Published in final edited form as:

Can J Stat. 2021 September ; 49(3): 731–753. doi:10.1002/cjs.11583.

Evaluation of competing risks prediction models using polytomous discrimination index

Maomao DING¹, Jing NING^{2,*}, Ruosha LI^{3,*}

¹Department of Statistics, Rice University, Houston, TX 77005, U.S.A.

²Department of Biostatistics, the University of Texas MD Anderson Cancer Center, Houston, TX 77030, U.S.A.

³Department of Biostatistics and Data Science, the University of Texas Health Science Center at Houston, Houston, TX 77030, U.S.A.

Abstract

For competing risks data, it is often important to predict a patient's outcome status at a clinically meaningful time point after incorporating the informative censoring due to competing risks. This can be done by adopting a regression model that relates the cumulative incidence probabilities to a set of covariates. To assess the performance of the resulting prediction tool, we propose an estimator of the polytomous discrimination index applicable to competing risks data, which can quantify a prognostic model's ability to discriminate among subjects from different outcome groups. The proposed estimator allows the prediction model to be subject to model misspecification and enjoys desirable asymptotic properties. We also develop an efficient computation algorithm that features a computational complexity of $O(n \log n)$. A perturbation resampling scheme is developed to achieve consistent variance estimation. Numerical results suggest that the estimator performs well under realistic sample sizes. We apply the proposed methods to a study of monoclonal gammopathy of undetermined significance.

Résumé:

Lorsque des données comportent des risques concurrents, il est souvent important de prédire l'issue pour un patient à un temps significatif d'un point de vue clinique après avoir incorporé la censure informative due aux risques concurrents, ce qui peut être fait en adaptant un modèle de régression qui lie les probabilités cumulatives d'incidence à un ensemble de covariables. Pour évaluer la performance des outils de prévision qui en découlent, les auteurs proposent un estimateur de l'indice de discrimination polytomique applicable aux données avec des risques concurrents, et qui permet de quantifier l'habileté d'un modèle de pronostic à discriminer entre les sujets de différents groupes de dénouement. L'estimateur proposé permet au modèle de prévision un certain nombre de mauvaises classifications et exhibe des propriétés asymptotiques souhaitables. Les auteurs développent un algorithme efficace dont la complexité de calcul est d'ordre $O(n \log n)$. Ils proposent un schéma de rééchantillonnage pour obtenir une estimation convergente de la variance et présentent des résultats numériques qui suggèrent que leur estimateur

* Author to whom correspondence may be addressed. jning@mdanderson.org (J. N.), ruosha.li@uth.tmc.edu (R. L.).

Additional Supporting Information may be found in the online version of this article at the publisher's website.

offre de bonnes performances avec des tailles d'échantillons réalistes. Ils appliquent finalement leur méthode à une étude de gammopathie monoclonale dont la significativité n'est pas établie

Key words and phrases:

Competing risks; cumulative incidence; Fine & Gray model; polytomous discrimination index; predictive discrimination; prognostic model

1. INTRODUCTION

Competing risks data emerge when patients are at risk of several terminal events, such that the occurrence of one event precludes the observation of other events. For example, in a study of monoclonal gammopathy of undetermined significance (MGUS), patients may experience progression to a plasma cell malignancy (PCM) or die without experiencing PCM. Development of PCM and death without PCM represent different courses of disease progression and form a competing risks structure. To design better management strategies for the patients, an important step is to build a prognostic tool that assesses the risk of each event. The risks of the events are often quantified by the covariate-dependent cumulative incidence function (CIF), namely the probability that a specific event occurs by time t in the presence of other competing events, given covariates.

Several regression strategies have been developed to facilitate regression modelling and prediction for the CIF. For example, Fine & Gray (1999) proposed a proportional hazards model for the subdistribution hazard function, which has a direct relationship with the CIF. The modelling framework is similar to a Cox proportional hazards model for standard survival data and enjoys convenient implementation and neat asymptotic properties. Andersen, Klein & Rosthøj (2003) and Klein & Andersen (2005) adopted the pseudo-observation approach for the cumulative incidence regression. To capture the time-varying effects of covariates, Scheike, Zhang & Gerds (2008) proposed estimating CIF using a direct binomial regression. More recently, Bellach et al. (2019) estimated the CIF under a nonparametric maximum likelihood estimation framework.

