_\mu(t^*)$, $\beta_A(t^*)$, and $Z(t^*)$ because both AIC and BIC statistics are functions of the log-likelihood function. In the simulation Section 4 and the real study Section 5, we use AIC since it puts smaller penalty than BIC on the number of parameters. The set of knots that minimizes the AIC statistic will be our set of optimal knots.

4. SIMULATION

Model performance was evaluated by a simulation study where 100 datasets were randomly generated with 200 patients in each dataset. The binary treatment variable A_i was generated using a Bernoulli distribution with parameter $p = 0.5$. Two covariates were also generated with one from the Uniform distribution $U(0, 1)$, denoted by X^1 , and the other one from the Bernoulli distribution

with the parameter $p = 0.5$, denoted by X^2 . The death time and dropout time were generated based on the following two-piece piecewise hazard:

$$\lambda_i^\delta(t) = \exp\left(\alpha_0^\delta(t) + A_i\alpha_A^\delta(t) + \alpha_U^\delta U_i + \alpha_1^\delta X^1 + \alpha_2^\delta X^2\right), \text{ for } \delta = 1, 2,$$

where $\alpha_0^1(t) = \eta_0^1 1_{(0 \leq t < t_1^\alpha)} + \eta_0^2 1_{(t_1^\alpha \leq t)}$, $\alpha_A^1(t) = \eta_A^1 1_{(0 \leq t < t_1^\alpha)} + \eta_A^2 1_{(t_1^\alpha \leq t)}$, $\alpha_0^2(t) = \gamma_0^1 1_{(0 \leq t < t_1^\gamma)} + \gamma_0^2 1_{(t_1^\gamma \leq t)}$, $\alpha_A^2(t) = \gamma_A^1 1_{(0 \leq t < t_1^\gamma)} + \gamma_A^2 1_{(t_1^\gamma \leq t)}$, $t_1^\alpha = t_1^\gamma = 3$, $\eta_0^1 = -2.5$, $\eta_0^2 = -3$, $\eta_A^1 = -1$, $\eta_A^2 = -0.5$, $\gamma_0^1 = -2.5$, $\gamma_0^2 = -3$, $\gamma_A^1 = -1$, $\gamma_A^2 = -0.5$, $\alpha_U^1 = -1.5$, $\alpha_U^2 = 1.5$, $\alpha_1^1 = -0.2$, $\alpha_1^2 = 0.1$, $\alpha_2^1 = -0.2$, and $\alpha_2^2 = 0.1$. The parameter values were chosen to mimic the distribution of the six groups in the real data.

Here, η_0^1 and η_0^2 are the hazards of death for the first piece and second piece respectively in the control arm, and $\eta_0^1 + \eta_A^1$ and $\eta_0^2 + \eta_A^2$ are the hazards of death for the first piece and second piece respectively in the treatment arm. The parameters γ_0^1 and γ_0^2 can be interpreted as the hazards of dropout for the first piece and the second piece respectively in the control arm, and $\gamma_0^1 + \gamma_A^1$ and $\gamma_0^2 + \gamma_A^2$ are the hazards of dropout for the first piece and the second piece respectively in the treatment arm. The length of the study was set to be 25 months. Enrollment times are generated from the Uniform distribution $U(0, 20)$ meaning that enrollment stops at 20 months and patients are entering the study at a stable speed which implies that the administrative censoring time follows the Uniform distribution $U(5, 25)$.

After the death time was generated, the longitudinal data Y_i was generated with the following model:

$$Y_i(t^*) = \beta_\mu(t^*) + A_i\beta_A(t^*) + \psi_1 X^1 + \psi_2 X^2 + Z(t^*)U_i + \epsilon_i(t^*),$$

