Statistica Sinica Preprint No: SS-2021-0051	
Title	Sieve Estimation of a Class of Partially Linear Transformation Models With Interval-Censored Competing Risks Data
Manuscript ID	SS-2021-0051
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0051
Complete List of Authors	Xuewen Lu, Yan Wang, Dipankar Bandyopadhyay and Giorgos Bakoyannis
Corresponding Author	Xuewen Lu
E-mail	xlu@ucalgary.ca

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

Statistica Sinica

Sieve estimation of a class of partially linear transformation models with interval-censored competing risks data

Xuewen Lu[1], Yan Wang [1], Dipankar Bandyopadhyay[2], and Giorgos Bakoyannis[3]

[1] *University of Calgary, Calgary, AB, T2N 1N4, Canada*

[2] *Virginia Commonwealth University, VA 23298-0032, USA*

[3] *Indiana University, Indianapolis, IN, USA*

Abstract: We consider a class of partially linear transformation models with interval-censored competing risks data. Under a semiparametric generalized odds rate specification for the cause-specific cumulative incidence function, we obtain optimal estimators of the large number of parametric and nonparametric model components by maximizing the likelihood function over a joint B-spline and Bernstein polynomial spanned sieve space. Our specification considers a relatively simpler finite-dimensional parameter space, approximating the infinite-dimensional parameter space as $n \to \infty$. This allows us to study the almost sure consistency and rate of convergence for all parameters, and the asymptotic distributions and efficiency of the finite-dimensional components. We study the finite-sample performance of our method using simulation studies under a variety of scenarios.

Corresponding author. Email: lux@math.ucalgary.ca. Phone: (1-403)220-6620. Fax: (1-403)282-5150

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

Furthermore, we illustrate our methodology by applying it to a data set on HIV-infected individuals from sub-Saharan Africa.

Key words and phrases: Bernstein polynomials, Competing risks, Cumulative incidence function, Interval censoring, Partially linear transformation model, Semiparametric efficiency.

1. Introduction

In biomedical studies with time-to-event outcomes, there could be several distinct causes of failure, referred to as competing risks (Crowder, 2001). For example, when studying 137 bone marrow transplant (BMT) patients (Klein and Moeschberger, 2006), patients may relapse or die while in remission during the follow-up period. If we consider relapse to be the event of interest, then death is a competing risk/event, because it impedes the occurrence of leukemia relapse. Competing risks data are often subject to interval censoring, implying that the event time is not observed precisely, but the interval in which it lies is known. Another example is our motivating HIV data set, generated from a large study of HIV care and treatment programs in sub-Saharan Africa (Egger et al., 2012) and analyzed in this study. HIV-infected individuals receiving care in these programs may die while in care, or become lost to care. These are the two competing risks

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

under consideration, with the corresponding time to events available as interval endpoints. Patients who are lost to care typically do not receive treatment and, thus, are more likely to die and further contribute to the expansion of the HIV epidemic.

Statistical models under the umbrella of interval-censored competing risks data (Hudgens et al., 2014) can be broadly classified into (a) the cause-specific hazards (CSH) modeling framework, or (b) the cumulative incidence function (CIF) modeling framework. For CIF modeling, a number of approaches have been proposed. For example, Li (2016) considered a sieve maximum likelihood approach (sieve-ML) for the Fine–Gray model (Fine and Gray, 1999) under interval censoring and possible left-truncation. In addition, Mao et al. (2017) proposed a broad class of semiparametric regression models accommodating both proportional and non-proportional sub-distribution hazards, and devised a fast and stable EM-type estimation framework. Bakoyannis et al. (2017) considered a class of semiparametric generalized odds rate (GOR) transformation models (Scharfstein et al., 1998) using the sieve-ML approach based on B-splines, showing that the estimator for the (finite-dimensional) regression parameter is semiparametrically efficient. All of these works assume that the covariates are related linearly to the time-to-event responses, which precludes an assessment of

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

potential nonlinear and nonparametric patterns. This linearity assumption is too ideal to apply in many situations, thereby making the aforementioned methods inconsistent. For example, in HIV studies (such as our motivating data set), age is usually considered an important predictor of the HIV-1 disease progression. The progress to acquired immunodeficiency syndrome (AIDS), a chronic and potentially life-threatening condition, is more rapid in older adults than it is in younger patients, with a higher mortality among older patients developing an AIDS-defining illness (Nguyen and Holodniy, 2008; Pirrone et al., 2013). Hence, a partially linear model (Lu and Song, 2015) for the patients' ages seems more plausible for fitting the data.

From the context of HIV data modeling, we propose a semiparametric partially linear transformation model, with a GOR specification for the CIF function. Our model includes some commonly used models as special cases, such as the linear transformation model and nonparametric additive models. We obtain optimal estimators of the large number of parametric and nonparametric model components by maximizing the likelihood function over a joint B-spline and Bernstein polynomial (BP; Lorentz, 1986) spanned sieve space. The BP approach, which we employ to estimate the unknown nonparametric risk functions, enjoys several merits from the perspective of implementation; it often requires only a few parameters for a de-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

cent approximation, and one does not need to specify the interior knots, as in B-splines (Eilers and Marx, 1996). Our specification transfers the setup consisting of both finite- and infinite-dimensional parameters into a relatively simpler finite-dimensional framework that approximates the infinite-dimensional parameter space as $n \to \infty$. This allows us to study the almost sure consistency and rate of convergence for all parameters, and the asymptotic distributions and efficiency of the finite-dimensional components.

The rest of the paper proceeds as follows. In Section 2, we describe the statistical framework of our partially linear transformation GOR model. The associated sieve-ML estimation method, related large-sample results, and the implementation are presented in Section 3. In Section 4, we study the finite-sample performance of our proposed method using simulation studies under a variety of scenarios with synthetic data. Furthermore, we illustrate our methodology by applying it to the motivating HIV data set in Section 5. Finally, Section 6 concludes the paper. Detailed derivations and proofs of the theoretical results presented in Section 3 and tables and figures showing all the simulation results in Section 4 are relegated to the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

2. Statistical Model

We assume that there are a finite number k of competing risks, with the cause of failure and the (true) failure time denoted by C and T, respectively. The two covariate vectors $Z = (Z_1, \ldots, Z_d)^\top \in R^d$ and $W = (W_1, \ldots, W_q)^\top \in R^q$ have potential effects on the survival probability of T, where the effects of Z are modeled parametrically, the effects of W are modeled nonparametrically, and both the parametric and nonparametric components are of interest. For the competing risks data, the cause-specific CIF is defined as

$$F_j(t; z, w) = \Pr(T \le t, C = j) | Z = z, W = w), \quad j = 1, \ldots, k.$$

Then, for modeling F_j, we propose the following partially linear transformation model:

$$g_j[F_j(t; z, w)] = \phi_j(t) + \beta_j^\top z + \sum_{e=1}^{q} \psi_{je}(w_e), \quad j = 1, \ldots, k, \tag{2.1}$$

where g_j is a known increasing cause-specific link function, ϕ_j is an unspecified, strictly increasing, and invertible function of time t, β_j is a vector of parameters for the parametric components, z is a d-dimensional covariate vector, and ψ_{je} are unknown smooth regression functions of w_e, with $e = 1, \ldots, q$.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

