

# Semi-parametric time-to-event modelling of lengths of hospital stays

Yang Li<sup>1</sup>  | Hao Liu<sup>2</sup> | Xiaoshen Wang<sup>3</sup> | Wanzhu Tu<sup>1</sup> 

<sup>1</sup>Department of Biostatistics and Health Data Science, Indiana University, Indianapolis, Indiana, USA

<sup>2</sup>Department of Biostatistics and Epidemiology, Rutgers School of Public Health, Piscataway, New Jersey, USA

<sup>3</sup>Department of Mathematics and Statistics, University of Arkansas at Little Rock, Little Rock, Arkansas, USA

## Correspondence

Yang Li, Department of Biostatistics and Health Data Science, Indiana University, 410 W. 10th St. Suite 3000, Indianapolis, IN 46202, USA.  
Email: [yli18@iu.edu](mailto:yli18@iu.edu)

## Funding information

Indiana Clinical and Translational Sciences Institute, Grant/Award Number: UL1TR002529; National Institutes of Health, Grant/Award Number: U24 AA026969

## Abstract

Length of stay (LOS) is an essential metric for the quality of hospital care. Published works on LOS analysis have primarily focused on skewed LOS distributions and the influences of patient diagnostic characteristics. Few authors have considered the events that terminate a hospital stay: Both successful discharge and death could end a hospital stay but with completely different implications. Modelling the time to the first occurrence of discharge or death obscures the true nature of LOS. In this research, we propose a structure that simultaneously models the probabilities of discharge and death. The model has a flexible formulation that accounts for both additive and multiplicative effects of factors influencing the occurrence of death and discharge. We present asymptotic properties of the parameter estimates so that valid inference can be performed for the parametric as well as nonparametric model components. Simulation studies confirmed the good finite-sample performance of the proposed method. As the research is motivated by practical issues encountered in LOS analysis, we analysed data from two real clinical studies to showcase the general applicability of the proposed model.

## KEYWORDS

additive-multiplicative hazards models, length of hospital stay, semi-parametric joint modelling, survival probability

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* published by John Wiley & Sons Ltd on behalf of Royal Statistical Society.

# 1 | INTRODUCTION

Length of hospital stay (LOS), that is, the number of inpatient days, is an important indicator of patient's health and care status. The concept is extensively used in clinical practice and health services research for quantification of resource consumption and quality of care (Chima et al., 1997; Freitas et al., 2012). Prolonged hospitalisation adds care cost, reduces operational efficiency and increases patient stress and risks for hospital-acquired infections (Clarke & Rosen, 2001; Graves et al., 2005). Reducing LOS has become one of the generally accepted policy goals of health care organisations (Pearson et al., 2001). In the United States, Centers for Medicare and Medicaid Services (CMS) use a prospective payment system that reimburses care cost by procedures performed, regardless of hospital days (CMS.gov, 2021), thus creating a great incentive for hospitals to reduce LOS.

Data analyses involving LOS are abundant in health services research. A simple Google Scholar search of 'hospital length of stay analysis' yields more than 2.1 million published items. Published analyses of LOS data tended to focus on skewed LOS distributions and employed common statistical models such as logistic regression and time-to-event analysis (Faddy et al., 2009; Marazzi et al., 1998; Vekaria et al., 2021; von Cube et al., 2017). Most do not differentiate the events that terminate a hospital stay. Discharge and death are treated the same, whichever comes first defines the LOS, despite the obvious clinical differences between the two. The situation resembles that described by various competing risk or semi-competing risk models, where the former depicts the occurrence of the *first* event by cause-specific hazards (Poguntke et al., 2018; Vekaria et al., 2021), and the latter quantifies the transition rates among multiple states with pre-defined subject-specific frailties or copula-based correlation structures (Hsieh et al., 2008; Lee et al., 2015; Peng & Fine, 2007; Xu et al., 2010).

With LOS data, analysts face two challenges: (1) Estimating the hazard function of discharge among survivors in the presence of censoring; (2) accommodating independent variables that are not adequately modelled by the Cox regression model. We present a model that addresses both challenges. Specifically, we describe a structure that simultaneously models the survival probabilities and discharge hazards, without pre-specifying the correlation structure between the two, by extending an approach as advocated by Sun et al. (2004, 2006), Zhao et al. (2011, 2013), Lee et al. (2015), and Wang et al. (2015). To enhance the model's flexibility, we allow independent variable effects to modify both the linear predictor and also the baseline hazard function, as suggested by Lin et al. (1995), Liu et al. (2010), Martinussen and Scheike (2007), and Sarker et al. (2015). To illustrate the proposed method, we analyse data from two clinical applications, one is a randomised clinical trial of anti-thrombotic therapies in patients suffering from stroke, and the other is a group of patients hospitalised for treatment of COVID.

The rest of this paper is organised as follows: In Section 2, we introduce the model, describe the estimation and inference procedures, as well as related theoretical results. In Section 3, we present simulation results. In Section 4, we use the proposed method to analyse data from two clinical studies. We conclude the paper in Section 5 with a few remarks.

## 2 | METHOD

### 2.1 | Model specification

Let  $T$  and  $U$  represent the event times for successful discharge and death, both are defined from the date of hospital admission. LOS is the time to  $T$  or  $U$ , whichever occurs first, that is,

$LOS = T \wedge U$  where  $a \wedge b = \min(a, b)$ . When  $LOS = U$ , in-hospital death occurs and the hazard of  $T$  is terminated. If  $LOS = T$  and the clinical follow-up continues,  $U$  could still be observed at a future time  $t > LOS$ . We write the counting processes for  $T$  and  $U$  as  $N^*(t) = I(T \leq t)$  and  $N^d(t) = I(U \leq t)$ , respectively, where  $I(\cdot)$  is the indicator function.

The conventional hazard function for  $T$  is

$$\lambda(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t | t \leq T)}{\Delta t} \tag{1}$$

For LOS analysis, the above hazard function is appropriate if  $T$  is subject only to censoring. But with death, the formulation becomes questionable because  $T$  cannot happen after  $U$ , that is,  $\lambda(t) = 0$  whenever  $t > U$ . This is a fact that (1) does not reflect.

To describe the hazard of discharge conditional on survival, we write the discharge intensity among survivors as  $E\{dN^*(t) | t \leq U\} = \lambda(t) | t \leq U dt$ , where

$$\lambda(t | t \leq U) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t | t \leq T, t \leq U)}{\Delta t} \tag{2}$$

represents the corresponding hazard function. The conditional hazard function in (2) is the same as the ‘cause-specific’ hazard discussed in the competing-risks literature (Prentice et al., 1978). However, unlike the the competing-risks models that focus on the occurrence of  $T$  or  $U$ , whichever occurs first, in LOS data,  $U$  could still happen after  $T$  if  $LOS = T$ .

To assess the effects of risk factors on  $\lambda(t | t \leq U)$ , we let  $X(t)$  and  $Z(t)$  be random covariate vectors of dimensions  $p$  and  $q$ , respectively. Letting  $\mathcal{X}(t) = \{X(s), 0 \leq s \leq t\}$ ,  $\mathcal{Z}(t) = \{Z(s), 0 \leq s \leq t\}$ , we present a general model as follows:

$$\lambda(t | \mathcal{X}(t), \mathcal{Z}(t), t \leq U) = [\lambda_0(t) + g\{\beta'X(t)\}] h\{\theta'Z(t)\} \tag{3}$$

where  $\lambda_0(t)$  is an unknown baseline hazard function,  $g(\cdot)$  and  $h(\cdot)$  are pre-specified twice differentiable functions representing the additive and multiplicative covariate effects, and  $\beta$  and  $\theta$  are the corresponding regression parameters. We assume  $h(x) > 0$  for any  $x \in \mathbb{R}$ , and  $g(0) = 0$ ,  $h(0) = 1$  to avoid identifiability issues. To reflect the fact that no live discharge is possible after death, we define  $\lambda(t | \mathcal{X}(t), \mathcal{Z}(t), t > U) = 0$ . In this model,  $X(t)$  modifies the baseline hazard function, whereas  $Z(t)$  influences the linear predictor. In practical data analysis, the composition of  $X(t)$  and  $Z(t)$  should be determined by scientific consideration and hypotheses of interest, as we shall demonstrate in the real data applications.

Model (3) differs from most of the existing models that depict time to the *first* occurrence of either  $T$  or  $U$  (Poguntke et al., 2018; Vekaria et al., 2021), and also from the semi-competing risk models (Lee et al., 2015; Peng & Fine, 2007; Xu et al., 2010) as Model (3) does not require specification of a correlation structure between  $T$  and  $U$ . The discharge hazards associated with the covariates are defined by (2), instead of by the marginal distribution of  $T$ , as done by Peng and Fine (2007) among others. Of note, the existing semi-competing risk models typically require copula-based correlations or shared frailties among successful discharge, in-hospital death, and death after discharge; estimation and inference, therefore, are susceptible to the risk of correlation misspecification. To avoid superimposing an arbitrary dependency structure or distributional assumptions, we extend an approach from event history analyses (Cheng et al., 2015; Sun et al., 2004, 2006; Zhao et al., 2011, 2013). Model (3) describes the conditional hazard

function of  $T$  given  $t \leq U$ , allowing for dependent  $T$  and  $U$  without requiring correlation structure specification.

For a special case of model (3), the survival function of  $T$  can involve a latent variable  $\nu$

$$Q(t|\nu, \mathcal{X}(t), \mathcal{Z}(t)) = \exp \left\{ - \int_0^t [\lambda_0(s) + g\{\beta'X(s)\}] h\{\theta'Z(s)\} ds \right\} \omega(\nu, t),$$

where  $\nu$  depends on  $U$ , and  $\omega(\cdot)$  satisfies  $E[\omega(\nu, t)|t \leq U] = 1$ . Other correlation structures between  $T$  and  $U$  are also possible but the marginal survival function of  $T$  would take a different form. In addition to its flexibility in accommodating correlation structures, model (3) provides a desired conditional hazard interpretation, which is not always ascertainable from either the copula (Peng & Fine, 2007) or the frailty semi-competing risk models (Lee et al., 2015; Sun & Shen, 2009; Wang et al., 2015; Zhang et al., 2005, 2007).

The simultaneous inclusion of  $g(\cdot)$  and  $h(\cdot)$  creates a structure that covers a broader class of models. Choices of  $g(\cdot)$ ,  $h(\cdot)$ ,  $X(t)$ , and  $Z(t)$  can be made based on investigators understanding of the variables' influences, and the intended interpretations of the resulting parameters. For example, when  $g(\cdot) = 0$  or  $h(\cdot) = 1$ , model (3) is reduced to the Cox proportional hazards model or the Aalen additive hazards model, respectively (Sun et al., 2004).