where $\beta_\mu(t^*) = 20 - \frac{4}{1 + \exp(0.3t^*)}$, $\beta_A(t^*) = 10 + 4 \exp(-0.3t^* - 0.5)$, $Z(t^*) = 10 - \frac{4}{1 + \exp(0.3t^*)}$, $\psi_1 = 2$, and $\psi_2 = 1$. And U_i and $\epsilon_i(t^*)$ were generated from the standard normal distribution $N(0, 1)$ and $N(0, 2^2)$ respectively. Here, the curves $\beta_\mu(t^*)$ and $\beta_A(t^*)$ were chosen such that they proportionally mimic trajectories in a real study (Bakitas and others, 2009; Li and others, 2013). The QOL measurements are assumed to be collected every 2 months until death, dropout, or study end. On average, there are 38.8%, 38.8%, 19.6%, 1.4%, 1.5%, and 0.0% of subjects in groups 1, 2, 3, 4, 5, and 6, respectively in the simulated data sets. In the estimation, the optimal number of knots for estimating $\beta_\mu(t^*)$ and $\beta_A(t^*)$ were selected using AIC (Akaike, 1974) within the range $[2, 7]$. We compared the proposed approach with the naive approach that only analyzed subjects whose death time were observed and another method (Li and others, 2017) assuming noninformative censoring (referred to as NIC herein). The average bias magnitude was calculated as the average of the bias magnitudes at 100 points equally taken within $(0, 25)$.

Figures 2 show that the estimates from the proposed approach (green curves) were virtually unbiased with the average bias magnitude over the interval $(0, 25)$ being 0.321 (1.63%) for $\hat{\beta}_\mu$ and 0.380 (3.69%) for $\hat{\beta}_A$, whereas the naive approach (blue curves) had larger average bias magnitudes: 5.75 (29.3%) for $\hat{\beta}_\mu$ and 1.59 (15.7%) for $\hat{\beta}_A$, and NIC approach (red curves) also had much larger bias magnitudes: 1.36 (6.90%) for $\hat{\beta}_\mu$ and 0.91 (8.89%) for $\hat{\beta}_A$. The naive approach also had 35.3% wider point-wise 95% confidence intervals for β_A . The CP of point-wise 95% CI showed that proposed approach had a CP closer to the nominal level 95% than the NIC approach (0.900 vs 0.767 for β_μ , 0.925 vs 0.854 for β_A). Hence, the proposed approach was more efficient than naive approach, and more accurate than both the naive approach and the NIC approach. The underestimated coverage probabilities from the naive method and the NIC method were likely due to their estimation bias.

For the parametric part of the model, estimates showed similar pattern (see Table 1) that the proposed approach produced nearly unbiased estimates with the CP of 95% CI around the nominal level of 95% whereas parameter estimates from the naive approach and the NIC approach had large