We consider a special subset of the class of partially linear transformation models, specifically, the class of partially linear GOR (PLGOR) transformation models, with link functions given by:

$$g_j\left(F_j;\alpha_j\right)=\begin{cases}\log\left[\frac{(1-F_j)^{-\alpha_j}-1}{\alpha_j}\right] & \text{if } 0<\alpha_j<\infty,\\ \\ \log\left[-\log(1-F_j)\right] & \text{if } \alpha_j=0,\end{cases}$$

for $j=1,\ldots,k$. This class of models includes the linear GOR transformation models (Dabrowska and Doksum, 1988; Scharfstein et al., 1998; Fine, 2001; Jeong and Fine, 2006) as special cases, including the proportional odds (PO) model and proportional subdistribution hazards model (Fine, 1999) with $\alpha_j=1$ and $\alpha_j=0$, respectively (Jeong and Fine, 2006). Note that the link functions are allowed to vary with the causes of failure. Following prior studies, we assume that the true link functions are known (Scharfstein et al., 1998; Fine and Gray, 1999; Fine, 2001; Mao and Wang, 2010; Bakoyannis et al., 2017). This assumption facilitates our estimation, given that the estimation of $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_k)^\top$ may be hindered by identifiability concerns, as in the non-competing-risk setting (Zeng et al., 2006).

In practice, the observation times (e.g., the time to clinic visits, laboratory tests, etc.) could be interval censored. Let $(U_1,\ldots,U_m)$ denote $m\in(0,\infty)$ distinct observation times, which may vary from subject to

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

subject. Let $V \in \{0, U_1, \ldots, U_m\}$ correspond to the last observation time prior to the failure, and $U \in \{U_1, \ldots, U_m, \infty\}$ be the first observation time after the failure; then, the observed interval is (V, U). Using this notation, a left-censored observation corresponds to $(V, U) = (0, U_1)$, and a right-censored observation corresponds to $(V, U) = (U_m, \infty)$. For $j = 1, \ldots, k$, if a subject fails from the jth cause of failure before the first observation time U_1 (i.e., it is left censored), we observe $\delta_j^1 = 1$. If the subject fails between $V > U_1$ and $U \leq U_m$, we observe $\delta_j = 1$. However, if the subject is right censored (i.e., $T > U_m$), then we observe $\delta = \sum_{j=1}^{k}(\delta_j + \delta_j^1) = 0$. We assume the observation interval to be $[a, b]$, that is, $a = U_1 < U_m = b$. Including the two covariate vectors Z and W, the observed data are U, V, Z, W, $\{\delta_j\}_{j=1}^k$, and $\{\delta_j^1\}_{j=1}^k$. We further assume the following two fundamental conditions:

A1. The observation times $(U_1, \ldots, U_m)$ are independent of (T, C), conditional on (Z, W).

A2. The distribution of $(U_1, \ldots, U_m)$ does not contain the parameters that govern the distribution of (T, C) (non-informative interval censoring).

Given the two assumptions above and the observed data

$$D = \{v_i, u_i, z_i, w_i, \{\delta_{ij}\}_{j=1}^k, \{\delta_{ij}^1\}_{j=1}^k, \delta_i = \sum_{j=1}^{k}(\delta_{ij} + \delta_{ij}^1), i = 1, \ldots, n\},$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

the likelihood function in terms of the cause-specific CIFs is

$$\begin{aligned} L(\theta; D) \quad \propto \quad & \prod_{i=1}^{n} \left\{ \prod_{j=1}^{k} \left[F_j(u_i; z_i, w_i, \theta_j) - F_j(v_i; z_i, w_i, \theta_j) \right]^{\delta_{ij}} \right\} \\ & \left\{ \prod_{j=1}^{k} \left[F_j(u_i; z_i, w_i, \theta_j) \right]^{\delta_{ij}^1} \right\} \\ \times \quad & \left[1 - \sum_{j=1}^{k} F_j(v_i; z_i, w_i, \theta_j) \right]^{1-\delta_i}, \end{aligned} \tag{2.2}$$

where $\theta = (\theta_1^\top, \ldots, \theta_k^\top)^\top$ is a parameter that includes, in our case, regression coefficients for the effect of the covariate vector Z on each cause-specific CIF, and unspecified functions of time, after adjusting the nonlinear effect of the covariate W.

In our model, there are two sets of nonparametric components. The first set consists of $\{\phi_j, 1 \le j \le k\}$, and the second set consists of $\{\psi_{je}, 1 \le j \le k, 1 \le e \le q\}$. We choose to use B-splines to model the nonparametric ϕ_j functions in the first set, and BPs to model the nonparametric components ψ_{je} in the second set. One may also use B-splines to model the nonparametric functions in the second set. The motivation for using two different series is that one can easily rely on the `R` package `intccr` (Park et al., 2019) to implement the B-spline modeling of the ϕ_j functions, naturally preserving the nonnegativity and monotonicity. On the other hand, BPs possess the optimal shape-preserving property among all approximation

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

polynomials (Carnicer and Peña, 1993). This provides a flexible estimation strategy free of knot specifications for the ψ_{je} functions, which represent the nonparametric effect. BPs have been proved to be effective in modeling the nonparametric components in various semiparametric models; see, for example, Zhou et al. (2017).

The main contributions of this study are as follows. First, we present a class of PLGOR transformation models for interval-censored competing risks data that extends the linear GOR transformation models (Mao et al., 2017; Bakoyannis et al., 2017) and the nonparametric additive transformation models. Second, our sieve-ML proposal is a pragmatic compromise between a purely B-spline approach, with a faster convergence rate under the same smoothness conditions for different nonparametric functions, and a purely BP approach, with better computability and a better shape-preserving property. Finally, with regard to theoretical contributions, our proofs of the almost sure consistency, rate of convergence, and asymptotic normality use the theory of empirical processes and functional analysis. Furthermore, they are established using new techniques, such as the symmetrization inequality (Panchenko, 2003), Hoeffding's inequality, and the Riesz representation theorem (Goodrich, 1970). These are more challenging than the techniques used for the linear transformation models in Mao et al.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

(2017) and Bakoyannis et al. (2017).