Model (3) can be used to assess proportional hazards across different levels of  $Z(t)$ ; within each level, the baseline hazard rate can be further modified by  $X(t)$ , whose effects may deviate from the exponential form. This structure lends enhanced flexibility to parameter interpretation in the LOS analyses. For example, including the comorbidity and treatment indicators in  $Z(t)$  and continuous characteristics (such as age) in  $X(t)$  gives the comorbidity/treatment indicators' hazard ratio interpretations at a given age. Similarly, for subjects of the same comorbidity profile,  $\beta$  represents an additive effect associated with per unit change of age on the baseline hazard, as one year change in age typically does not alter the hazard exponentially. More generally, the model depicts the proportional hazard of  $Z(t)$  at different levels of  $X(t)$ . In particular, when  $g(x) = x$  and  $h(x) = \exp(x)$ , for one-dimensional covariates  $X(t)$  and  $Z(t)$ , one unit change in  $Z(t)$  leads to the interpretation of  $\theta$  as a hazard ratio:

$$\frac{\lambda(t|U \geq t, X(t), Z(t) + 1)}{\lambda(t|U \geq t, X(t), Z(t))} = \exp(\theta).$$

Similarly, one unit change in  $X(t)$  while keeping  $Z(t)$  constant gives  $\beta \exp(\theta'Z(t))$  a covariate-dependent hazard difference interpretation:

$$\lambda(t|U \geq t, X(t) + 1, Z(t)) - \lambda(t|U \geq t, X(t), Z(t)) = \beta \exp(\theta'Z(t)).$$

In the absence of death or when death functions as a noninformative censoring event, model (3) is reduced to the Cox-Aalen regression model (Scheike & Zhang, 2002), except for the additive term  $\beta'X(t)$ . But as we shall demonstrate in the simulation study, unbiased inference cannot be derived from the Cox-Aalen regression by treating death time as noninformative censoring, even with the time-varying covariate effects averaged out over time.

The terminal event time  $U$  can be modelled by a standard Cox model:

$$\lambda^d(t|\mathcal{X}(t), \mathcal{Z}(t)) = \lambda^d(t) \exp\{\gamma'W(t)\}, \quad (4)$$

where  $\lambda^d(t)$  is the baseline hazard function,  $\gamma$  is a vector of regression coefficients for the covariates  $W(t) \subseteq \{X(t), Z(t)\}$ .

Here, we let  $C$  be a non-informative censoring time that is independent of both  $T$  and  $U$  given  $\mathcal{X}(t)$  and  $\mathcal{Z}(t)$ . Letting  $T_1 = U \wedge T \wedge C$ ,  $\delta_1 = I(T \leq (U \wedge C))$ ,  $T_2 = U \wedge C$ ,  $\delta_2 = I(U \leq C)$ , we write  $n$  independent and identically distributed quadruplets as  $\{(\delta_{1i}, T_{1i}, \delta_{2i}, T_{2i}), i = 1, \dots, n\}$ . Let  $N_i(t) = N_i^*(t \wedge C_i)$  be a counting process that jumps by one at  $t$  if and only if  $T_i = t$  and  $t \leq (U_i \wedge C_i)$ . The survival information for the  $i$ th subject is contained in  $\{\delta_{2i}, T_{2i}\}$ , regardless of  $T_{1i}$  and whether  $U_i$  and  $T_i$  are dependent.

## 2.2 | Estimation and inference

For narrative convenience, we discuss estimation and inference of the proposed model when  $g(x) = x$  and  $h(x) = \exp(x)$ , while noting that the estimation and inference procedures described below are readily extendable to other forms of  $g(\cdot)$  and  $h(\cdot)$ .

### 2.2.1 | Estimation

To estimate the unknown parameters associated with model (3), we define

$$dM_i(t) = \phi_i(t) [dN_i(t) \exp\{-\theta'Z_i(t)\} - \{\beta'X_i(t) + \lambda_0(t)\}dt],$$

where  $\phi_i(t) = \frac{I(t \leq T_{1i})}{S_i^d(t)}$  and  $S_i^d(t)$  represents the subject-specific survival function, that is,

$$S_i^d(t) = P(t \leq U_i | \mathcal{X}_i(t), \mathcal{Z}_i(t)).$$

Under model (3), it is straightforward to show  $M_i(t)$ 's are zero-mean stochastic processes. Hence one could estimate  $\beta$ ,  $\theta$ , and  $\lambda_0(t)$  by solving the following estimating equations:

$$\sum_{i=1}^n \int_0^t \phi_i(s) [ \exp\{-\theta'Z_i(s)\} dN_i(s) - \{\beta'X_i(s) + \lambda_0(s)\} ds ] = 0, \tag{5}$$

and

$$\sum_{i=1}^n \int_0^\tau V_i(t) \phi_i(t) [ \exp\{-\theta'Z_i(t)\} dN_i(t) - \{\beta'X_i(t) + \lambda_0(t)\} dt ] = 0, \tag{6}$$

where  $V_i(t) = (X_i'(t), Z_i'(t))'$  and  $\tau$  a pre-specified value representing a known finite maximum duration of the study. Combining Equations (5) and (6), estimates of  $\beta$  and  $\theta$  can be obtained by solving

$$\sum_{i=1}^n \int_0^\tau \{V_i(t) - \bar{V}(t)\} \phi_i(t) [ \exp\{-\theta'Z_i(t)\} dN_i(t) - \beta'X_i(t) dt ] = 0, \tag{7}$$

where

$$\bar{X}(t) = \frac{\sum_{i=1}^n X_i(t)\phi_i(t)}{\sum_{i=1}^n \phi_i(t)}, \quad \bar{Z}(t) = \frac{\sum_{i=1}^n Z_i(t)\phi_i(t)}{\sum_{i=1}^n \phi_i(t)}, \quad \bar{V}(t) = (\bar{X}'(t), \bar{Z}'(t))'$$

Let  $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$  be the cumulative baseline hazard function. When  $\beta = \hat{\beta}$  and  $\theta = \hat{\theta}$ ,  $\Lambda_0(t)$  can be estimated by

$$\hat{\Lambda}(t; \hat{\beta}, \hat{\theta}) = \int_0^t \frac{\sum_{i=1}^n \phi_i(s) \left[ e^{-\hat{\theta}'Z_i(s)} dN_i(s) - \hat{\beta}'X_i(s)ds \right]}{\sum_{i=1}^n \phi_i(s)}, \tag{8}$$

based on Equation (5).

We note that corresponding to model (3), the estimation involves  $S_i^d(t)$ , which is in general unknown. On the other hand, it can be estimated under model (4) with right-censored survival data from the study cohort. Specifically,  $S_i^d(t)$  can be estimated by  $\widehat{S}_i(t)$ , where

$$\widehat{S}_i(t) = \exp \left\{ - \int_0^t \exp \{ \hat{\gamma}'_n W_i(s) \} d\hat{\Delta}_0(s) \right\},$$

where  $d\hat{\Delta}_0(t) = \frac{\sum_{i=1}^n I(t \leq T_{2i}) dN_i^d(t)}{\sum_{i=1}^n I(t \leq T_{2i}) \exp \{ \hat{\gamma}'_n W_i(t) \}}$ ,  $\hat{\gamma}_n$  is the solution to the equation

$$\sum_{i=1}^n \int_0^\tau I(t \leq T_{2i}) \{ W_i(t) - \bar{W}(t; \gamma) \} dN_i^d(t) = 0, \tag{9}$$

$\bar{W}(t; \gamma) = \frac{R^{(1)}(t; \gamma)}{R^{(0)}(t; \gamma)}$ ,  $R^{(k)}(t; \gamma) = \frac{1}{n} \sum_{i=1}^n I(t \leq T_{2i}) e^{\gamma' W_i(t)} W_i^{\otimes k}$ ,  $a^{\otimes 0} = 1$ ,  $a^{\otimes 1} = a$ ,  $a^{\otimes 2} = aa'$  for any vector  $a$  and  $k = 0, 1, 2$ .

We denote the true values of  $\beta$  and  $\theta$  by  $\beta_0$  and  $\theta_0$ , and their estimators by  $\hat{\beta}_n$  and  $\hat{\theta}_n$ , solved from (7) with  $S_i^d(t)$  replaced by  $\widehat{S}_i(t)$ . Let  $\zeta = (\beta', \theta')'$ ,  $\zeta_0 = (\beta'_0, \theta'_0)'$ ,  $\hat{\zeta}_n = (\hat{\beta}'_n, \hat{\theta}'_n)'$  and  $\hat{\phi}_i(t) = \frac{I(t \leq T_{1i})}{\widehat{S}_i(t)}$ . The estimate of  $\Lambda_0(t)$  given by Equation (8) is now updated by

$$\hat{\Lambda}(t) = \int_0^t \frac{\sum_{i=1}^n \hat{\phi}_i(s) \left[ e^{-\hat{\theta}'_n Z_i(s)} dN_i(s) - \hat{\beta}'_n X_i(s)ds \right]}{\sum_{i=1}^n \hat{\phi}_i(s)}. \tag{10}$$

Since the baseline cumulative hazard estimator  $\hat{\Lambda}(t)$  is not necessarily non-decreasing in  $t$ . Without altering the asymptotic properties of  $\hat{\Lambda}(t)$ , we modify the estimator as  $\tilde{\Lambda}^*(t) = \max_{s \leq t} \hat{\Lambda}(s)$ . Following arguments similar to Lin et al. (1994), we show that  $\tilde{\Lambda}^*(t)$  and  $\hat{\Lambda}(t)$  are asymptotically equivalent as  $\tilde{\Lambda}^*(t) - \hat{\Lambda}(t) = o_p(n^{-1/2})$ .

### 2.2.2 | Prediction

Survival and discharge probabilities are both important clinical endpoints. Patient-specific trajectories of predictive probabilities are useful in aiding hospital management and patient care planning (Li et al., 2021; Li & Cheng, 2016; Sun et al., 2006). Although prediction is not the focus of the current research, we note that the existing semi-competing risk models with

shared frailty terms do not lend themselves to a straightforward predictive probability interpretation unless the frailty variables are integrated out over pre-specified prior distributions (Lee et al., 2015). From models (3) and (4), predictive probabilities can be ascertained in a straightforward fashion. More specifically, for a patient with covariates  $X = x, Z = z$ , the probability of death can be calculated from  $\widehat{G}(t|x, z) = 1 - \widehat{S}(t|x, z)$ . Suppose the patient is alive at time  $t$ , the cumulative hazard that the patient has a successful discharge at time  $T = t$  can be estimated from:

$$\widehat{H}(t|x, z) = \int_0^t \left[ \exp \{ \widehat{\theta}' z(s) \} d\widetilde{\Lambda}^*(s) + \widehat{\beta}' x(s) \exp \{ \widehat{\theta}' z(s) \} ds \right], \tag{11}$$

where  $\widehat{S}(t|x, z), \widetilde{\Lambda}^*(t), \widehat{\beta}, \widehat{\theta}$  are the estimators obtained from Section 2.2.1. The predicted cumulative incidence probability is then given by  $\widehat{F}(t|x, z) = 1 - \exp \{ -\widehat{H}(t|x, z) \}$ .

### 2.2.3 | Asymptotic properties

The following theorems state the asymptotic properties of  $\widehat{\zeta}_n$  and  $\widehat{\Lambda}(t)$ .