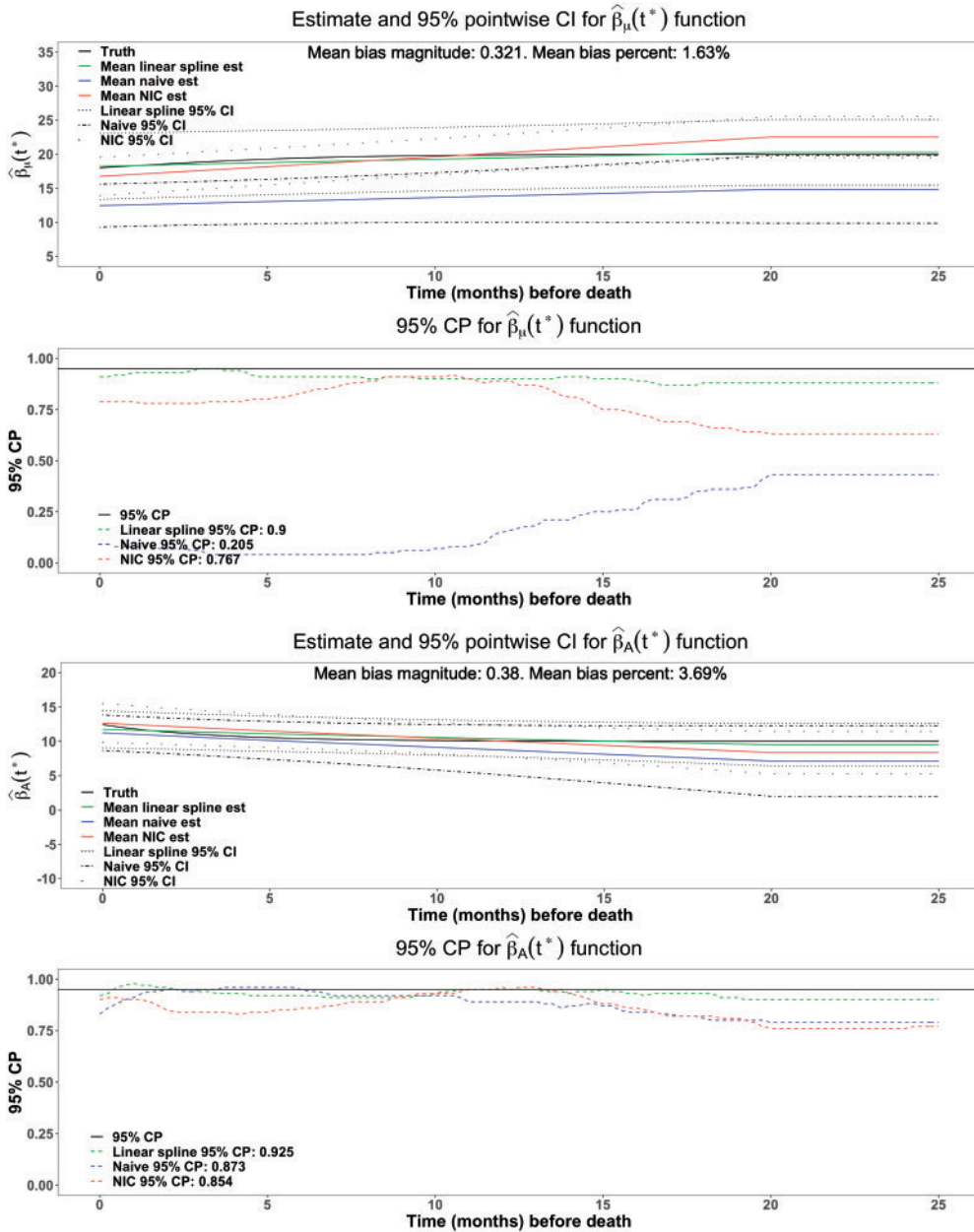


Figure 2. Comparison with the naive method and NIC method for the estimate of β_μ and β_A . The proposed method, naive method and NIC method were colored as green, blue, and red, respectively. The mean bias, bias percentage, and CP were calculated from the average of 100 equally taken points.

biases and several CP's of their 95% CI's were significantly less than 95% although some are around or above 95%.

5. APPLICATION

As described in Section 1, we use the ENABLE III study (Bakitas and others, 2015) to illustrate the proposed model. It was a two-arm randomized clinical trial to investigate the effect of early

Table 1. Simulation results for parameters in the model

Parameters	True	Bias	Naive bias	NIC bias	Bias (%)	Naive bias (%)	NIC bias (%)	CP (%)	Naive CP (%)	NIC CP (%)
η_0^1	-2.5	0.023	0.644	0.412	0.920	25.760	16.480	90	37	65
η_A^1	-1.0	-0.062	0.140	0.259	-6.200	14.000	25.900	91	88	87
α_U^1	-1.5	-0.008	—	—	-1.733	—	—	90	—	—
τ	2.0	0.003	-0.003	0.028	0.150	-0.150	1.400	93	93	92
γ_0^1	-2.5	-0.044	0.664	—	-1.760	26.560	—	95	35	—
γ_A^1	-1.0	-0.117	0.089	—	-11.700	8.900	—	96	92	—
α_U^2	1.5	0.063	—	—	4.200	—	—	96	—	—
η_0^2	-3.0	-0.040	0.090	-0.789	-1.333	3.000	-26.300	91	89	33
η_A^2	-0.5	0.049	0.347	0.729	9.800	69.400	145.800	94	77	44
γ_0^2	-3.0	-0.011	0.324	—	-0.367	10.800	—	91	79	—
γ_A^2	-0.5	-0.108	0.067	—	-21.600	13.400	—	89	93	—
α_1^1	-0.2	-0.041	-0.016	0.051	-20.500	-8.000	25.500	94	95	94
α_1^2	0.1	-0.013	-0.020	-0.034	-13.000	-20.000	-34.000	91	95	89
α_2^1	-0.2	0.030	0.053	—	15.000	26.500	—	92	97	—
α_2^2	0.1	-0.009	-0.041	—	-9.000	-41.000	—	87	95	—
ψ_1	2.0	0.106	-0.276	0.048	5.300	-13.800	2.400	90	100	91
ψ_2	1.0	0.004	0.080	0.074	0.400	8.000	7.400	87	100	86