3. Estimation and Implementation

3.1 Sieve Maximum Likelihood Estimation

As in Hu and Xiang (2016), Zhang et al. (2010), Lu and Song (2015), and Li (2016), we avoid imposing parametric assumptions on ϕ_j and ψ_j in (2.1) and, thus, the likelihood involves $(q+1)k$ infinite-dimensional or functional parameters. In general, maximizing the likelihood function with an infinite-dimensional parameter $\theta \in \Theta$ over Θ may lead to inconsistent maximum likelihood estimates (Shen and Wong, 1994). One approach to overcome this problem is to use a sieve-ML estimation. A sieve (Shen and Wong, 1994) is a sequence $\{\Theta_n\}_{n\geq 1}$ of parameter spaces that approximate (in a certain sense) the original parameter space Θ, with the approximation error tending to zero as $n \to \infty$. A sieve-ML estimate is the estimate obtained by maximizing the likelihood function over Θ_n. Another practical advantage of using the sieve-ML approach is that it reduces the dimensionality of the optimization problem, and thus the computational burden, compared with a fully semiparametric likelihood approach (Zhang et al., 2010). This is because the dimension of Θ_n is significantly smaller (i.e., it involves fewer parameters to be estimated) than that of the full parameter space Θ in

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

finite samples. The computational advantage of the sieve-ML approach over a fully semiparametric maximum likelihood approach for interval-censored survival data is shown in a simulation study by Zhang et al. (2010). Denote $\beta = (\beta_1^\top, \ldots, \beta_k^\top)^\top$. We define the sieve space as

$$\begin{aligned}\Theta_n &= \Big\{\theta = (\beta, \phi_{1n}, \ldots, \phi_{kn}, \psi_{(11)n}, \ldots, \psi_{(1q)n}, \ldots, \psi_{(k1)n} \ldots, \psi_{(kq)n}) \\ &\in \mathcal{B} \otimes \mathcal{M}_n^1 \otimes \cdots \otimes \mathcal{M}_n^k \otimes \mathcal{W}_n^{(11)} \otimes \cdots \otimes \mathcal{W}_n^{(1q)} \otimes \cdots \otimes \mathcal{W}_n^{(k1)} \otimes \cdots \otimes \mathcal{W}_n^{(kq)}\Big\}.\end{aligned}$$

In our case, the sequence of approximating functional parameter spaces for $\{\phi_j, 1 \le j \le k\}$ is chosen to be spaces of monotone (owing to the monotonicity of the CIF) B-spline functions, which are, for $j = 1, \ldots, k$, defined by

$$\mathcal{M}_n^j = \left\{\phi_{jn}(t) = \sum_{s=1}^{m} \gamma_{js} B_s(t, m, a, b) : \max_{0 \le s \le m} |\gamma_{js}| \le M_n,\ 0 \le \gamma_{j0} \le \gamma_{j1} \le \cdots \le \gamma_{jm}\right\}.$$

On the other hand, the sequence of approximating functional parameter spaces for $\{\psi_{je}, 1 \le j \le k, 1 \le e \le q\}$ is chosen to be spaces of BPs without constraints of monotonicity, which are, for each combination (j, e), for $1 \le j \le k$ and $0 \le e \le q$, defined by

$$\mathcal{W}_n^{(je)} = \left\{\psi_{(je)n}(w_e) = \sum_{s=0}^{m_w} \alpha_{(je)s}\{B_s^e(w_e, m_w, a_w^e, b_w^e) - B_s^e(a_w^e, m_w, a_w^e, b_w^e)\} : \max_{0 \le s \le m_w} |\alpha_{(je)s}| \le M_n\right\},$$

where for each $e = 1, \ldots, q$, $\{B_s^e(w_e, m_w, a_w^e, b_w^e)\}_{s=0}^{m_w}$ are Bernstein basis

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

polynomials defined as

$$B_s^e(w_e, m_w, a_w^e, b_w^e) = \binom{m_w}{s} \left\{ \frac{w_e - a_w^e}{b_w^e - a_w^e} \right\}^s \times \left\{ 1 - \frac{w_e - a_w^e}{b_w^e - a_w^e} \right\}^{m_w - s}, s = 0, \ldots, m_w,$$

with degree $m_w = o(n^\nu)$, for some $\nu \in (0, 1)$. Here, we use the same degree m_w for each ψ_{je}. For identifiability, we assume $\psi_{(je)n}(w_e) = 0$ when $w_e = a_w^e$, i.e., $\psi_{(je)n}(a_w^e) = 0$. This is why we subtract the term $B_s^e(a_w^e, m_w, a_w^e, b_w^e)$ from each summand.

A major computational advantage of using a BP for ψ_{je} is that the sieve space defined by $\mathcal{W}_n^{(je)}$ takes the simplest form satisfying the identifiability condition $\psi_{je}(a_w^e) = 0$, because for all $s = 1, \ldots, m_w$, the Bernstein basis polynomials automatically satisfy $B_s^e(w_e, m_w, a_w^e, b_w^e)|_{w_e = a_w^e} = 0$. One can show that the size of the sieve spaces defined above can be controlled by $M_n = O(n^\mu)$, with μ being a positive constant (Lorentz, 1986; Shen et al., 1997). During the likelihood maximization, we impose the monotonicity constraints $\gamma_{js} \leq \gamma_{j(s+1)}$, for every $s = 0, \ldots, (m-1)$ and $j = 1, \ldots, k$. Additionally, the constraint

$$\max_{z,w} \left\{ \sum_{j=1}^{k} F_j(t; z, w, \theta_j) \right\} < 1 \tag{3.3}$$

is needed to ensure that the sum of the estimated CIFs at the maximum follow-up time t is bounded above by one, where the maximum is over all the observed covariate patterns.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

Let the true parameter values be denoted by $\theta_0 = (\beta_0^\top, \phi_0^\top, \psi_0^\top)^\top$, where $\beta_0 = (\beta_{1,0}^\top, \ldots, \beta_{k,0}^\top)^\top$, $\phi_0 = (\phi_{1,0}, \ldots, \phi_{k,0})^\top$, and $\psi_0 = (\psi_{1q,0}, \ldots, \psi_{1q,0}, \ldots, \psi_{k1,0}, , \ldots, \psi_{kq,0})^\top$, and let the corresponding sieve-ML estimator be denoted by $\hat{\theta}_n = (\hat{\beta}_n^\top, \hat{\phi}_n^\top, \hat{\psi}_n^\top)^\top$, where $\hat{\beta}_n = (\hat{\beta}_{1,n}^\top, \ldots, \hat{\beta}_{k,n}^\top)^\top$, $\hat{\phi}_n = (\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{k,n})^\top$, and $\hat{\psi}_n = (\hat{\psi}_{11,n}, \ldots, \hat{\psi}_{1q,n}, \ldots, \hat{\psi}_{k1,n}, \ldots, \hat{\psi}_{kq,n})^\top$. In addition, define the L_2-metric for the distance between two parameters $\theta_1 = (\beta^{(1)\top}, \phi^{(1)\top}, \psi^{(1)\top})^\top$ and $\theta_2 = (\beta^{(2)\top}, \phi^{(2)\top}, \psi^{(2)\top})^\top$ as