**Theorem 1.** *Under regularity conditions C1–C6 in the Appendix,  $\widehat{\beta}_n$  and  $\widehat{\theta}_n$  are consistent estimators of  $\beta_0$  and  $\theta_0$ , respectively. Additionally,  $n^{1/2}(\widehat{\zeta}_n - \zeta_0)$  converges weakly to a zero-mean normal distribution with a covariance matrix  $\Sigma$  consistently estimated by  $\widehat{\Sigma} = \widehat{A}^{-1} \widehat{\Sigma}^* \widehat{A}^{-1}$ , where*

$$\begin{aligned} \widehat{\Sigma}^* &= \frac{1}{n} \sum_{i=1}^n (\widehat{u}_{1i} + \widehat{u}_{2i})^{\otimes 2}, \\ \widehat{A} &= \begin{pmatrix} \widehat{A}_{11} & \widehat{A}_{12} \\ \widehat{A}_{21} & \widehat{A}_{22} \end{pmatrix}, \\ \widehat{A}_{11} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau (X_i(t) - \overline{X}(t)) \widehat{\phi}_i(t) X_i'(t) dt, \\ \widehat{A}_{12} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau (X_i(t) - \overline{X}(t)) \widehat{\phi}_i(t) e^{-\widehat{\theta}'_n Z_i(t)} Z_i'(t) dN_i(t), \\ \widehat{A}_{21} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau (Z_i(t) - \overline{Z}(t)) \widehat{\phi}_i(t) X_i'(t) dt, \\ \widehat{A}_{22} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau (Z_i(t) - \overline{Z}(t)) \widehat{\phi}_i(t) e^{-\widehat{\theta}'_n Z_i(t)} Z_i'(t) dN_i(t), \\ \widehat{u}_{1i} &= \int_0^\tau (V_i(t) - \overline{V}(t)) d\widehat{M}_i(t), \\ d\widehat{M}_i(t) &= \widehat{\phi}_i(t) [e^{-\widehat{\theta}'_n Z_i(t)} dN_i(t) - \{ \widehat{\beta}'_n X_i(t) + \widehat{\lambda}_0(t) \} dt], \\ \widehat{u}_{2i} &= \int_0^\tau \left[ \frac{\widehat{\Psi}(u)}{R^{(0)}(u)} + \widehat{B} \left\{ W_i(u) - \frac{R^{(1)}(u; \widehat{\gamma}_n)}{R^{(0)}(u; \widehat{\gamma}_n)} \right\} \right] d\widehat{M}_i^d(u), \\ \widehat{\Psi}(u) &= \frac{1}{n} \sum_{i=1}^n e^{\widehat{\gamma}'_n W_i(u)} \int_0^\tau I(t > u) (V_i(t) - \overline{V}(t; \widehat{\gamma}_n)) d\widehat{M}_i(t), \end{aligned}$$

$$\begin{aligned}
 d\hat{M}_i^d(t) &= dN_i^d(t) - I(t \leq T_{2i})e^{\hat{\gamma}'_n W_i(t)} d\hat{\Delta}_0(t), \quad d\hat{\Delta}_0(t) = \frac{\sum_{i=1}^n dN_i^d(t)}{\sum_{i=1}^n I(t \leq T_{2i})e^{\hat{\gamma}'_n W_i(t)}}, \\
 \hat{B} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left( V_i(t) - \bar{V}(t; \hat{\gamma}_n) \right) \hat{J}(t, W_i)' \Omega_\gamma^{-1} d\hat{M}_i(t), \\
 \hat{\Omega}_\gamma &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left[ \frac{R^{(2)}(t; \hat{\gamma}_n)}{R^{(0)}(t; \hat{\gamma}_n)} - \left( \frac{R^{(1)}(t; \hat{\gamma}_n)}{R^{(0)}(t; \hat{\gamma}_n)} \right)^{\otimes 2} \right] dN_i^d(t), \\
 \hat{J}(t; W_i) &= \int_0^t e^{\hat{\gamma}'_n W_i(u)} \left[ W_i(u) - \bar{W}(u; \hat{\gamma}_n) \right] d\hat{\Delta}_0(u).
 \end{aligned}$$

**Theorem 2.** Under regularity conditions C1–C6 in the Appendix,  $\hat{\Lambda}(t)$  converges in probability to  $\Lambda_0(t)$  uniformly in  $t \in [0, \tau]$  and  $\sqrt{n}\{\hat{\Lambda}(t) - \Lambda_0(t)\}$  converges weakly to a zero-mean Gaussian process with a covariance function that can be consistently estimated by  $\hat{Y}(s, t) = \frac{1}{n} \sum_{i=1}^n \{\hat{g}_i(s)\hat{g}_i^T(t)\}$  for  $s, t \in [0, \tau]$ , where

$$\begin{aligned}
 \hat{g}_i(t) &= \hat{\Gamma}(t)\hat{A}^{-1}(\hat{u}_{1i} + \hat{u}_{2i}) + \hat{u}_{3i}(t) + \hat{u}_{4i}(t) + \hat{u}_{5i}(t), \\
 \hat{\Gamma}(t) &= \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\hat{\phi}_i(s)}{\hat{\kappa}(s)} \left[ -X'_i(s)ds, -Z'_i(s)e^{-\hat{\theta}'_n Z_i(s)} dN_i(s) \right], \\
 \hat{u}_{3i}(t) &= \int_0^t \frac{d\hat{M}_i(s)}{\hat{\kappa}(s)}, \quad \hat{u}_{4i}(t) = \int_0^t \frac{\hat{G}(u, t)d\hat{M}_i^d(u)}{R^{(0)}(u; \hat{\gamma}_n)}, \\
 \hat{u}_{5i}(t) &= \hat{O}(t) \int_0^\tau \left\{ W_i(u) - \frac{R^{(1)}(u; \hat{\gamma}_n)}{R^{(0)}(u; \hat{\gamma}_n)} \right\} d\hat{M}_i^d(u), \quad \hat{\kappa}(s) = \frac{1}{n} \sum_{i=1}^n \hat{\phi}_i(s), \\
 \hat{G}(u, t) &= \frac{1}{n} \sum_{i=1}^n e^{\hat{\gamma}'_n W_i(u)} \int_0^t \frac{I(s > u)}{\hat{\kappa}(s)} d\hat{M}_i(s), \\
 \hat{O}(t) &= \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{1}{\hat{\kappa}(s)} \hat{J}(s, W_i)' \Omega_\gamma^{-1} d\hat{M}_i(s),
 \end{aligned}$$

besides other quantities defined in Theorem 1. In particular,  $\sqrt{n}\{\hat{\Lambda}(t) - \Lambda_0(t)\} \xrightarrow{D} N(0, \Sigma_\Lambda(t))$ , and the covariance  $\Sigma_\Lambda(t)$  can be estimated consistently by  $S_\Lambda(t) = n^{-1} \sum_{i=1}^n \{\hat{g}_i(t)\}^2$ . Theorems 1 and 2 imply that the SEs of  $\hat{\beta}_n, \hat{\theta}_n$  and  $\hat{\Lambda}_0(t)$  can be consistently estimated by  $\sqrt{S_\beta}, \sqrt{S_\theta}$  and  $\sqrt{S_\Lambda(t)}$ , respectively, where  $S_\beta, S_\theta$  are diagonal entries of  $\hat{\Sigma}$  in Theorem 1.

### 3 | SIMULATION STUDIES

We conducted an extensive simulation study to assess the finite sample performance of the proposed estimators  $\hat{\beta}_n, \hat{\theta}_n$  and  $\tilde{\Lambda}^*(t)$ . We evaluated the estimation performance under the sample sizes of  $n = 100$  and  $200$ , with 1000 replication for each parameter setting.

#### 3.1 | Data generation

We considered two major scenarios:



(i)  $T$  and  $U$  were independent.  $T_i$  for subject  $i$  was generated with the hazard function:

$$\lambda_i(t; X_i, Z_i) = (\sqrt[4]{t} + \beta X_i) \exp\{\theta Z_i\}, \tag{12}$$

where  $X_i$  and  $Z_i$  were generated from the uniform (0, 1) and Bernoulli distribution with a probability of success of 0.5, respectively.  $U_i$  was generated from the Cox model with covariates  $Z_i$  with a baseline hazard  $\lambda_0^d$ .

(ii)  $T$  and  $U$  were correlated. Besides sharing covariates  $X$  and  $Z$  generated in the same way as in Scenario (i), we involved in the distribution functions of  $T$  and  $U$  a latent variable  $v$  from the positive stable distribution with Laplace transformation  $L(s) = E[\exp\{-sv_i\}] = \exp\{-s^\rho\}$  for  $0 \leq \rho \leq 1$  for subject  $i$  (Luong & Doray, 2009). From  $v_i$  and  $Z_i$ ,  $U_i$  had a hazard function  $\lambda^d(t; Z_i, v_i) = 0.1v_i \exp\{\gamma Z_i\}$ . It can be shown that  $U_i$  satisfies model (4) with  $S(t; Z_i) = \exp\{-(0.1te^{\gamma Z_i})^\rho\}$ . The discharge time  $T_i$  was generated from survival function

$$Q(t; X_i, Z_i, v_i) = \exp\{-(1 + \beta X_i) \exp\{\theta Z_i\}t\} \omega(v_i, t), \tag{13}$$

where

$$\omega(v_i, t) = \exp\left(kv_it + t^\rho \left[\{\lambda_0^d e^{\gamma Z_i} - k\}^\rho - \{\lambda_0^d e^{\gamma Z_i}\}^\rho\right]\right),$$

$0 \leq k \leq \lambda_0^d$ . One can show that  $E[\omega(v_i, t)|Z_i, t \leq U_i] = 1$ . Hence among survivors,

$$S(t|X_i, Z_i, t \leq U_i) = \exp\left[-(1 + \beta X_i) \exp\{\theta Z_i\}t\right] \quad \text{and} \quad \lambda(t|X_i, Z_i, t \leq U_i) = (1 + \beta X_i) \exp\{\theta Z_i\},$$

which satisfies model (3). Different values of  $k$ ,  $\rho$  and  $\lambda_0^d$  yield event times  $T$  and  $U$  with various magnitudes of correlation. In all simulated settings,  $T$  and  $U$  were negatively correlated with the empirical estimates of the Spearman correlation coefficient ranged around  $-0.1$  to  $-0.3$  (Schemper et al., 2013). The censoring time  $C_i$  was taken to be a common value at  $\tau = 2$ , which represented the study duration.

### 3.2 | Metrics of evaluation

We calculated summaries of estimation bias (Bias), sample standard error (SSE), averaged standard error estimate (SEE) according to the theorems, and empirical 95% coverage probability (CP). When the estimators worked as the theorems dictated, the evaluation metrics were expected to show a bias approaching zero, strong agreement between SSE and SEE, and CP close to the 95% nominal level.