introduction of palliative care (treatment arm) for advanced cancer patients versus delayed introduction (control arm). We jointly analyzed the longitudinal QOL data and the time-to-event data using the proposed approach. Unlike the previous analysis in (Li and others, 2017) where no dropout was defined because death times were extracted from cancer registries and study completion was used as the censoring time for subjects who survived beyond the study end, here a patient was considered a dropout if s/he did not have any data at the last visit and the dropout time was defined as the last visit time that had longitudinal QOL data. Based on the censoring status and the availability of longitudinal QOL data, there were 33.8%, 28.0%, 13.0%, 10.6%, 14.5%, and 0.0% patients in groups 1, 2, 3, 4, 5, and 6 respectively (as defined in Section 3.1). Two covariates, baseline QOL (QOL_0) and sex (Sex), were included in the submodels. The data were analyzed using the following submodels:

$$Y_i(t^*) = \beta_\mu(t^*) + A_i\beta_A(t^*) + \psi_1 QoL_0 + \psi_2 \text{Sex} + Z(t^*)U_i + \epsilon_i(t^*),$$

$$\lambda_i^\delta(t) = \exp\left(\alpha_0^\delta(t) + A_i\alpha_A^\delta(t) + \alpha_U^\delta U_i + \alpha_1^\delta QoL_0 + \alpha_2^\delta \text{Sex}\right), \text{ for } \delta = 1, 2,$$

where the survival model is a three-piece exponential model with tertiles of observed times being the breakpoints for the pieces, $\alpha_0^1(t) = \sum_{j=1}^3 \eta_0^j \mathcal{B}_j(t)$, $\alpha_A^1(t) = \sum_{j=1}^3 \eta_A^j \mathcal{B}_j(t)$, $\mathcal{B}_j(t) = 1_{(t_{j-1}^\alpha \leq t < t_j^\alpha)}$, $t_1^\alpha = 6.51$, and $t_2^\alpha = 10.26$ are the tertiles of the observed death times, and $\alpha_0^2(t) = \sum_{j=1}^3 \gamma_0^j \mathcal{B}_j'(t)$, $\alpha_A^2(t) = \sum_{j=1}^3 \gamma_A^j \mathcal{B}_j'(t)$, $\mathcal{B}_j'(t) = 1_{(t_{j-1}^\gamma \leq t < t_j^\gamma)}$, $t_1^\gamma = 3.19$, and $t_2^\gamma = 9.15$ are the tertiles of observed dropout times.

We followed the estimation procedure as described in Section 3 and determined the optimal number of knots for estimating $\beta_\mu(t^*)$, $\beta_A(t^*)$, and $Z(t^*)$ over the range $[2, 7]$ for the linear splines in the longitudinal submodel using AIC. The entire analysis took 8 h. The final model had 2 knots for the longitudinal submodel. The estimated trajectory of the longitudinal submodel is shown in Figure 3; QOL declines faster near end of life which is consistent with what has been observed in the literature (Bakitas and others, 2009; El-Jawahri and others, 2017; Bakitas and others, 2015). It also showed that the early treatment arm had better QOL but did not reach statistical significance based on pointwise Wald test which was consistent with the main finding (Bakitas and others, 2015) of the ENABLE III study.