$$d(\theta_1, \theta_2) = \left(\sum_{j=1}^{k} \left\| \beta_j^{(1)} - \beta_j^{(2)} \right\|^2 + \sum_{j=1}^{k} \left\| \phi_j^{(1)} - \phi_j^{(2)} \right\|_\Phi^2 + \sum_{j=1}^{k} \left\| \psi_j^{(1)} - \psi_j^{(2)} \right\|_\Psi^2 \right)^{\frac{1}{2}},$$

where

$$\left\| \phi_j^{(1)} - \phi_j^{(2)} \right\|_\Phi^2 = E\left[\phi_j^{(1)}(V) - \phi_j^{(2)}(V) \right]^2 + E\left[\phi_j^{(1)}(U) - \phi_j^{(2)}(U) \right]^2, \quad j = 1, \ldots, k$$

and

$$\left\| \psi_j^{(1)} - \psi_j^{(2)} \right\|_\Psi^2 = \sum_{e=1}^{q} \left\| \psi_{je}^{(1)} - \psi_{je}^{(2)} \right\|_\Psi^2 = \sum_{e=1}^{q} E\left[\psi_{je}^{(1)}(W_e) - \psi_{je}^{(2)}(W_e) \right]^2, \quad j = 1, \ldots, k,$$

and $\|\cdot\|$ denotes the Euclidean norm. Under the conditions given in the Supplementary Material, we obtain the following theorems about the asymptotic properties of the proposed estimators.

Theorem 1. *Assume Conditions (C1)–(C6) given in the Supplementary Material hold. Then,*

$$d(\hat{\theta}_n, \theta_0) \xrightarrow{a.s.} 0.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

Therefore, the combined B-spline and BP-based sieve-ML estimator is strongly consistent.

Theorem 2. *Assume Conditions (C1)–(C7) given in the Supplementary Material hold. Then*

$$d(\hat{\theta}_n, \theta_0) = O_p\left\{n^{-\min[\nu\sigma,(1-\nu)/2]}\right\},$$

where $\nu \in (0,1)$ *such that* $m = O(n^{\nu})$ *and* $m_w = O(n^{\nu})$*, and* $\sigma = \min(p, r/2)$*, with* p *and* r *defined in Condition (C4).*

This theorem implies that the convergence rate of the estimator for the functional parameters is slower than the usual $\sqrt{n}$ rate. This estimator achieves the optimal convergence rate for nonparametric regression estimators, which is $n^{-\sigma/(1+2\sigma)}$ when one chooses $\nu = 1/(1+2\sigma)$. Furthermore, when $p \geq r/2$, we obtain $\sigma = r/2$, and the convergence rate becomes $n^{-r/\{2(1+r)\}}$. A similar result was obtained by Zhou et al. (2017) when estimating the unknown functions using BPs, but only in a regression analysis of bivariate interval-censored failure time data. In our case, because we use both B-splines and BPs, the convergence rate is dominated by the smoothness level of the regression risk functions ψ_{je}, which are modeled by BPs. In fact, if we used B-splines to model ϕ_j and ψ_{je} simultaneously, then, under the same smoothness level $p = r$ for ϕ_j and ψ_{je}, we would obtain a

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

faster convergence rate $n^{-r/\{1+2r\}}$, the same as that obtained by Lu and Song (2015) for the partially linear additive hazards model with current status data. Although a purely B-spline based estimator has some better theoretical large-sample properties under the same smooth conditions, we choose BPs to model the nonparametric risk functions, because the resultant estimator has some superior finite properties, as discussed in Section 2. On the other hand, if $p \leq r/2$ or $r \geq 2p$, the convergence rate becomes $n^{-p/\{1+2p\}}$, and we still obtain the same optimal convergence rate as that of a purely B-spline based estimator.

Theorem 3. *Assume Conditions (C1)–(C8) given in the Supplementary Material hold. Then,*

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N[0, I^{-1}(\beta_0)],$$

which implies that the convergence rate of the estimator for the Euclidean parameter β_0 is $\sqrt{n}$. This also points to the efficiency of this estimator because the corresponding variance matrix attains the semiparametric efficiency bound $I(\beta_0)$.

Detailed proofs of the theorems, using necessary regularity conditions (Zhang et al., 2010; Zhou et al., 2017), are relegated to Section SM1 of the accompanying Supplementary Material. The information matrix $I(\beta_0)$ is

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

also defined there. Because finding $I(\beta_0)$ involves solving an integral equation with no explicit solution, estimating $I(\beta_0)$ using $I(\hat{\beta}_n)$ is not straightforward (Zhang et al., 2010; Li, 2016). Consequently, one can either use the least squares method of Zhang et al. (2010) and Li (2016) for a standard error estimation, or rely on the computationally simpler nonparametric bootstrap method. The validity of the bootstrap for the Euclidean parameter estimates in general semiparametric M-estimation problems has been verified by Cheng et al. (2010).

3.2 Implementation

The proposed methodology can be implemented easily using the function `ciregic`, available in the `R` package `intccr` (Park et al., 2019). The BP uses fixed knots once the degree m_w and the interval $[a_w^e, b_w^e]$, that is, the support of W_e, are determined. This nice property of a BP allows us bypass the knot selection procedure during the estimation. In practice, we can set $m_w = [n^{1/3}]$, or use the BIC (shown in the data analysis) to select m_w.

In practice, the true link functions are usually unknown. Hence, selecting the link function parameters is necessary. We use a grid search over a plausible combination of $\alpha_1, \ldots, \alpha_k$, and select the combination of α and the degrees m and m_w of the B-splines and BPs, respectively, using the

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

BIC. The final step is to obtain the parameter estimates of the model by maximizing the likelihood function (2.2) using a constrained optimization algorithm. The R package `alabama` provides a useful set of functions for optimization under both linear and nonlinear inequality constraints. The `ciregic` function uses this package to impose the monotonicity and boundedness constraints automatically. We implement this routine in our subsequent data analysis. The associated `R` code is available from the following `GitHub` link: `https://github.com/bandyopd/PLTM-ICCR`.