### 3.3 | Simulation results

The estimation results for  $\beta$  and  $\theta$  are in Tables 1 and 2 and Tables S1–S3, where Table 1 corresponds to settings of independent  $T$  and  $U$  in Scenario (i) with  $\lambda_0^d = 0.1$ , and Tables 2 and S1–S3 represent correlated  $T$  and  $U$ . The Spearman correlation coefficient between  $T$  and  $U$  ranged

**TABLE 1** Simulation results on  $\hat{\beta}_n$  and  $\hat{\theta}_n$  for Scenario (i) when  $T$  and  $U$  were generated independently

$n$	$\hat{\beta}_n$				$\hat{\theta}_n$			
	Bias	SEE	SSE	CP	Bias	SEE	SSE	CP
$(\beta, \theta, \gamma) = (0, 0, 0)$								
100	-0.016	0.341	0.358	0.944	-0.006	0.197	0.200	0.953
200	-0.015	0.238	0.237	0.955	0.008	0.138	0.143	0.948
$(\beta, \theta, \gamma) = (0.2, 0, 0)$								
100	0.028	0.374	0.391	0.945	-0.002	0.201	0.220	0.929
200	0.008	0.260	0.269	0.948	0.003	0.140	0.143	0.945
$(\beta, \theta, \gamma) = (0, 0.5, 0)$								
100	0.005	0.330	0.356	0.939	-0.023	0.214	0.225	0.929
200	-0.007	0.234	0.239	0.943	-0.032	0.151	0.155	0.936
$(\beta, \theta, \gamma) = (0.2, 0.5, 0)$								
100	0.012	0.361	0.397	0.923	-0.016	0.215	0.227	0.938
200	0.005	0.255	0.267	0.937	-0.026	0.152	0.155	0.947
$(\beta, \theta, \gamma) = (0, 0, 0.3)$								
100	-0.027	0.341	0.361	0.932	0.004	0.199	0.207	0.943
200	0.005	0.238	0.240	0.953	0.007	0.139	0.141	0.952
$(\beta, \theta, \gamma) = (0.2, 0, 0.3)$								
100	0.017	0.376	0.398	0.937	0.006	0.202	0.200	0.957
200	-0.004	0.261	0.254	0.942	-0.002	0.141	0.137	0.956
$(\beta, \theta, \gamma) = (0, 0.5, 0.3)$								
100	-0.007	0.333	0.356	0.922	-0.016	0.217	0.225	0.932
200	0.012	0.234	0.240	0.937	-0.031	0.152	0.157	0.940
$(\beta, \theta, \gamma) = (0.2, 0.5, 0.3)$								
100	0.029	0.362	0.389	0.932	-0.003	0.218	0.233	0.930
200	0.009	0.254	0.255	0.947	-0.010	0.153	0.152	0.954

Note: Summaries of biases, sample standard errors (SSE), standard error estimates (SEE) and coverage probabilities (CP) for  $n = 100$  and  $200$ , based on 1000 simulations.

around  $-0.1$  to  $-0.3$  as described in Scenario (ii). All proposed estimators performed well with small biases, values of SSEs and SEEs close to each other, and empirical CPs approached the nominal level. The SEEs decreased with an increasing sample size, showing improved estimation efficiency.

In addition to the parametric estimators, we also calculated the pointwise SSE, SEE and CP for the non-parametric estimator  $\tilde{\Lambda}^*(t)$ , which is asymptotically equivalent to  $\hat{\Lambda}(t)$  as defined in Equation (10). Observations showed a similar performance as with the parametric estimators. For example, Figure 1 shows the pointwise estimates on  $(0, \tau)$  when  $(\beta, \theta, \gamma) = (0, 0.5, 0)$ , as one set-up in Table 2. We noted that  $\tilde{\Lambda}^*(t)$  had generally accurate SEEs, and the empirical coverage probabilities fluctuated closely around the nominal-level. When the sample size increased, the estimates showed improved efficiency with reduced standard errors. To gauge the magnitude of

**TABLE 2** Simulation results on  $\hat{\beta}_n$  and  $\hat{\theta}_n$  for Scenario (ii) when correlated  $T$  and  $U$  were generated,  $k = 0.1$ ,  $\rho = 0.8$  and  $\lambda_0^d = 0.1$

$n$	$\hat{\beta}_n$				$\hat{\theta}_n$			
	Bias	SEE	SSE	CP	Bias	SEE	SSE	CP
$(\beta, \theta, \gamma) = (0, 0, 0)$								
100	0.014	0.401	0.419	0.939	-0.004	0.203	0.208	0.952
200	-0.011	0.281	0.284	0.950	0.002	0.142	0.145	0.942
$(\beta, \theta, \gamma) = (0.2, 0, 0)$								
100	0.026	0.437	0.453	0.941	-0.007	0.205	0.210	0.956
200	-0.008	0.306	0.311	0.950	0.001	0.144	0.148	0.945
$(\beta, \theta, \gamma) = (0, 0.5, 0)$								
100	0.009	0.400	0.424	0.930	-0.016	0.218	0.223	0.943
200	-0.010	0.283	0.288	0.939	-0.014	0.154	0.159	0.940
$(\beta, \theta, \gamma) = (0.2, 0.5, 0)$								
100	0.025	0.435	0.463	0.935	-0.010	0.218	0.222	0.947
200	-0.004	0.308	0.311	0.941	-0.006	0.154	0.158	0.936
$(\beta, \theta, \gamma) = (0, 0, 0.3)$								
100	0.015	0.405	0.421	0.940	-0.018	0.205	0.210	0.950
200	-0.010	0.284	0.287	0.952	-0.011	0.143	0.146	0.946
$(\beta, \theta, \gamma) = (0.2, 0, 0.3)$								
100	0.030	0.440	0.455	0.945	-0.021	0.207	0.213	0.951
200	-0.007	0.308	0.315	0.947	-0.011	0.145	0.150	0.943
$(\beta, \theta, \gamma) = (0, 0.5, 0.3)$								
100	0.012	0.402	0.425	0.935	-0.026	0.220	0.225	0.939
200	-0.009	0.284	0.290	0.943	-0.025	0.155	0.161	0.933
$(\beta, \theta, \gamma) = (0.2, 0.5, 0.3)$								
100	0.029	0.437	0.461	0.937	-0.019	0.219	0.223	0.948
200	-0.002	0.308	0.313	0.941	-0.016	0.155	0.160	0.933

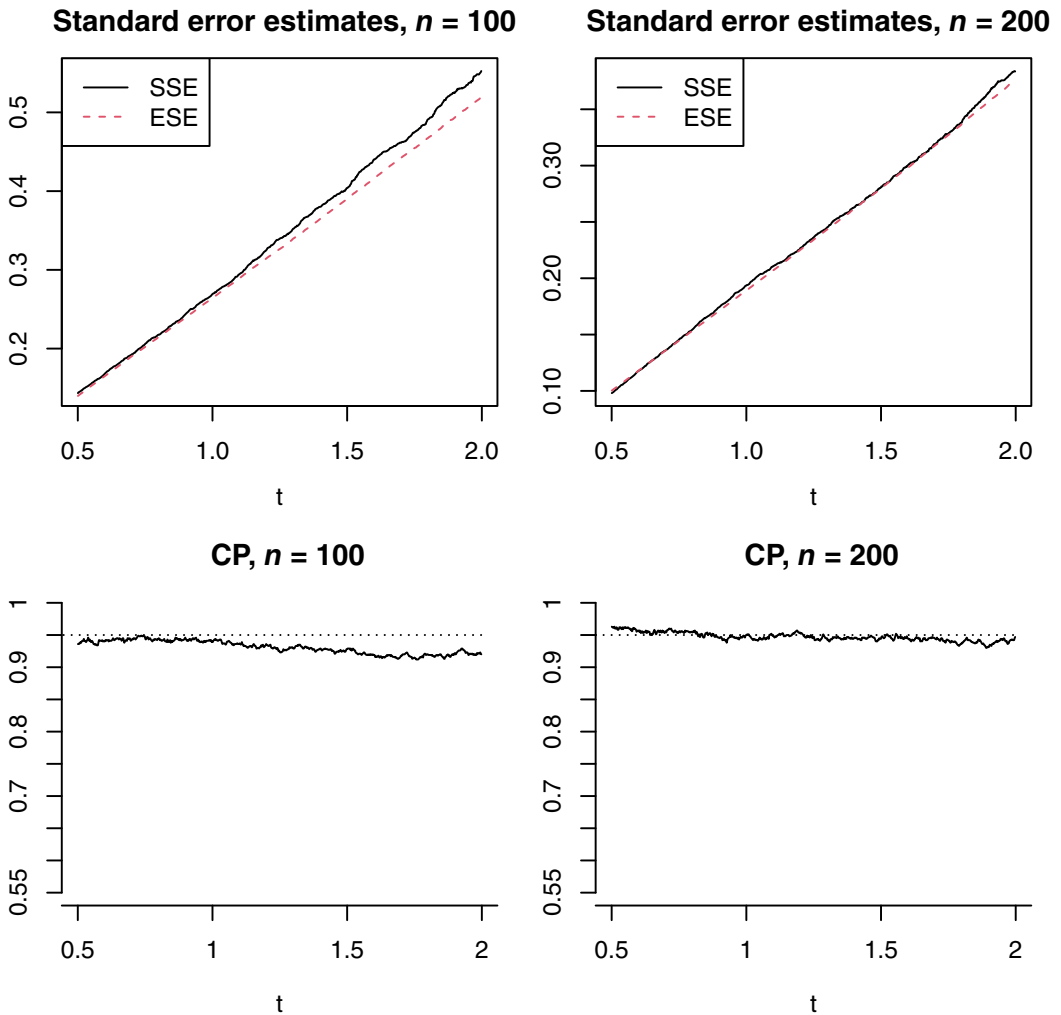
Note: Summaries of biases, sample standard errors (SSE), standard error estimates (SEE) and coverage probabilities (CP) for  $n = 100$  and  $200$ , based on 1000 simulations.

bias over time, we calculated the integrated squared bias (ISB) defined as

$$ISB = \sum_{j=0}^M (\overline{\tilde{\Lambda}^*}(t_j) - \Lambda_0(t_j))^2 \Delta t,$$

where  $t_0 = 0, t_1, \dots, t_M$  are equally spaced time points separated by  $\Delta t = \tau/M, M = 999$ ,

$$\overline{\tilde{\Lambda}^*}(t_j) = \frac{1}{1000} \sum_{m=1}^{1000} \tilde{\Lambda}^{*(m)}(t_j),$$



**FIGURE 1** Pointwise SE estimates and the coverage probabilities for Scenario (ii) when  $T$  and  $U$  are correlated,  $n = 100$  and  $200$ ,  $\beta = 0$ ,  $\theta = 0.5$ ,  $\gamma = 0$ , based on 1000 simulations. CP, coverage probabilities; SEE, standard error estimates; SSE, sample standard errors [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

and  $\tilde{\Lambda}^{*(m)}(t_j)$  is the  $m$ th replicate of  $\tilde{\Lambda}^*(t_j)$ . The ISB values were 0.0022 and 0.0019, respectively, when  $n = 100$  and  $n = 200$ . The results indicated the  $\tilde{\Lambda}^*(t)$ 's were unbiased.

We also note that when there were no deaths or when  $U$  was treated as non-informative censoring, Scheike and Zhang (2002) proposed the following model:

$$\lambda(t|X, Z) = [\lambda_0(t) + \beta(t)'X] \exp\{\theta'Z\}, \tag{14}$$

coupled with an estimation procedure for  $\theta$  and cumulative time-varying effects  $A(t)$ , where  $A(t) = \int_0^t \beta(s)ds$ . This model yields a parameter estimator  $\hat{\beta}^* = \hat{A}(L)/L$ , where  $L$  is the largest  $T$  observed. We refer to  $\hat{\beta}^*$  as the SZ2002 estimator, which is different from the estimator we proposed under the conditional hazard in (2). We presented simulation results showing that the SZ2002 estimator has considerably larger biases whereas our estimator performed well in both

**TABLE 3** Comparisons of the biases associated with the proposed  $\beta$  estimator and a modified SZ2002 estimator, for Scenarios (i) and (ii) when  $\gamma = 0.3$

	Proposed		Modified SZ2002	
	$(\beta, \theta) = (0, 0)$	$(\beta, \theta) = (0.2, 0.5)$	$(\beta, \theta) = (0, 0)$	$(\beta, \theta) = (0.2, 0.5)$
Scenario (i)				
$n = 100$	-0.027	0.029	0.820	1.241
$n = 200$	-0.005	0.009	-0.511	0.275
Scenario (ii)				
$n = 100$	0.015	0.029	-0.260	0.264
$n = 200$	-0.010	-0.002	-0.777	-0.625

simulation scenarios (i and ii). See Table 3. The results, therefore, confirm that treating deaths as noninformative censoring events leads to biased inference for model (3).