The estimation results for parameters were presented in Table 2 and showed the longitudinal QOL was significantly associated with the baseline QOL ($\psi_1 = 0.54, p < 0.05$). Patients with early

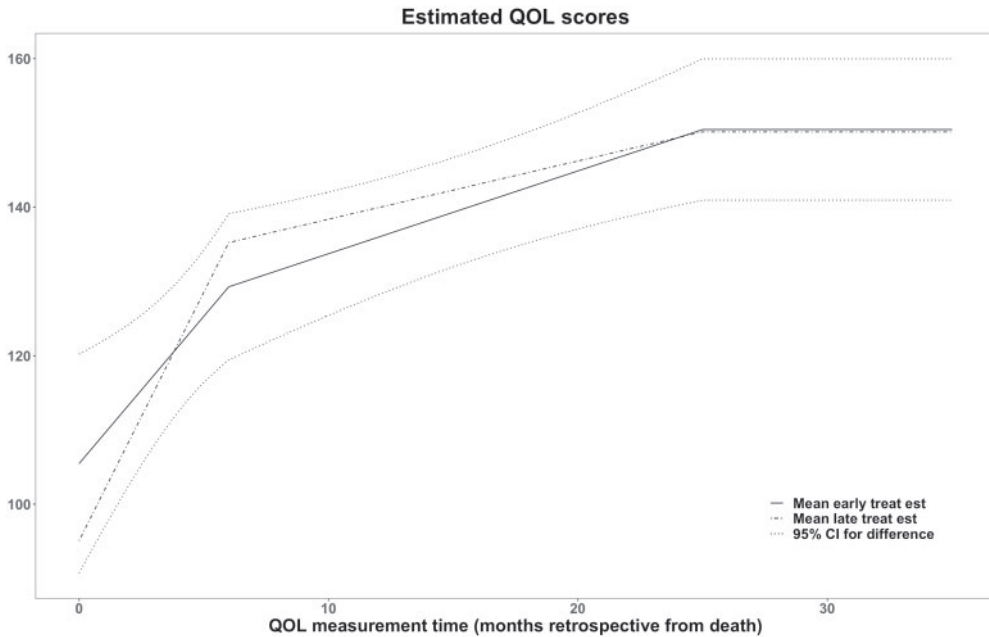


Figure 3. Fitted QOL trajectories for early and late treatment arms with 95% in ENABLE III data.

Table 2. Estimation of model parameters

Longitudinal model		Survival model		Dropout model	
Parameters	Estimate(SE)	Parameters	Estimate(SE)	Parameters	Estimate(SE)
ψ_1	0.54 (0.07)	η_0^1	-1.56 (0.60)	γ_0^1	-2.18 (0.70)
ψ_2	4.31 (2.97)	η_A^1	-0.29 (0.32)	γ_A^1	0.79 (0.37)
τ	13.31 (0.42)	α_U^1	0.08 (0.03)	α_U^2	-0.07 (0.04)
		η_0^2	-0.78 (0.63)	γ_0^2	-2.94 (0.75)
		η_0^3	-0.97 (0.64)	γ_0^3	-2.99 (0.83)
		η_A^2	-0.32 (0.43)	γ_A^2	0.19 (0.57)
		η_A^3	-0.27 (0.43)	γ_A^3	0.30 (0.62)
		α_1^1	-0.015 (0.004)	α_1^2	-0.010 (0.005)
		α_2^2	0.61 (0.24)	α_2^3	0.51 (0.25)

treatment had lower hazard of death compared to late treatment group although this impact was not statistically significant according to the Wald tests for η_A^1 , η_A^2 , and η_A^3 . Patients in the early treatment group tended to drop out faster compared to late treatment group, and this effect was statistically significant during the first 3 months but not significant after that according to the Wald tests for γ_A^1 , γ_A^2 , and γ_A^3 .