4. Simulation study

In order to evaluate the finite-sample performance of our method, we perform simulation studies in which we generate synthetic data under three scenarios. In scenario 1, under a correct model specification, we compare the performance of the estimation procedure when the nonlinear regression functions are approximated by BPs and B-splines. In scenario 2, we evaluate the effect of covariate confounding and model misspecification. Finally, in scenario 3, we assess the model performance under a more complex nonparametric regression function than those of scenarios 1 and 2. Scenarios 2 and 3 and tables and figures summarizing the results from the three scenarios are presented in Section SM2 of the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

Scenario 1: We considered two causes of failure and three covariates in the model. The first two covariates $Z = (Z_1, Z_2)^\top$ have linear effects, and the third covariate W has a nonlinear effect, where $Z_1 \sim$ Bernoulli(p = 0.4), $Z_2 \sim N(0, 1)$, and $W \sim$ Unif$(0, 2\pi)$. The CIFs for causes 1 and 2 have a proportional odds (PO) form, given by

$$F_j(t) = \frac{\exp\left[\phi_j(t) + \beta_j^\top Z + \psi_j(W)\right]}{1 + \exp\left[\phi_j(t) + \beta_j^\top Z + \psi_j(W)\right]}, \quad j = 1, 2,$$

where $\exp[\phi_1(t)] = 0.4\,[1 - \exp(-0.6t)]\,/0.6$ and $\exp[\phi_2(t)] = 0.75\,[1 - \exp(-0.5t)]\,/0.5$ follow a baseline cumulative subdistribution hazard function from a Gompertz distribution (Jeong and Fine, 2007). Under this setting, the true nonlinear regression functions corresponding to the first and second causes of failure are $\psi_1(W) = \sin(W)$ and $\psi_2(W) = -\sin(W)$, respectively. The true values for the regression parameters are $\beta_1 = (\beta_{11}, \beta_{12})^\top = (0.5, -0.3)$ and $\beta_2 = (\beta_{21}, \beta_{22})^\top = (-0.5, 0.3)$. The parameters and functions were chosen such that $\lim_{t\to\infty}\{\text{CIF}_1(\text{t}) + \text{CIF}_2(\text{t})\} = 1$. Based on this model, we simulated the failure times and causes of failure. The first observation time U_{1i} (e.g., clinic visit) was simulated from an Exponential(3). The following observation times were placed at times V apart from the previous observation, where $V \sim$ Exponential(3), with an upper bound of three years. This choice led to an average time of four

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

months between two consecutive observations. The baseline cumulative incidence functionals $\phi_j(t)$ were approximated by B-spline functions, and the nonlinear regression functions $\psi_j(w)$ were approximated by BP functions, denoted as "B-spline+Bernstein polynomials." As suggested by an anonymous referee, we also included the "B-spline+B-spline" approach in our comparison, that is, where the nonlinear regression functions $\psi_j(w)$ are approximated by B-splines as well.

For the implementation, we set $k = 1$ in the R function *ciregic*. This implies that cubic B-splines with $[N^{1/3}]$ internal knots are used to approximate $\phi_j(t)$, where $[N^{1/3}]$ is the largest integer up to and including $N^{1/3}$, and N is the total number of distinct time points V_i and U_i for the non-right-censored subjects, plus the number of right-censored subjects. For the "B-spline+Bernstein polynomials" approach, we use BPs with $m = 5$ degrees for the approximation of $\psi_j(w)$ (i.e., six basis functions). For the "B-spline+B-spline" approach, we use six cubic B-spline basis functions. We consider two sample sizes, 100 and 500, with the number of simulated data sets for each scenario being 200. For estimating the standard error of $\hat{\beta}$, `ciregic` has two options: `nboot=0`, or `nboot=a positive integer`, where `nboot` indicates the number of bootstrap samples for estimating the variances and covariances of the estimated regression coefficients. When

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

`nboot = 0`, the least-squares (LS) method is used. Pilot simulation studies revealed that, although the LS method is significantly faster, it leads to biased variance estimates. Hence, we used the bootstrap method with `nboot = 500` bootstrap samples in our simulation studies.

The simulation results, which report the AEST (average estimate of β), MCSD (Monte Carlo standard deviation of estimates), ASE (average estimated standard errors of estimates), and ECP (empirical coverage probability), are presented in Table S1. For the "B-spline+Bernstein polynomials" approach, we observe that the MCSDs mostly agree with the ASEs, and the ECPs remain very close to the nominal 95% level, implying that the bootstrap method works well. Figures S1 and S2 display histograms of the estimates of β for $n = 100$ and 500, respectively, and Figures S3 and S4 present plots of the true and estimated baseline CIFs and nonlinear regression functions, respectively, also for sample sizes 100 and 500. The histograms reveal the satisfactory asymptotic normality of the estimator of β, even for relatively smaller sample sizes. The differences between the true and estimated functions are quite small, and decrease with an increasing sample size. Furthermore, the biases of the estimates of β are also small, revealing satisfactory performance of the approximations to the unknown nonlinear regression functions by the BPs.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

Table S1 also presents the corresponding statistics for the "B-spline+B-spline" approach, and Figures S5–S8 present histograms of the corresponding β estimates, and plots of the true and estimated baseline CIFs and nonlinear regression functions, respectively, also for sample sizes 100 and 500. We observe that the "B-spline+B-spline" approach yields overestimated coverage probabilities and a larger bias for the nonlinear regression function estimates, as compared with the "B-spline+Bernstein polynomials" approach. This indicates that the BPs possess a better shape-preserving property than that of the B-splines, and provides an improved estimation of the coverage probability.

5. Application: HIV Data

Based on the HIV data set (Bakoyannis et al., 2017), our goal is to employ the PLGOR transformation model to evaluate the factors potentially associated with the cumulative incidence of being lost to care and of death while in care (i.e., under continuous HIV care coverage and prior to becoming lost to care). These are the two competing risk endpoints that are interval censored. The former outcome (lost to care) is associated with non-retention in care, which can lead to increased mortality for the patient by virtue of not receiving antiretroviral treatment (ART). The data on 3053 patients

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

came from the East Africa IeDEA (International Epidemiologic Databases to Evaluate AIDS) Regional Consortium (Zaniewski et al., 2018), which involves HIV care and treatment programs in Kenya, Uganda, and Tanzania. Loss to care is defined by the clinicians as having no clinic visits for a three-month period. This cutoff was chosen because patients without care for three months are expected to have run out of ART supplies for at least one month. Such a treatment interruption is clinically significant because it is associated with increased viremia.

Table 1: HIV data analysis: Results from fitting the FGPO model, when $\alpha = (0, 1)$ with a new term Age^2

	A. Loss to care	**B. Death while in care**
Covariate	mRR[1]/$\hat{\beta}$ (p-value)	mRR[1]/$\hat{\beta}$ (p-value)
Age at ART initiation		
per 10 years	0.689[2]/-0.373 (<0.001)	2.040[3]/0.713 (<0.001)
Age^2 at ART initiation		
per 100 years	1.022[2]/0.022 (<0.001)	0.947[3]/-0.055 (<0.001)
Gender		
Male vs Female	1.175[2]/0.161 (0.087)	1.493[3]/0.401 (0.003)
CD4 at ART initiation		
per 100 cells/μl	1.032[2]/0.031 (0.264)	0.616[3]/-0.484 (0.003)

[1] Measure of relative risk [2] Subdistribution hazard ratio [3] Odds ratio

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

To motivate our PLGOR transformation model, we add a new covariate term Age^2 to the model that assumes the popular Fine–Gray proportional subdistribution hazards model for the *loss to care* outcome, and the PO model for the *death while in care* outcome. This model is obtained by setting $\boldsymbol{\alpha} = (0, 1)$, and is henceforth referred to as the FGPO model. The results are shown in Table 1. We observe that both Age and Age^2 are statistically significant in the model, which indicates a possible nonlinear effect of Age on the cause-specific CIF. Hence, a partially linear model, accommodating the nonlinear effect of Age, is preferred.