### 3.4 | Sensitivity analyses

Finally, we assessed the estimation robustness against covariate mis-specification by simulating various situations where  $W$  in model (4) is incorrectly specified. We generated correlated  $T$  and  $U$  for Scenario (ii) with  $k = 0.1$ ,  $\rho = 0.8$  and  $\lambda_0^d = 0.1$ ;  $Z \sim \text{Bernoulli}(p = 0.5)$ .  $W$  was misspecified in various ways: (i) Inclusion of an extra variable  $X \sim \text{Uniform}(0, 1)$ , that is, using  $W = (Z, X)'$  when the true  $W = Z$ . (ii) Inclusion of a wrong variable. We used  $W = (X_1, Z)'$  when the true  $W = (X_1, X_2)'$ , where  $X_1 \sim \text{Uniform}(0, 1)$  and  $X_2 \sim \text{Uniform}(1, 3)$ . (iii) Leaving out a variable in  $W$ . We used  $W = Z$  or  $W = X$  when the true  $W = (Z, X)'$ , where  $X \sim \text{Uniform}(1, 3)$ . Simulation results are presented in Tables S4 and S5. Briefly, estimates of  $\beta$  and  $\theta$  were not strongly influenced by the misspecification of  $W$ , probably because only the survival function estimate ( $\widehat{S}_i(t)$ ) but not  $\hat{\gamma}$  was needed in the estimating equations for  $\beta$  and  $\theta$ . For all tested settings, the performance appeared to improve with sample size.

## 4 | REAL DATA APPLICATIONS

The model proposed in the paper is general enough to be applicable to a broad range of LOS analyses. Here we present analyses of data from two clinical studies, each of which requires some of the features of the proposed model. As discussed in earlier sections, although existing frailty and copula-based semi-competing risk models can be used to analyse such data, those methods do not accommodate deaths that occurred after discharge, nor do they offer the conditional hazard interpretation, and often require pre-specification of correlation structures that are not easily verifiable in practical data analysis.

### 4.1 | LOS with the International Stroke Trial

We first considered the International Stroke Trial (IST), a randomised clinical trial of anti-thrombotic therapies in patients suffering from strokes (Group, 1997; Sandercock et al., 2011).

The study followed a factorial design, testing the effects of various combinations of aspirin and heparin at different dose levels on the survival of stroke patients. Treatment started upon enrollment and continued for 14 days or until hospital discharge or death, whichever came first. Among the 6415 trial participants whose medical images showed visible infarctions at randomisation, 1351 had died before discharge; 3159 had been successfully discharged, of which 296 died after discharge. The maximum follow-up length was 567 days, at which time patients without a discharge or death record were censored.

The IST trial has four treatment groups. Among the 6415 patients, 1662 were on aspirin, 1564 on heparin, 1675 on both, and 1514 on neither. In this secondary analysis, we focused on the time from treatment initiation to hospital discharge, with death as the terminal event. Indicators of treatment groups were the main independent variables of interest, with other patient characteristics such as age and conscious state at enrollment as relevant covariates (Kwah & Diong, 2014).

We first examined the treatment effects on death while adjusting for the influences of age and the baseline level of consciousness by using the following Cox model:

$$\lambda_i^d(t|W_i) = \lambda^d(t) \exp\{\gamma'W_i\}, \quad (15)$$

where  $\lambda_i^d$  represented the hazard of death. Because of the wide age range (16–98 years) of the trial participants, we categorised age by tertiles. A total of six covariates were included in  $W_i$ : Age  $\leq 67$ ,  $67 < \text{Age} \leq 77$ , full consciousness at randomisation (CONSC), aspirin alone (ASP), heparin alone (HEP), and both aspirin and heparin (Both). All independent variables were binary. Table 4 shows that being in the younger age groups and being fully alert at randomisation were associated with reduced hazards of death. Aspirin reduced the hazards of death slightly but its effect did not reach the level of statistical significance.

We then examined LOS with successful discharge by fitting the model below:

$$\lambda(t|X, Z, t \leq U) = [\lambda_0(t) + \beta'X] \exp\{\theta'Z\}. \quad (16)$$

We determined the composition of  $X$  and  $Z$  primarily based on the objective of the analysis, which is to determine the effects of the four treatments on LOS with successful discharge. We, therefore, included the treatment indicators in  $Z$  to retain the hazard ratio interpretation. Age and levels of consciousness were included in  $X$  as adjustment factors to the baseline hazard. Interestingly, previous studies suggested that age was a poor predictor of functional recovery

**TABLE 4** Results from Cox regression analysis containing independent variable effects on death hazards of the 6415 patients whose medical images showed visible infarctions at randomisation in International Stroke Trial

	EST (SE)	p-value	95%CI
Age $\leq 67$	-1.3028 (0.0748)	0.0000	(-1.4495, -1.1561)
$67 < \text{Age} \leq 77$	-0.5111 (0.0554)	0.0000	(-0.6195, -0.4024)
CONSC	-1.2831 (0.0495)	0.0000	(-1.3801, -1.1861)
ASP	-0.0270 (0.0696)	0.6978	(-0.1634, 0.1093)
HEP	0.0082 (0.0684)	0.9045	(-0.1258, 0.1422)
Both	-0.0258 (0.0699)	0.7124	(-0.16298, 0.1113)

Abbreviations: 95% CI, 95% confidence interval; EST, point estimate; SE, standard error estimate.

**TABLE 5** Results from the proposed model containing independent variable effects on length of stay among survivors from 6415 International Stroke Trial participants

	Proposed method			Scheike and Zhang (2002)		
	EST (SE)	p-value	95% CI	EST (SE)	p-value	95% CI
Age ≤ 67	0.0010 (0.0002)	0.0000	(0.0006, 0.0013)	—	—	—
67 < Age ≤ 77	0.0004 (0.0002)	0.0132	(0.0001, 0.0007)	—	—	—
CONSC	0.0012 (0.0001)	0.0000	(0.0009, 0.0015)	—	—	—
ASP	0.0993 (0.0534)	0.0631	(−0.0054, 0.2041)	0.1105 (0.0502)	0.0276	(0.0122, 0.2088)
HEP	0.0256 (0.0534)	0.6319	(−0.0790, 0.1301)	0.0282 (0.0360)	0.4336	(−0.0423, 0.0986)
Both	0.0099 (0.0540)	0.8542	(−0.0959, 0.1158)	0.0249 (0.0359)	0.4890	(−0.0456, 0.0953)

Notes: Age groups and conscious status were included in *X* as additive effects, and treatment groups included in *Z* as multiplicative effects. The left panel presents the results from the proposed model and the right panel from Scheike and Zhang’s (2002) model by treating deaths as noninformative censoring events.

Abbreviations: 95% CI, 95% confidence interval; EST, point estimate; SE, standard error estimate.

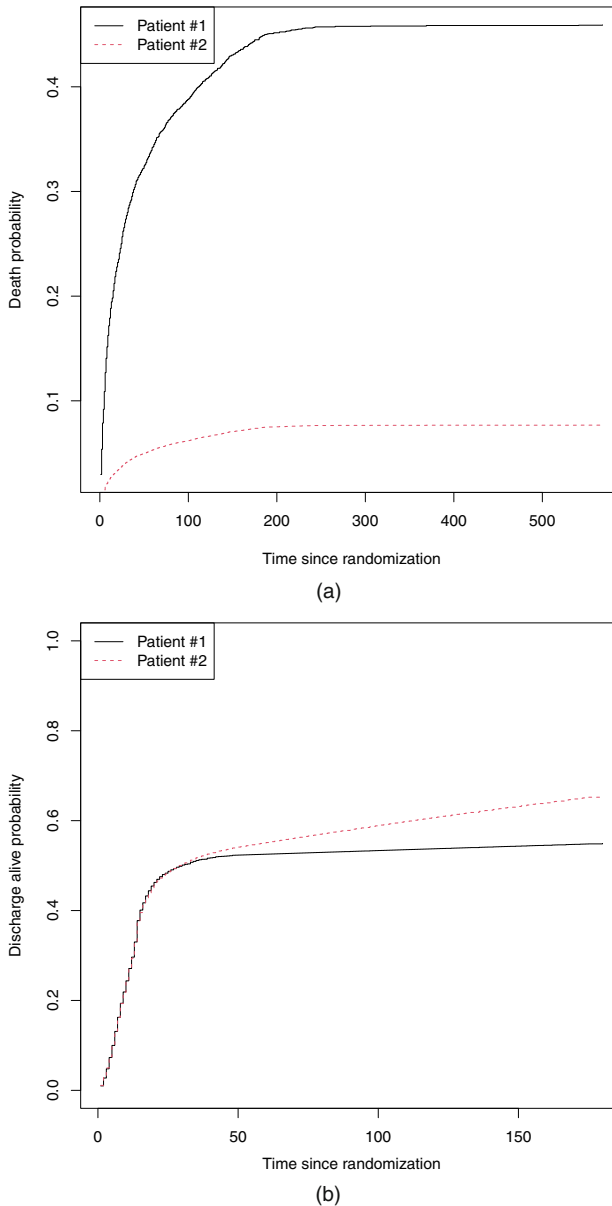
from stroke (Kugler et al., 2003) and levels of consciousness correlated with various stroke outcomes but their predictive power was not high (Li et al., 2020). The studies have not tested hypotheses whether age and conscientious state alter the baseline hazards. To test, we included them in *X* as modifiers of the baseline hazards. Analytical results supported the hypothesis that younger patients and those that had been fully alert at the randomisation had greater baseline hazards for a successful discharge (shorter LOS). For the treatment effects, the analysis showed that the average treatment effects of aspirin and heparin corresponded to shorter LOS, but the effects did not reach the level of significance. The findings were generally consistent with the findings of the primary paper, which did not find significant effects for either heparin or aspirin on mortality (Group, 1997). Detailed results are presented in the left panel of Table 5.

For comparison, we also analysed the LOS data by treating death as non-informative censoring by using the model described by Scheike and Zhang (2002):

$$\lambda(t|X, Z) = [\lambda_0(t) + \beta(t)'X] \exp\{\theta'Z\}, \tag{17}$$

where *X* and *Z* were defined similarly as in model (16). The cumulative coefficient estimates of  $A(t) = \int_0^t \beta(s)ds$  are presented in Figures S1 and S2, and the estimates of  $\theta$  elements in the right panel of Table 5. The analysis showed a significant effect of aspirin in reducing LOS. The finding, however, had a questionable validity because death was treated as non-informative censoring, and the hazards of discharge should be interpreted in the full sample which included people that had died. To characterise the hazards of discharge, it is more justified to use the conditional hazards in model (16) per definition (2). On the other hand, the hazards of discharge could be underestimated among survivors if death is ignored. This may explain why most positive estimates from the right panel are slightly larger than the corresponding estimates on the left.

As an illustration for prediction, we presented the trajectories of survival and successful discharge probabilities for patients with otherwise comparable health conditions and mortality rates as within the IST cohort. For example, for a patient of 69 years of age, drowsy at admission and assigned to aspirin (Patient #1), and another patient aged 63, fully alert at admission and assigned heparin (Patient #2), Figure 2 present their death and successful discharge probabilities within 90



**FIGURE 2** Predictive trajectories of (a) death probabilities and (b) successful discharge probabilities for observed International Stroke Trial covariates [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.com)]

days. It can be seen the two sets of different covariates lead to quite different trajectories, where Patient #2 had overall lower death risks and higher probabilities of successful discharge.