6. DISCUSSION

In this article, we proposed a novel and flexible semiparametric joint modeling approach under the TTM framework (Li and others, 2013, 2017) to account for informative censoring in palliative care studies where longitudinal QOL trajectories are often truncated by informative dropouts. The proposed method could serve as a useful data analysis tool for investigators in palliative research. The model was developed for analyzing data in two-arm clinical trials. However, it can be easily extended to trials with more than two arms and can also be used in observational studies to compare

two longitudinal trajectories because potential confounders can be adjusted for in the model. The longitudinal QOL data are modeled on the retrospective time scale from death which can provide a direct estimate of the terminal trajectory and thus a convenient interpretation for palliative care investigators. The submodels for death time and dropout time are frailty models that share the same random effect with the submodel of longitudinal QOL. The piecewise hazard for both the death time and dropout time is a flexible feature of our approach that can accommodate a variety of different survival distributions and dropout patterns. It is worth noting that QOL data missing due to missed patients' visits before dropout is assumed to be missing at random.

An improvement of the proposed approach over existing retrospective joint models (Chan and Wang, 2017; Kong and others, 2018; Li and others, 2017) is that it can handle informative censoring, such as informative dropouts, that are associated with low QOL. And the improvements were demonstrated in the simulation results where the proposed model had better performance with respect to bias and coverage probability in the presence of informative censoring. A possible extension of the proposed model is to use cubic splines (e.g., B-splines) instead of linear splines for the longitudinal submodel since cubic splines have more degrees of freedom. However, with cubic splines, it would be more time-consuming to calculate the integrals for groups 2 and 3 as described in [Supplementary material](#) available at *Biostatistics* online because there would be no explicit formulas for calculating the nested integrations. A second possible extension is to conduct sensitivity analysis to study the unverifiable assumption of conditional independence between the longitudinal trajectory and informative dropout in joint models (Daniels and Hogan, 2008; Harel and Schafer, 2009). There are existing approaches on sensitivity analysis (Creemers and others, 2010; Su and others, 2019) for traditional shared parameter models and those may be adapted under the TTM framework. Another possible extension is to remove the normality assumption on the distribution of the random effect U_i such that its distribution is unspecified. That way, the longitudinal model would have a fully nonparametric random effect (Wu and Zhang, 2006) which would give a little more flexibility over the current model.

There are two other factors that have significant impact on the computing time: the number of knots for the longitudinal model and the number of random effects. Due to the computational challenge, we only included one random effect U_i in the longitudinal submodel in (2.1). In theory, more than one random effects can be included like a typical mixed effect model. For example, we can use the longitudinal submodel as follows:

$$Y_i(t^*) = \beta_\mu(t^*) + A_i\beta_A(t^*) + X_i^T\psi_X + Z_i(t^*)^T b_i + \epsilon_i(t^*),$$

where $Z_i(t^*)$ is the design matrix for the random effect b_i which could be a vector including random intercept and slopes. Correlations between b_i and U_i can be allowed to account for the dependence between the longitudinal outcome and the events (i.e., death and dropout). However, this will substantially increase the already-heavy computational burden due to the added integrations with respect to the different random effects. Note that we are also not treating the piecewise survival submodel as semiparametric because of the computational burden that would be added by selecting the optimal breakpoints for the survival submodel. In theory, the piecewise survival submodel can also be semiparametric if the breakpoints for the pieces are selected similarly as how the knots are selected for the longitudinal submodel.

Our model can be used for a variety of QOL measures. In addition to FACIT-pal, common QOL measures such as EORTC-QLQ-C30, McGill QOL questionnaire and EQ-5D can be modeled similarly. Furthermore, other continuous measures such as Edmonton Symptom Assessment Scale (Hui and Bruera, 2017), Karnofsky Performance Scale (Mor and others, 1984), and HADS Depression and Anxiety Scales (Snaith, 2003) can be also modeled. Notice that although our model can accommodate a broad range of patterns for the longitudinal trajectories by employing the linear splines, it assumes continuous longitudinal trajectories for the outcome variables. In the presence of sudden changes in the outcomes (e.g., for heart failure patients) which may cause noncontinuous

longitudinal trajectories, the model may not provide optimal results in which case wavelet-based models (Morris and Carroll, 2006) could be adopted for the analysis.