Now, to fit the PLGOR model, we employ BPs to model a nonparametric function of Age. The CIFs are fitted using cubic B-splines, with the number of knots controlled by the argument $k = 1$ in the R function `ciregic`. This results in 19 B-spline basis functions for each CIF. The knots are placed at the corresponding percentiles of the distribution of the observed times (V_i, U_i). We estimate the variance of the regression parameters using the bootstrap method with 500 replications. For practical implementation, we need to select the parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$, indexing the link functions for the competing events, and the degree m_w of the BP. To do so, we use the following grid-search BIC:

$$BIC(\boldsymbol{\alpha}, m_w) = -2\ell_n(\hat{\theta}; \boldsymbol{\alpha}) + \log(n)\{4 + 2(m_w + 1)\}.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

We perform a grid search over all possible combinations of $\alpha_1 \in \{0, 0.5, \ldots, 4\}$ and $\alpha_2 \in \{0, 0.5, \ldots, 4\}$. The minimum BIC value (6666.91) is achieved at $m_w = 2$, with $\boldsymbol{\alpha} = (2, 0)$, and the corresponding maximum log-likelihood value for the fitted PLGOR model is -3140.88. Despite the two additional parameters, this provides a better fit than the optimal linear transformation model (LTM) of Bakoyannis et al. (2017), where the maximum log-likelihood (-3144.47) was achieved at $\boldsymbol{\alpha} = (2, 1.2)$. For this comparison, we chose the argument $k = 1$ within the `ciregic` function, keeping the control on the number of knots in the B-spline structure unchanged. Our model also outperforms the popular Fine–Gray proportional subdistribution hazards model for both competing risks (FGFG, henceforth), which corresponds to $\boldsymbol{\alpha} = (0, 0)$, with the maximized log-likelihood value $= -3147.05$, and the FGPO model with the maximized log-likelihood value $= -3146.91$.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

Table 2: Comparison between the proposed model with $\boldsymbol{\alpha} = (2, 0)$, the Fine–Gray proportional odds model (FGPO) with $\boldsymbol{\alpha} = (0, 1)$, and the linear transformation model with $\boldsymbol{\alpha} = (2, 1.2)$

		FGPO: $\boldsymbol{\alpha} = (0, 1)$	$\boldsymbol{\alpha} = (2, 1.2)$	$\boldsymbol{\alpha} = (2, 0)$	
Outcome	**Covariate**	mRR[1] / $\hat{\beta}$ (p-value)	$\hat{\beta}$ (p-value)	$\hat{\beta}$ (p-value)	
A. Loss	**Age**				
to care	*per 10 years*	0.783[2] / -0.244 (<0.001)	-0.348 (<0.001)	$\psi_1(Age)$	$B_1(Age)$ 1.353(< 0.001)
					$B_2(Age)$ −0.992(0.117)
					$B_3(Age)$ 0.185(0.683)
	Gender				
	Male vs Female	1.170[2] / 0.157 (0.031)	0.273 (0.005)	0.282 (0.022)	
	CD4				
	per 100 cells/μl	1.002[2] / 0.002 (0.947)	0.001 (0.976)	< 0.001 (0.940)	
B. Death	**Age**				
	per 10 years	1.360[3] / 0.308 (<0.001)	0.329 (<0.001)	$\psi_2(Age)$	$B_1(Age)$ −1.038(0.005)
					$B_2(Age)$ 1.417(0.100)
					$B_3(Age)$ 0.168(0.808)
	Gender				
	Male vs Female	1.758[3] / 0.564 (0.008)	0.567 (0.008)	1.677[2] / 0.517 (0.004)	
	CD4				
	per 100 cells/μl	0.699[3] / -0.358 (0.003)	-0.346 (0.006)	0.997[2] / -0.0034 (0.026)	

[1] Measure of relative risk [2] Subdistribution hazard ratio [3] Odds ratio

The resultant parameter estimates (and p-values) obtained from fitting the FGPO model, the LTM of Bakoyannis et al. (2017) with $\boldsymbol{\alpha} = (2, 1.2)$,

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

and our proposed PLGOR model with $\boldsymbol{\alpha} = (2, 0)$ (the best-fitting model) are presented in Table 2. We observe that for both outcomes, males have a significantly higher CIF than that of females for all models. Although there is no evidence of an association between CD4 cell count at ART initiation and the CIF of loss to care (from all models), lower CD4 counts are significantly associated with increased CIF of death while in care, as revealed by the appropriate summary quantities (odds and relative risks) for the models. This agrees with previous studies (Lawn et al., 2009) exploring the effect of low CD4 cell counts on mortality risks. Finally, exploring the association of Age with the two competing risks revealed interesting findings. For proper interpretation, we plot the two fitted nonparametric risk functions in Figure 1. We observe that a lower age (<63 years) at ART initiation is associated with a decreased CIF of loss to care and an increased CIF of death while in care. However, for Age ≥ 63, the respective directions of the association are reversed, that is, an increased CIF for loss to care and a decreased CIF of death while in care. These new findings quantifying the time-varying effects of age were not revealed from the FGPO and LTM fits.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