## 4.2 | LOS in patients with COVID

The second application is a study of LOS of patients hospitalised for COVID in a midwestern hospital. In this application, censoring happened when a patient remained in the hospital at the



time of data extraction. Among the patients that were not censored, six died with survival times ranging from 0.8 to 30.5 days since admission; 480 survived and were discharged within 20 days. The maximum length of the observational window was 31 days.

In this application, LOS was calculated from the date of admission to discharge or death, whichever came first. No patients were followed up after LOS. Patient information was ascertained at the time of admission, as recorded by an electronic health record system: Age (AGE), smoking status (SMO), and a set of pre-specified medical conditions associated with COVID-related disease severity (CDC, 2020), including dyspnea (DYS), diabetes (DIA), hypertension (HYP), chronic obstructive pulmonary disease (COPD), asthma (AST), cancer (CANC), chronic kidney disease (CKD) and congestive heart failure (CHF). In this analysis, we assessed the influences of these comorbidities on successful discharge with death as the terminal event. For the convenience of interpretation, we characterised AGE by decade (ranged from 0 to 9). Binary indicators were created for all comorbidities. Because of the low number of mortality in the data set, we used the following model to screen the potential correlates of death:

$$\lambda_i^d(t|W_i) = \lambda^d(t) \exp\{\gamma'W_i\}. \tag{18}$$

Our analysis showed age was the only significant factor associated with increased hazard of death ( $\exp(\hat{\gamma}) = 1.008$ ;  $p < 0.0001$ ), which is consistent with the existing literature (Weiss & Murdoch, 2020).

We then examined the comorbidity effects on LOS with discharge alive, while adjusting for the influence of AGE as an additive effect in model (3),

$$\lambda(t|X, Z, t \leq U) = [\lambda_0(t) + \beta'X] \exp\{\theta'Z\}. \tag{19}$$

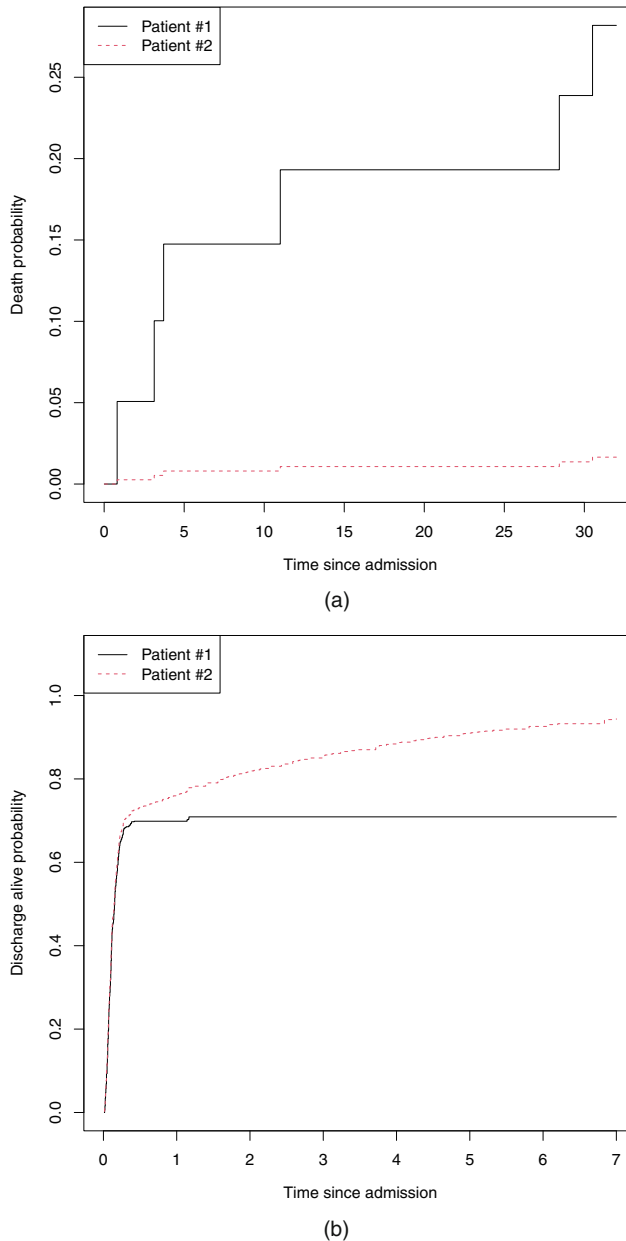
The independent variables of primary interest were comorbidity and smoking status. We included them in  $Z$  to retain the hazard ratio interpretation. Age was included in  $X$  as a modifier to the baseline hazard because we do not believe it would affect the hazard of discharge exponentially. Here we coded age as  $X_i = AGE_i - 9$ , representing the age difference from the oldest age category (9 in decades) to ensure  $\hat{\beta}'X_i > 0$  ( $i = 1, 2, \dots, 487$ ). We then screened the comorbidity and smoking indicators one at a time. Significant variables were included in  $Z_i$  for the final model (19),

**TABLE 6** Results of the length of stay analysis for patients that tested positive for COVID and were admitted to the hospital

	The proposed method			Scheike and Zhang (2002)		
	EST (SE)	p-value	95% CI	EST (SE)	p-value	95% CI
AGE	-0.0872 (0.0182)	0.0000	(-0.1229, -0.0515)	—	—	—
SMO	-0.3750 (0.1716)	0.0288	(-0.7113, -0.0387)	-0.1900 (0.1605)	0.2365	(-0.5045, 0.1246)
DIA	-0.2171 (0.0835)	0.0094	(-0.3808, -0.0534)	-0.1295 (0.1229)	0.2919	(-0.3704, 0.1113)
CKD	-0.3311 (0.1280)	0.0097	(-0.5820, -0.0803)	-0.3359 (0.2073)	0.1051	(-0.7421, 0.0704)
CHF	-0.2743 (0.1104)	0.0129	(-0.4907, -0.0580)	-0.3435 (0.2222)	0.1221	(-0.7790, 0.0920)

*Notes:* The left panel presents results from the proposed model. Significant multiplicative effects  $Z$  included smoking status (SMO), diabetes (DIA), chronic kidney disease (CKD), and congestive heart failure (CHF). AGE was included in  $X$  as an additive effect. The right panel are result from Scheike and Zhang’s (2002) model where deaths are treated as noninformative censoring events.

Abbreviations: 95% CI: 95% confidence interval; EST, point estimate; SE, standard error estimate.



**FIGURE 3** Predictive trajectories of (a) death probabilities and (b) successful discharge probabilities for observed COVID covariates [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

which also included age in  $X_i$ . We presented the model fitting results in the left panel of Table 6. Briefly, SMO, DIA, CKD, CHF and AGE were significantly associated with longer LOS. Based on the point estimates, a decade increase in age was associated with a reduced baseline hazard for discharge by 8.7%. SMO, DIA, CKD and CHF reduced discharge hazards by 31.3%, 19.5%, 28.2%, 24.0%, respectively. Among patients that were successfully discharged but had LOS greater than the median value (0.15 days), 15.8% were smokers, 35.8% had DIA, 13.3% CKD, 11.7% CHF, and the mean age was 50. A vast majority of the patients (98.3%) were successfully discharged and had LOS in less than 10 days.

We also analysed the COVID LOS data by treating death as a form of noninformative censoring by applying the method from Scheike and Zhang (2002) to fit model (17), where  $X$  and  $Z$  are defined as the same as with (19). The results are presented in Figure S3 for the cumulative coefficient estimates of  $A(t) = \int_0^t \beta(s)ds$  from AGE, and the right panel of Table 6 for the estimates of  $\theta$  elements representing comorbidity and smoking effects. Again, the model fitting results were different from those obtained from the proposed model. In particular, all comorbidities lost their significance. Such comparisons showed that ignoring the terminating events of death and treating them as noninformative censoring could lead to different and possibly questionable conclusions.

The median discharge time is 0.15 days, suggesting a significant portion of the patients were admitted for observation instead of treatment. We visually compared the predictive trajectories of death and discharge probabilities up to 7 days from two hypothetical patients as displayed by Figure 3. Suppose Patients #1 and #2 were both nonsmokers, free of DIA, CKD and CHF at admission, but one was 90 and the other one was 63 years of age. The younger patient is expected to be discharged sooner with a much lower probability of death.

## 5 | DISCUSSION

LOS is a critically important metric for the health status of hospitalised patients and a widely used indicator for care quality. A major challenge in analysing LOS with successful discharge is the handling of patient death, a terminating event that prevents the full observation of a hospital stay. LOS analyses in the medical literature have often ignored the influence of terminal events, and as a result, provided no interpretation for the factors associated with LOS among survivors. On the other hand, the competing risk or semi-competing risk models have focused on analysing either the first occurrence of the discharge and death events or the transition rates among different multi-states under unverifiable correlation assumptions. In this research, by modelling the survival probabilities and conditional discharge hazards, we present an alternative analysis that is more closely aligned with the practical interest of LOS analysis. The proposed model is flexible enough to accommodate a more general correlation structure instead of operating on assumed dependency. The accommodation of the additive and multiplicative effects has extended the model's flexibility. We show that the regression estimates also can be used for predicting both deaths and discharge probability trajectories. Coupled with asymptotic properties, the proposed method lays the foundation for formal statistical inference.

Methodologically, this research represents a more comprehensive approach for modelling hospital LOS. The model that we presented has a flexible structure that allows for further extensions. For example, one could extend the model to include other functional forms for  $g(\cdot)$  and  $h(\cdot)$ , and thus further enhancing its flexibility. An intuitive formulation is to replace  $\exp\{-\theta'Z_i(t)\}$  and  $\beta'X_i(t)$  with  $1/h\{\theta'Z_i(t)\}$  and  $g\{\beta'X_i(t)\}$ , respectively, in Equations (5) and (6), as long as  $h\{\theta'Z_i(t)\}$  and  $g\{\beta'X_i(t)\}$  yield positive hazard values on the time domain. Although this extension will not fundamentally alter the modelling structure, it will provide alternative interpretations for the additive and multiplicative effects. Another extension is to include time-varying effects in the additive or multiplicative components, by using appropriate smoothing techniques.

Through two real clinical studies, we have illustrated the use of the proposed model. We wish to emphasise that the model provides not only a structure for incorporating patient deaths when analysing LOS but also enhances the analysts' ability to account for the effects of different patient characteristics. This said, as with other sophisticated statistical models, increased modelling flexibility typically comes at the expense of model justification. For practical data analysts, a

question of foremost importance is what factors should be included in  $X$  and what factors should be included in  $Z$ . Such decisions ultimately rest with the investigators. How each factor exerts its influence on the care process and what hypotheses the investigators intend to test should be the primary consideration. As we have shown in the example data analysis, one might want to include the variables of primary interest in  $Z$  to retain the usual hazard ratio interpretation.  $X$ , on the other hand, can be viewed as adjustments to the baseline hazard. When independent variables are of higher dimensions, variable selection may become necessary. Penalisation or regularisation steps prior to employing the proposed inference could help alleviate the challenge (Li et al., 2021). All things considered, we have put forward a new analytical approach for studying the LOS, one that could be used in a broad range of clinical investigations. Computational code for method implementation is available upon request.

## ACKNOWLEDGEMENTS

The authors are grateful to the Editor, the Associate Editor, and two anonymous reviewers for their many insightful comments and suggestions, which have greatly improved this manuscript. The work was partially supported by the National Institutes of Health Grant U24 AA026969 and the Indiana Clinical and Translational Sciences Institute UL1TR002529.