SOFTWARE

Software in the form of R code, together with a sample input data set and complete documentation is available on Github: <https://github.com/quranwu/RetroJM>.

SUPPLEMENTARY MATERIAL

Supplementary material is available at <http://biostatistics.oxfordjournals.org>.

FUNDING

This work was supported by National Institutes of Health (R01GM123014, R01NR011871, R01HL158963, R01CA530191).

REFERENCES

- AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723.
- BAKITAS, M., LYONS, K. D., HEGEL, M. T., BALAN, S., BROKAW, F. C., SEVILLE, J., HULL, J. G., LI, Z., TOSTESON, T. D., BYOCK, I. R. and others. (2009). Effects of a palliative care intervention on clinical outcomes in patients with advanced cancer: the project ENABLE II randomized controlled trial. *JAMA* **302**, 741–749.
- BAKITAS, M. A., TOSTESON, T. D., LI, Z., LYONS, K. D., HULL, J. G., LI, Z., DIONNE-ODOM, J. N., FROST, J., DRAGNEV, K. H., HEGEL, M. T., AZUERO, A. and others. (2015). Early versus delayed initiation of concurrent palliative oncology care: patient outcomes in the ENABLE III randomized controlled trial. *Journal of Clinical Oncology* **33**, 1438–1445.
- CHAN, K. C. G. AND WANG, M.-C. (2010). Backward estimation of stochastic processes with failure events as time origins. *Annals of Applied Statistics* **4**, 1602–1620.
- CHAN, K. C. G. AND WANG, M.-C. (2017). Semiparametric modeling and estimation of the terminal behavior of recurrent marker processes before failure events. *Journal of the American Statistical Association* **112**, 351–362.
- CREEMERS, A., HENS, N., AERTS, M., MOLENBERGHS, G., VERBEKE, G. AND KENWARD, M. G. (2010). A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal* **52**, 111–125.
- DANIELS, M. J. AND HOGAN, J. W. (2008). *Missing Data in Longitudinal Studies*, Volume 109, Monographs on Statistics and Applied Probability. Boca Raton, FL: Chapman & Hall/CRC. Strategies for Bayesian modeling and sensitivity analysis.
- DEMPSEY, W. AND MCCULLAGH, P. (2018). Survival models and health sequences. *Lifetime Data Analysis* **24**, 550–584.
- DUMANOVSKY, T., AUGUSTIN, R., ROGERS, M., LETTANG, K., MEIER, D. E. AND MORRISON, R. S. (2016). The growth of palliative care in U.S. hospitals: a status report. *Journal of Palliative Medicine* **19**, 8–15.
- EL-JAWAHRI, A., TRAEGER, L., VANDUSEN, H., GREER, J. A., JACKSON, V. A., PIRL, W. F., TELLES, J., FISHMAN, S., RHODES, A., SPITZER, T. R., MCAFEE, S. L., CHEN, Y.-B. A. and others. (2017). Effect of inpatient palliative care during hematopoietic stem cell transplantation (HCT) hospitalization on psychological distress at six months post-HCT. *Journal of Clinical Oncology* **35**, 10005.
- ELASHOFF, R. M., LI, G. AND LI, N. (2017). *Joint Modeling of Longitudinal and Time-to-Event Data*, Volume 151, Monographs on Statistics and Applied Probability. Boca Raton, FL: CRC Press.
- HAREL, O. AND SCHAFFER, J. L. (2009). Partial and latent ignorability in missing-data problems. *Biometrika* **96**, 37–50.
- HENDERSON, R., DIGGLE, P. AND DOBSON, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics (Oxford, England)* **1**, 465–480.
- HOGAN, J. W. AND LAIRD, N. M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* **16**, 239–257.
- HUI, D. AND BRUERA, E. (2017). The edmonton symptom assessment system 25 years later: past, present, and future developments. *Journal of Pain and Symptom Management* **53**, 630–643.
- KELLEY, A. S. AND MORRISON, R. S. (2015). Palliative care for the seriously ill. *New England Journal of Medicine* **373**, 747–755.