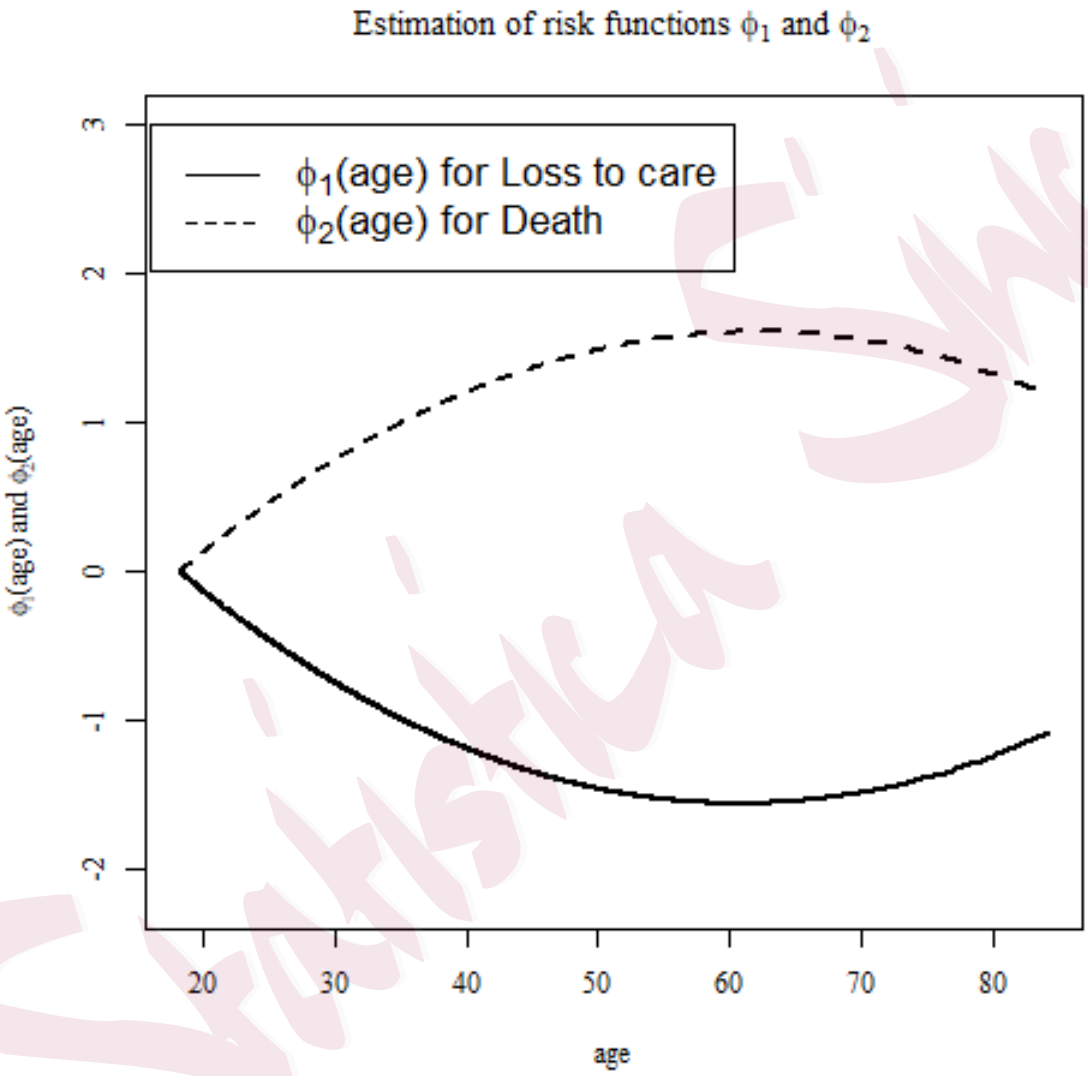

Figure 1: Estimated nonparametric risk functions of age at ART initiation, derived from fitting the PLGOR model to the HIV data. The solid and dashed lines represent the *loss to care* and *death in care*, respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

6. Conclusion

The central contribution of this study is to introduce covariate nonlinearity into GOR transformation models for interval-censored competing risks data. This general class includes many other semiparametric models as special cases, such as the PO (Jeong and Fine, 2006; Shi et al., 2013) and the proportional subdistribution hazards models (Fine and Gray, 1999). Our method comprises a purely B-spline approach, with a faster convergence rate under the same smoothness conditions for different nonparametric functions, and a purely BP approach to enhance the computational scalability, enforcing an optimal shape-preserving property among all approximating polynomials (Carnicer and Peña, 1993). Unlike other methods (Li, 2016), our proposal explicitly incorporates the boundedness of the cause-specific CIF constraint into the optimization. With regard to theoretical contributions, the sieve MLE for the regression (Euclidean) parameter based on the combined B-spline and BP sieves is shown to be consistent, semiparametrically efficient, and asymptotically normal, using newer techniques, such as the symmetrization inequality, Hoeffding's inequality, and the Riesz representation theorem. We further show that the estimators for the functional parameters (baseline CIFs and nonparametric risk functions) are consistent, almost surely in an L_2-metric, and converge at the optimal rate for

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

nonparametric regressions. We also provide an easy implementation of our model using the `R` function `ciregic`.

A practical issue during our model implementation is to decide between parametric and nonparametric choices for our covariates, or equivalently, how to determine q, representing the number of unknown smooth regression functions. To the best of our knowledge, there are no theoretical investigations in this regard within the interval-censored competing risks framework. However, for ordinary additive partially linear regressions models, Zhang et al. (2011) developed a method to distinguish linear and nonlinear terms for partially linear models, automatically and consistently. Their ideas can be adapted to our model, although a thorough investigation of this topic is beyond the scope of the current work. Hence, we suggest two pragmatic strategies in applications. One is simply to put discrete covariates in the linear part and continuous ones in the nonlinear part. Another more reasonable approach is called the screening method, which involves conducting an initial preliminary univariate analysis, and then separating the continuous covariates based on the shape of the estimated nonparametric functions. Specifically, one can include all the categorical covariates in the linear part, and then add each continuous covariate sequentially to construct a partially linear model. If the univariate, continuous covariate, partially linear model

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

indicates nonlinearity, we can assign this covariate to a nonlinear function; otherwise, we assign it to the linear part, and eventually determine q. This is what we did in the HIV data analysis.

There are a number of future directions to pursue. For example, the proposed methodology applies to the case of current-status data (Groeneboom et al., 2008b,a), which represents a more severe form of interval censoring (Zhang et al., 2010). It is also straightforward to show that our results hold for other general classes of partially linear semiparametric transformation models, such as the Box–Cox transformation models (Ji et al., 2017) for survival outcomes. Furthermore, we assumed that the true link functions are known, as in Scharfstein et al. (1998), Fine (1999, 2001), and Mao and Wang (2010). In practical applications, this assumption may not hold. Another practical issue is the selection of the degree of the BP. To this end, and following Mao and Wang (2010), we performed a grid search over a plausible set of combinations of $(\alpha_1, \ldots, \alpha_k)$ and a range of m_w values, and determined the combination using the BIC. However, the issues related to investigating the additional variability due to this type of model selection remain an open problem in semiparametric modeling, and require further investigation.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

Supplementary Material

The Supplementary Material consists of two sections. Section SM1 contains detailed derivations and proofs of the theoretical results presented in Section 3. Section SM2 contains scenarios 2 and 3, and tables and figures summarizing all the simulation results in Section 4.

Acknowledgments

Lu's research was partially supported by a Discovery Grant (RG/PIN06466-2018) from the National Science and Engineering Research Council (NSERC) of Canada. Bandyopadhyay acknowledges partial support from grants R01DE024984 and P30CA016059 awarded by the United States National Institutes of Health. Bakoyannis acknowledges partial support by R21AI145662 and R01AI140854 awarded by the United States National Institutes of Health. The authors would also like to thank the editor, associate editor, and two anonymous referees for their constructive comments. The research reported in this publication was supported by the National Institute Of Allergy And Infectious Diseases (NIAID), Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD), National Institute On Drug Abuse (NIDA), National Cancer Institute (NCI), and the National Institute Of Mental Health (NIMH), in accordance with the regulatory requirements

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

of the National Institutes Of Health under Award Number U01AI069911 East Africa IeDEA Consortium. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes Of Health. The research reported in this publication was also supported by the President's Emergency Plan for AIDS Relief (PEPFAR) through the United States Agency for International Development (USAID), under the terms of Cooperative Agreement No. AID-623-A-12-0001, made possible by the joint support of USAID. The contents of this paper are the sole responsibility of AMPATH, and do not necessarily reflect the views of USAID or the United States Government.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

Xuewen Lu and Yan Wang

Department of Mathematics & Statistics

University of Calgary

Calgary, Alberta, T2N 1N4, Canada

E-mail: lux@math.ucalgary.ca

E-mail: yan.wang1@ucalgary.ca

Dipankar Bandyopadhyay

Department of Biostatistics

Virginia Commonwealth University, VA 23298-0032, USA

E-mail: bandyopd@gmail.com

Giorgos Bakoyannis

Department of Biostatistics

Indiana University, Indianapolis, IN, USA

E-mail: gbakogia@iu.edu

References

Bakoyannis, G., Yu, M., and Yiannoutsos, C. T. (2017). Semiparametric regression on cumulative incidence function with interval-censored com-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

peting risks data. *Statistics in Medicine*, 36(23):3683–3707.