## DATA AVAILABILITY STATEMENT

The IST data that partly support the findings of this study are available from the International Stroke Trial database published by Edinburgh DataShare: <https://datashare.ed.ac.uk/handle/10283/124>. The authors do not currently have the permission to publish or share the raw COVID data that were analysed in Section 4.2.

## ORCID

Yang Li  <https://orcid.org/0000-0001-7024-7103>

Wanzhu Tu  <https://orcid.org/0000-0002-4236-9135>

## REFERENCES

- Andersen, P.K. & Gill, R.D. (1982) Cox's regression model for counting processes: a large sample study. *The annals of statistics*, 10(4), 1100–1120.
- CDC. (2020) *People with certain medical conditions*.
- Cheng, X.-M., Kang, F.-Y., Zhou, J. & Wang, X. (2015) A general additive-multiplicative rates model for recurrent and terminal events. *Acta Mathematicae Applicatae Sinica, English Series*, 31(4), 1115–1130.
- Chima, C.S., Barco, K., Dewitt, M.L., Maeda, M., Teran, J.C. & Mullen, K.D. (1997) Relationship of nutritional status to length of stay, hospital costs, and discharge status of patients hospitalized in the medicine service. *Journal of the American Dietetic Association*, 97(9), 975–978.
- Clarke, A. & Rosen, R. (2001) Length of stay: how short should hospital care be? *The European Journal of Public Health*, 11(2), 166–170.
- CMS.gov. (2021) *Prospective payment systems - General information*. Available from: <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ProspectivePaymentSystem/ProspectivePaymentSystem> [Accessed 26th February 2021].
- Faddy, M., Graves, N. & Pettitt, A. (2009) Modeling length of stay in hospital and other right skewed data: comparison of phase-type, gamma and log-normal distributions. *Value in Health*, 12(2), 309–314.
- Fleming, T.R. & Harrington, D.P. (2011) *Counting processes and survival analysis*, Vol. 169. Hoboken, NJ: John Wiley & Sons.
- Freitas, A., Silva-Costa, T., Lopes, F., Garcia-Lema, I., Teixeira-Pinto, A., Brazdil, P. et al. (2012) Factors influencing hospital high length of stay outliers. *BMC Health Services Research*, 12(1), 1–10.
- Graves, N., Birrell, F. & Whitby, M. (2005) Effect of pressure ulcers on length of hospital stay. *Infection Control & Hospital Epidemiology*, 26(3), 293–297.

- Group, I. S. T. C. (1997) The international stroke trial (ist): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19 435 patients with acute ischaemic stroke. *The Lancet*, 349(9065), 1569–1581.
- Hsieh, J.-J., Wang, W. & Adam Ding, A. (2008) Regression analysis based on semicompeting risks data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1), 3–20.
- Kugler, C., Altenhöner, T., Lochner, P. & Ferbert, A. (2003) Does age influence early recovery from ischemic stroke? *Journal of Neurology*, 250(6), 676–681.
- Kwah, L.K. & Diong, J. (2014) National institutes of health stroke scale (nihss). *Journal of Physiotherapy*, 60(1), 61.
- Lee, K.H., Haneuse, S., Schrag, D. & Dominici, F. (2015) Bayesian semi-parametric analysis of semi-competing risks data: investigating hospital readmission after a pancreatic cancer diagnosis. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 64(2), 253.
- Li, J., Yu, T., Lv, J. & Lee, M.-L.T. (2021) Semiparametric model averaging prediction for lifetime data via hazards regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 70(5), 1187–1209.
- Li, J., Zhang, P., Wu, S., Yuan, R., Liu, J., Tao, W. et al. (2020) Impaired consciousness at stroke onset in large hemisphere infarction: incidence, risk factors and outcome. *Scientific Reports*, 10(1), 1–9.
- Li, R. & Cheng, Y. (2016) Flexible association modelling and prediction with semi-competing risks data. *Canadian Journal of Statistics*, 44(3), 361–374.
- Lin, D., Fleming, T. & Wei, L. (1994) Confidence bands for survival curves under the proportional hazards model. *Biometrika*, 81, 73–81.
- Lin, D. & Ying, Z. (1995) Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *The Annals of Statistics*, 23(5), 1712–1734.
- Liu, Y., Wu, Y., Cai, J. & Zhou, H. (2010) Additive–Multiplicative rates model for recurrent events. *Lifetime Data Analysis*, 16(3), 353–373.
- Luong, A. & Doray, L.G. (2009) Inference for the positive stable laws based on a special quadratic distance. *Statistical Methodology*, 6(2), 147–156.
- Marazzi, A., Paccaud, F., Ruffieux, C. & Beguin, C. (1998) Fitting the distributions of length of stay by parametric models. *Medical Care*, 36, 915–927.
- Martinussen, T. & Scheike, T.H. (2007) *Dynamic regression models for survival data*. New York, NY: Springer Science & Business Media.
- Pearson, S.D., Kleefield, S.F., Soukop, J.R., Cook, E.F. & Lee, T.H. (2001) Critical pathways intervention to reduce length of hospital stay. *The American Journal of Medicine*, 110(3), 175–180.
- Peng, L. & Fine, J.P. (2007) Regression modeling of semicompeting risks data. *Biometrics*, 63(1), 96–108.
- Poguntke, I., Schumacher, M., Beyersmann, J. & Wolkewitz, M. (2018) Simulation shows undesirable results for competing risks analysis with time-dependent covariates for clinical outcomes. *BMC Medical Research Methodology*, 18(1), 1–10.
- Pollard, D. (1990) Empirical processes: theory and applications. *Proceedings of the NSF-CBMS regional conference series in probability and statistics*, JSTOR, pp. i–86.
- Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V., Jr., Flournoy, N., Farewell, V.T. & Breslow, N.E. (1978) The analysis of failure times in the presence of competing risks. *Biometrics*, 34, 541–554.
- Sandercock, P., Niewada, M. & Czlonkowska, A. (2011) *International stroke trial database (version 2)*. University of Edinburgh. <https://doi.org/10.7488/ds/104>.
- Sarker, S., Min, D.-K., Black, T.R. & Lim, H.J. (2015) Application of multiplicative and additive hazards models to injury prevention among healthcare workers. *Advances in Research*, 5(1), 8–12.
- Scheike, T.H. & Zhang, M.-J. (2002) An additive–multiplicative Cox–Aalen regression model. *Scandinavian Journal of Statistics*, 29(1), 75–88.
- Schemper, M., Kaider, A., Wakounig, S. & Heinze, G. (2013) Estimating the correlation of bivariate failure times under censoring. *Statistics in Medicine*, 32(27), 4781–4790.
- Spiekerman, C.F. & Lin, D. (1998) Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, 93(443), 1164–1175.
- Sun, J. & Shen, J. (2009) Efficient estimation for the proportional hazards model with competing risks and current status data. *Canadian Journal of Statistics*, 37(4), 592–606.
- Sun, J., Sun, L. & Flournoy, N. (2004) Additive hazards model for competing risks analysis of the case-cohort design. *Communications in Statistics-Theory and Methods*, 33(2), 351–366.

- Sun, L., Liu, J., Sun, J. & Zhang, M.-J. (2006) Modeling the subdistribution of a competing risk. *Statistica Sinica*, 16, 1367–1385.
- van der Vaart, A.W. & Wellner, J.A. (1996) Weak convergence. In *Weak convergence and empirical processes* (pp. 16–28). New York, NY: Springer.
- Van der Vaart, A.W. (2000) *Asymptotic statistics*, Vol. 3. Cambridge, UK: Cambridge University Press.
- Vekaria, B., Overton, C., Wiśniowski, A., Ahmad, S., Aparicio-Castro, A., Curran-Sebastian, J. et al. (2021) Hospital length of stay for COVID-19 patients: data-driven methods for forward planning. *BMC Infectious Diseases*, 21(1), 1–15.
- von Cube, M., Schumacher, M. & Wolkewitz, M. (2017) Basic parametric analysis for a multi-state model in hospital epidemiology. *BMC Medical Research Methodology*, 17(1), 1–12.
- Wang, N., Wang, L. & McMahan, C.S. (2015) Regression analysis of bivariate current status data under the gamma-frailty proportional hazards model using the em algorithm. *Computational Statistics & Data Analysis*, 83, 140–150.
- Weiss, P. & Murdoch, D.R. (2020) Clinical course and mortality risk of severe COVID-19. *The Lancet*, 395(10229), 1014–1015.
- Xu, J., Kalbfleisch, J.D. & Tai, B. (2010) Statistical analysis of illness–Death processes and semicompeting risks data. *Biometrics*, 66(3), 716–725.
- Zhang, Z., Sun, J. & Sun, L. (2005) Statistical analysis of current status data with informative observation times. *Statistics in Medicine*, 24(9), 1399–1407.
- Zhang, Z., Sun, L., Sun, J. & Finkelstein, D.M. (2007) Regression analysis of failure time data with informative interval censoring. *Statistics in Medicine*, 26(12), 2533–2546.
- Zhao, H., Zhou, J. & Sun, L. (2013) A marginal additive rates model for recurrent event data with a terminal event. *Communications in Statistics-Theory and Methods*, 42(14), 2567–2583.
- Zhao, X., Zhou, J. & Sun, L. (2011) Semiparametric transformation models with time-varying coefficients for recurrent and terminal events. *Biometrics*, 67(2), 404–414.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Li, Y., Liu, H., Wang, X. & Tu, W. (2022) Semi-parametric time-to-event modelling of lengths of hospital stays. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 71(5), 1623–1647. Available from: <https://doi.org/10.1111/rssc.12593>

## APPENDIX

The following regularity conditions are needed in the proof:

- (C1)  $\{N_i(\cdot), X_i(\cdot), Z_i(\cdot), T_i, C_i, U_i, \delta_i\}_{i=1}^n$  are independent and identically distributed.
- (C2)  $P(U \geq \tau) > 0$ , and  $\int_0^\tau \lambda_0(t) dt < \infty$ .
- (C3) The covariates  $X(t)$  and  $Z(t)$  are almost surely bounded on  $t \in [0, \tau]$ .
- (C4) The quantity  $\Omega_\gamma = E \left[ \int_0^\tau \{W_i(t) - \bar{w}_{\gamma_0}(t)\}^{\otimes 2} e^{\gamma_0' W_i(t)} I(t \leq T_{2i}) d\Delta_0(t) \right]$  is non-singular, and  $\bar{w}_{\gamma_0}(t) = \frac{r^{(1)}(t)}{r^{(0)}(t)}$ ,  $r^{(k)}(t) = E \left[ I(t \leq T_{2i}) e^{\gamma_0' W_i(t)} W_i^{\otimes k} \right]$ ,  $k = 0, 1$ .
- (C5) There is a convex compact set  $\mathcal{B}$  containing  $\zeta_0 = (\beta_0', \theta_0)'$  such that  $(\beta_0, \theta_0)$  is the unique solution to the system  $u(\beta, \theta) = E \left\{ \int_0^\tau \phi_i(t) \{V_i(t) - \right.$

$$\bar{v}(t) \} \left[ \{ \lambda_0(t) + \beta_0 X_i(t) \} e^{(\theta_0 - \theta)' Z_i(t)} - \beta' X_i(t) \right] dt \} = 0, \quad \text{where } \bar{v}(t) = (\bar{x}'(t), \bar{z}'(t))', \quad \bar{x}(t) = \frac{E[X_i(t)\phi_i(t)]}{E[\phi_i(t)]}, \quad \bar{z}(t) = \frac{E[Z_i(t)\phi_i(t)]}{E[\phi_i(t)]}.$$

(C6)  $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$  is non-singular, where  $A_{11} = E \left[ \int_0^\tau (X_i(t) - \bar{x}(t)) \phi_i(t) X_i'(t) dt \right]$ ,  $A_{12} = E \left[ \int_0^\tau (X_i(t) - \bar{x}(t)) \phi_i(t) Z_i'(t) \{ \lambda_0(t) + \beta_0' X_i(t) \} e^{(\theta_0 - \theta)' Z_i(t)} dt \right]$ ,  $A_{21} = E \left[ \int_0^\tau (Z_i(t) - \bar{z}(t)) \phi_i(t) X_i'(t) dt \right]$  and  $A_{22} = E \left[ \int_0^\tau (Z_i(t) - \bar{z}(t)) \phi_i(t) Z_i'(t) \{ \lambda_0(t) + \beta_0' X_i(t) \} e^{(\theta_0 - \theta)' Z_i(t)} dt \right]$ .