- KONG, S., NAN, B., KALBFLEISCH, J. D., SARAN, R. AND HIRTH, R. (2018). Conditional modeling of longitudinal data with terminal event. *Journal of the American Statistical Association* **113**, 357–368.
- KURLAND, B. F. AND HEAGERTY, P. J. (2005). Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics (Oxford, England)* **6**, 241–258.
- KURLAND, B. F., JOHNSON, L. L., EGGLESTON, B. L. AND DIEHR, P. H. (2009). Longitudinal data with follow-up truncated by death: Match the analysis method to research aims. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* **24**, 211.
- LI, N., ELASHOFF, R. M., LI, G. AND SAVER, J. (2010). Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the ninds rt-PA stroke trial. *Statistics in Medicine* **29**, 546–557.
- LI, S., LI, N., WANG, H., ZHOU, J., ZHOU, H. AND LI, G. (2022). Efficient algorithms and implementation of a semiparametric joint model for longitudinal and competing risk data: with applications to massive biobank data. *Computational and Mathematical Methods in Medicine*.
- LI, Z., FROST, H. R., TOSTESON, T. D., ZHAO, L., LIU, L., LYONS, K., CHEN, H., COLE, B., CURROW, D. AND BAKITAS, M. (2017). A semiparametric joint model for terminal trend of quality of life and survival in palliative care research. *Statistics in Medicine* **36**, 4692–4704.
- LI, Z., TOSTESON, T. D. AND BAKITAS, M. A. (2013). Joint modeling quality of life and survival using a terminal decline model in palliative care studies. *Statistics in Medicine* **32**, 1394–1406.
- LIU, L., WOLFE, R. A. AND KALBFLEISCH, J. D. (2007). A shared random effects model for censored medical costs and mortality. *Statistics in Medicine* **26**, 139–155.
- MOR, V., LALIBERTE, L., MORRIS, J. N. AND WIEMANN, M. (1984). The Karnofsky Performance Status Scale: an examination of its reliability and validity in a research setting. *Cancer* **53**, 2002–2007.
- MORRIS, J. S. AND CARROLL, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**, 179–199.
- PARIKH, R. B., KIRCH, R. A., SMITH, T. J. AND TEMEL, J. S. (2013). Early specialty palliative care—translating data in oncology into practice. *The New England Journal of Medicine* **369**, 2347–2351.
- PUTTER, H., FIOCCO, M. AND GESKUS, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine* **26**, 2389–2430.
- RIZOPOULOS, D. (2012). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics & Data Analysis* **56**, 491–501.
- RIZOPOULOS, D. AND LESAFFRE, E. (2014). Introduction to the special issue on joint modelling techniques. *Statistical Methods in Medical Research* **23**, 3–10.
- SNAITH, R. P. (2003). The hospital anxiety and depression scale. *Health and Quality of Life Outcomes* **1**, 1–4.
- SU, L. AND HOGAN, J. W. (2010). Varying-coefficient models for longitudinal processes with continuous-time informative dropout. *Biostatistics (Oxford, England)* **11**, 93–110.
- SU, L., LI, Q., BARRETT, J. K. AND DANIELS, M. J. (2019). A sensitivity analysis approach for informative dropout using shared parameter models. *Biometrics* **75**, 917–926.
- TAKAHASHI, H. AND MORI, M. (1973). Double exponential formulas for numerical integration. *Publications of the Research Institute for Mathematical Sciences* **9**, 721–741.
- TEMEL, J. S., GREER, J. A., MUZIKANSKY, A., GALLAGHER, E. R., ADMANE, S., JACKSON, V. A., DAHLIN, C. M., BLINDERMAN, C. D., JACOBSEN, J., PIRL, W. F., BILLINGS, J. A. and others. (2010). Early palliative care for patients with metastatic non-small-cell lung cancer. *The New England Journal of Medicine* **363**, 733–742.
- TSIATIS, A. A. AND DAVIDIAN, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* **14**, 809–834.
- WU, H. AND ZHANG, J. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis*. Wiley.
- WULFSOHN, M. S. AND TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330–339.