Carnicer, J. M. and Peña, J. M. (1993). Shape preserving representations and optimality of the bernstein basis. *Advances in Computational Mathematics*, 1(2):173–196.

Cheng, G., Huang, J. Z., et al. (2010). Bootstrap consistency for general semiparametric m-estimation. *Annals of Statistics*, 38(5):2884–2915.

Crowder, M. J. (2001). *Classical Competing Risks*. CRC Press.

Dabrowska, D. M. and Doksum, K. A. (1988). Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association*, 83(403):744–749.

Egger, M., Ekouevi, D. K., Williams, C., Lyamuya, R. E., Mukumbi, H., Braitstein, P., Hartwell, T., Graber, C., Chi, B. H., Boulle, A., et al. (2012). Cohort Profile: The international epidemiological databases to evaluate AIDS (IeDEA) in sub-Saharan Africa. *International Journal of Epidemiology*, 41(5):1256–1264.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 11(2):89–121.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

Fine, J. (1999). Analysing competing risks data with transformation models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(4):817–830.

Fine, J. P. (2001). Regression modeling of competing crude failure probabilities. *Biostatistics*, 2(1):85–97.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509.

Goodrich, R. K. (1970). A Riesz representation theorem. *Proceedings of the American Mathematical Society*, 24(3):629–636.

Groeneboom, P., Maathuis, M. H., and Wellner, J. A. (2008a). Current status data with competing risks: Limiting distribution of the mle. *Annals of Statistics*, 36(3):1064–1089.

Groeneboom, P., Maathuis, M. H., Wellner, J. A., et al. (2008b). Current status data with competing risks: consistency and rates of convergence of the mle. *Annals of Statistics*, 36(3):1031–1063.

Hu, T. and Xiang, L. (2016). Partially linear transformation cure models

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

for interval-censored data. *Computational Statistics and Data Analysis*, 93:257–269.

Hudgens, M. G., Li, C., and Fine, J. P. (2014). Parametric likelihood inference for interval censored competing risks data. *Biometrics*, 70(1):1–9.

Jeong, J.-H. and Fine, J. (2006). Direct parametric inference for the cumulative incidence function. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(2):187–200.

Jeong, J.-H. and Fine, J. P. (2007). Parametric regression on cumulative incidence function. *Biostatistics*, 8(2):184–196.

Ji, Y., Wang, L., Zhang, H., and Zhou, Y. (2017). Semiparametric estimation of a Box-Cox transformation model with varying coefficients model. *Science China Mathematics*, 60(5):897–922.

Klein, J. P. and Moeschberger, M. L. (2006). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media, 2 edition.

Lawn, S. D., Little, F., Bekker, L.-G., Kaplan, R., Campbel, E., Orrell, C., and Wood, R. (2009). Changing mortality risk associated with CD4

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

cell response to antiretroviral therapy in South Africa. *AIDS (London, England)*, 23(3):335–342.

Li, C. (2016). The fine–gray model under interval censored competing risks data. *Journal of Multivariate Analysis*, 143:327–344.

Lorentz, G. G. (1986). *Bernstein Polynomials*. New York: Chelsea Publishing Co.

Lu, X. and Song, P. X.-K. (2015). Efficient estimation of the partly linear additive hazards model with current status data. *Scandinavian Journal of Statistics*, 42(1):306–328.

Mao, L., Lin, D.-Y., and Zeng, D. (2017). Semiparametric regression analysis of interval-censored competing risks data. *Biometrics*, 73(3):857–865.

Mao, M. and Wang, J.-L. (2010). Semiparametric efficient estimation for a class of generalized proportional odds cure models. *Journal of the American Statistical Association*, 105(489):302–311.

Nguyen, N. and Holodniy, M. (2008). Hiv infection in the elderly. *Clinical Interventions in Aging*, 3(3):453–472.

Panchenko, D. (2003). Symmetrization approach to concentration inequalities for empirical processes. *Annals of Probability*, 31(4):2068–2081.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

Park, J., Bakoyannis, G., and Yiannoutsos, C. T. (2019). Semiparametric competing risks regression under interval censoring using the r package intccr. *Computer Methods and Programs in Biomedicine*, 173:167–176.

Pirrone, V., Libon, D. J., Sell, C., Lerner, C. A., Nonnemacher, M. R., and Wigdahl, B. (2013). Impact of age on markers of HIV-1 disease. *Future Virology*, 8(1):81–101.

Scharfstein, D. O., Tsiatis, A. A., and Gilbert, P. B. (1998). Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime Data Analysis*, 4(4):355–391.

Shen, X. et al. (1997). On methods of sieves and penalization. *Annals of Statistics*, 25(6):2555–2591.

Shen, X. and Wong, W. H. (1994). Convergence rate of sieve estimates. *Annals of Statistics*, 22(2):580–615.

Shi, H., Cheng, Y., and Jeong, J.-H. (2013). Constrained parametric model for simultaneous inference of two cumulative incidence functions. *Biometrical Journal*, 55(1):82–96.

Zaniewski, E., Tymejczyk, O., Kariminia, A., Desmonde, S., Leroy, V.,

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0051

Ford, N., Sohn, A. H., Nash, D., Yotebieng, M., Cornell, M., et al. (2018). IeDEA–WHO Research-Policy Collaboration: Contributing real-world evidence to HIV progress reporting and guideline development. *Journal of Virus Eradication*, 4(Suppl 2):9–15.

Zeng, D., Yin, G., and Ibrahim, J. G. (2006). Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*, 101(474):670–684.

Zhang, H. H., Cheng, G., and Liu, Y. (2011). Linear or nonlinear? Automatic structure discovery for partially linear models. *Journal of the American Statistical Association*, 106(495):1099–1112.

Zhang, Y., Hua, L., and Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the cox model with interval-censored data. *Scandinavian Journal of Statistics*, 37(2):338–354.

Zhou, Q., Hu, T., and Sun, J. (2017). A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data. *Journal of the American Statistical Association*, 112(518):664–672.