The above regularity conditions are frequently used in the failure-time data analysis literature, and they are easily satisfied in many applications. Conditions (C1) and (C2) are similar to those of Andersen and Gill (1982, theorem 4.1) and are the standard conditions for failure-time data. Condition (C3) is about the boundedness and continuity and is commonly used in survival analysis. Conditions (C4)–(C6) are needed to ensure the existence and uniqueness of the proposed estimator. They are also needed to avoid technical distractions in the proofs, which are analogous to Scheike and Zhang (2002). For (C6), if  $A$  is positive definite,  $(\beta_0, \theta_0)$  will be the only solution to  $u(\beta, \theta) = 0$  so that the uniqueness in (C5) holds automatically.

**A.1 Proof of Theorem 1**

To prove the consistency and asymptotic normality of  $\hat{\zeta}_n$ , from Equations (5) and (6) we have

$$\sum_{i=1}^n \int_0^\tau \hat{\phi}_i(t) \left[ V_i(t) - \bar{V}(t) \right] \left[ \exp\{-\theta' Z_i(t)\} dN_i(t) - \beta' X_i(t) dt \right] = 0,$$

where  $\hat{\phi}_i(t) = \frac{I(t \leq T_{1i})}{\hat{S}_i(t)} = \frac{I(t \leq T_{1i})}{\exp \left\{ -\int_0^t \lambda_0^d(s) e^{\beta_0' W_i(s)} ds \right\}}$ . Define

$$U_n(\beta, \theta) = \sum_{i=1}^n \int_0^\tau \hat{\phi}_i(t) \left[ V_i(t) - \bar{V}(t) \right] \left[ \exp\{-\theta' Z_i(t)\} dN_i(t) - \beta' X_i(t) dt \right],$$

It follows that  $n^{-1} U_n(\beta, \theta) \xrightarrow{P} u(\beta, \theta)$  uniformly in  $\zeta \in \mathcal{B}$  by uniform consistency of  $\widehat{S}_i(t)$  to  $S_i^d(t)$  in probability (Fleming & Harrington, 2011) and the uniform strong law of large numbers (Pollard, 1990). The consistency of  $\hat{\zeta}_n$  follows under condition (C6) by theorem 5.9 of Van der Vaart (2000). To show that the asymptotic normality of  $\hat{\zeta}_n$ , we note that by the Taylor series expansion,

$$n^{1/2}(\hat{\zeta}_n - \zeta_0) = \left[ -n^{-1} \frac{\partial U(\zeta_0)}{\partial \zeta'} \Big|_{\zeta=\zeta_0} \right]^{-1} n^{-1/2} U(\zeta_0) + o_p(1) = A^{-1} n^{-1/2} U(\zeta_0) + o_p(1), \quad (A1)$$

where

$$\begin{aligned} U(\zeta_0) &= \sum_{i=1}^n \int_0^\tau \hat{\phi}_i(t) \left[ V_i(t) - \bar{V}(t) \right] \left[ \exp\{-\theta_0' Z_i(t)\} dN_i(t) - \beta_0' X_i(t) dt \right] \\ &= \sum_{i=1}^n \int_0^\tau \hat{\phi}_i(t) \left[ V_i(t) - \bar{V}(t) \right] \left[ \exp\{-\theta_0' Z_i(t)\} dN_i(t) - \beta_0' X_i(t) dt - \lambda_0(t) dt \right] \end{aligned}$$

$$= \sum_{i=1}^n \int_0^\tau (\phi_i(t) + \hat{\phi}_i(t) - \phi_i(t)) [V_i(t) - \bar{V}(t)] [\exp\{-\theta'_0 Z_i(t)\} dN_i(t) - \beta'_0 X_i(t) dt - \lambda_0(t) dt]. \tag{A2}$$

According to the functional delta method (van der Vaart & Wellner, 1996, theorem 3.9.4, p. 374), we have

$$\hat{\phi}_i(t) - \phi_i(t) = \frac{\phi_i(t)}{n} \left[ \sum_{i=1}^n \int_0^t \frac{e^{\gamma'_0 W_i(s)} dM_i^d(s)}{r^{(0)}(s)} + J(t; W_i)' \Omega_\gamma^{-1} \sum_{i=1}^n \int_0^\tau \left\{ W_i(u) - \frac{r^{(1)}(u)}{r^{(0)}(u)} \right\} dM_i^d(u) \right] + o_p(n^{-1/2}). \tag{A3}$$

Back to (A2),

$$\begin{aligned} U(\zeta_0) &= \sum_{i=1}^n \int_0^\tau [V_i(t) - \bar{v}(t)] dM_i(t) \\ &+ \sum_{i=1}^n \int_0^\tau [V_i(t) - \bar{v}(t)] [\exp\{-\theta'_0 Z_i(t)\} dN_i(t) - \beta'_0 X_i(t) dt - \lambda_0(t) dt] \frac{\phi_i(t)}{n} \\ &\times \left[ \sum_{i=1}^n \int_0^t \frac{e^{\gamma'_0 W_i(s)} dM_i^d(s)}{r^{(0)}(s)} + J(t; W_i)' \Omega_\gamma^{-1} \sum_{i=1}^n \int_0^\tau \left\{ W_i(u) - \frac{r^{(1)}(u)}{r^{(0)}(u)} \right\} dM_i^d(u) \right] + o_p(n^{1/2}) \\ &= \sum_{i=1}^n \int_0^\tau (V_i(t) - \bar{v}(t)) dM_i(t) + \sum_{i=1}^n \int_0^\tau \left[ \frac{\Psi(u)}{r^{(0)}(u)} + B \left\{ W_i(u) - \frac{r^{(1)}(u)}{r^{(0)}(u)} \right\} \right] dM_i^d(u) + o_p(n^{1/2}), \end{aligned} \tag{A4}$$

where

$$\begin{aligned} dM_i(t) &= \phi_i(t) [e^{-\theta'_0 Z_i(t)} dN_i(t) - \{\beta'_0 X_i(t) + \lambda_0(t)\} dt], \\ \Psi(u) &= E \left[ e^{\gamma'_0 W_i(u)} \int_0^\tau I(t > u) (V_i(t) - \bar{v}(t)) dM_i(t) \right] \\ dM_i^d(t) &= dN_i^d(t) - I(t \leq T_{2i}) e^{\gamma'_0 W_i(t)} d\Delta_0(t), \\ B &= E \left[ \int_0^\tau \{V_i(t) - \bar{v}(t)\} J(t, W_i)' \Omega_\gamma^{-1} dM_i(t) \right], \\ \Omega_\gamma &= E \left[ \int_0^\tau \{W_i(t) - \bar{w}_{\gamma_0}(t)\}^{\otimes 2} e^{\gamma'_0 W_i(t)} I(t \leq T_{2i}) d\Delta_0(t) \right], \\ J(t; W_i) &= \int_0^t e^{\gamma'_0 W_i(s)} \{W_i(s) - \bar{w}_{\gamma_0}(s)\} d\Delta_0(s), \end{aligned}$$

From (A1), (A4) and multivariate central limit theorem Theorem 1 holds.



**A.2 Proof of Theorem 2**

The uniform consistency of  $\hat{\Lambda}(t)$  follows using arguments similar to those in Spiekerman and Lin (1998) (Appendix B). To establish asymptotic normality of  $\hat{\Lambda}(t)$  for  $t \in [0, \tau]$ , where

$$\hat{\Lambda}(t) = \int_0^t \frac{\sum_{i=1}^n \hat{\phi}_i(s) \left[ e^{-\hat{\theta}'_n Z_i(s)} dN_i(s) - \hat{\beta}'_n X_i(s) ds \right]}{\sum_{i=1}^n \hat{\phi}_i(s)},$$

denote  $\hat{\Lambda}(t) = \hat{\Lambda}(t; \hat{\beta}_n, \hat{\theta}_n)$ . Consider

$$\hat{\Lambda}(t; \hat{\beta}_n, \hat{\theta}_n) = \hat{\Lambda}(t; \hat{\beta}_n, \hat{\theta}_n) - \hat{\Lambda}(t; \beta_0, \theta_0) + \hat{\Lambda}(t; \beta_0, \theta_0) - \Lambda_0(t),$$

where

$$\sqrt{n} \{ \hat{\Lambda}(t; \hat{\beta}_n, \hat{\theta}_n) - \hat{\Lambda}(t; \beta_0, \theta_0) \} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Gamma(t) A^{-1} (u_{1i} + u_{2i}) + o_p(1),$$

and

$$\begin{aligned} & \sqrt{n} \{ \hat{\Lambda}(t; \beta_0, \theta_0) - \Lambda_0(t) \} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \int_0^t \frac{dM_i(s)}{\hat{\kappa}(s)} + \int_0^t \frac{\hat{\phi}_i(s) - \phi_i(s)}{\hat{\kappa}(s)} \left\{ e^{-\theta'_0 Z_i(s)} dN_i(s) - \beta'_0 X_i(s) ds - \lambda_0(s) ds \right\} \right] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \int_0^t \frac{dM_i(s)}{\kappa(s)} + \int_0^t \frac{G(u, t) dM_i^d(u)}{r^{(0)}(u)} + O(t) \int_0^\tau \left\{ W_i(u) - \frac{r^{(1)}(u)}{r^{(0)}(u)} \right\} dM_i^d(u) \right] + o_p(1), \end{aligned}$$

where  $\kappa(s) = \frac{1}{n} \sum_{i=1}^n \phi_i(s)$ ,

$$\begin{aligned} \Gamma(t) &= E \left[ \int_0^t \frac{\phi_i(s)}{\kappa(s)} \left( -X'_i(s) ds, -Z'_i(s) e^{(\theta - \theta_0)' Z_i(s)} \{ \lambda_0(s) + \beta'_0 X_i(s) \} ds \right) \right], \\ G(u, t) &= E \left[ e^{\gamma'_0 W_i(u)} \int_0^t \frac{I(s > u)}{\kappa(s)} dM_i(s) \right], \\ \hat{O}(t) &= E \left[ \int_0^t \frac{1}{\kappa(s)} J(s, W_i)' \Omega_\gamma^{-1} dM_i(s) \right]. \end{aligned}$$

By the multivariate central limit theorem and following similar arguments in Zhao et al. (2013, supplementary materials), Theorem 2 holds.