These regression models, once fitted to the data, can readily facilitate a risk prediction tool based on the CIF. However, before such a prediction tool can be put into practical use, one first needs to evaluate its predictive performance. Research efforts have been devoted to assessing the predictive performance of a competing risks regression model. Schoop et al. (2011) extended the Brier score, a metric of predictive accuracy, to examine the prediction error of the CIF predictor. For assessing predictive discrimination, one needs to assess how well a prediction tool separates the cases, namely those who experience a specific event of interest, from the controls. For example, Saha & Heagerty (2010) used the group of patients who remained free of any event by time t as the control and studied the time-dependent ROC curve. An alternative definition of the control group was proposed in Zheng et al. (2012), which includes those who fail from other competing risks events by time t , as well as the event-free patients. Blanche, Dartigues & Jacqmin-Gadda (2013) proposed estimators of the area under curve (AUC) for each definition of the controls using the technique of inverse probability weighting. Wu & Li (2018) studied alternative estimators of these prediction

metrics using nonparametric weighting. Wolbers et al. (2014) extended the concordance index (Harrell et al., 1982; Uno et al., 2011) to the competing risks setting and proposed a consistent estimator.

The focus of this work is to examine the predictive discrimination of a competing risks prediction model by recognizing the polytomous nature of the outcome. The competing risks outcome, as of a clinically meaningful time point t , can be viewed as a polytomous outcome where the outcome categories include the specific failure types, as well as the event-free type. In many clinical scenarios, these different outcome categories have disparate clinical implications and require different healthcare needs. It is thus important to assess whether a prediction tool can differentiate among patients who fall into different outcome categories. In assessing prediction performance for a specific type of failure, we anticipate that a powerful prediction tool can discriminate the corresponding cases from each of the remaining outcome groups.

For uncensored data, Van Calster et al. (2012) extended the AUC for binary outcomes to the polytomous outcomes as the polytomous discrimination index (PDI). The category-specific PDI is defined as the probability of correctly identifying the case of a specific category among a set of k cases, each chosen from one category. The overall PDI is defined as average over k category-specific PDIs. The PDI reduces to the standard AUC for outcomes with two categories. The k category-specific PDIs inform discriminatory performance for individual outcome categories, while the overall PDI gives us an overall measure indicating the discriminatory ability of the polytomous model in all categories. More recently, Li et al. (2018) studied nonparametric estimation and inference for PDI. In de Jong et al. (2019), the authors used PDI to measure the performance of multinomial logistic regression.

In this article, we extend PDI to assess the prediction performance of a competing risks prediction tool. In Section 2, we provide the formal definition of PDI for competing risks data and develop an estimator based on the technique of inverse probability of censoring weighting. Large sample properties of PDI and the method for variance estimation are given in the subsequent subsections, where we allow the regression model that underlies the prediction tool to be subject to model misspecification. We design an algorithm that accelerates the computation of PDI estimation drastically to log-linear time. In Section 3, we perform simulation studies to demonstrate the finite sample behaviour of PDI under various scenarios. We apply PDI to the monoclonal gammopathy data in Section 4 and provide a brief discussion in Section 5. The details of proofs are provided in the Appendix.

2. METHOD

2.1. Competing Risks Data and Model

Let T and $\varepsilon \in \{1, \dots, K\}$ be the failure time of the earliest event and the corresponding cause, respectively. Let $\mathbf{Z} = (Z_1, Z_2, \dots, Z_d)^T$ denote a d -dimensional vector of baseline covariates. The failure time T may be censored by an independent random variable C . Define $X = T \wedge C$ and $\delta = \mathbb{I}(T \leq C)$. The observed data consist of n i.i.d. replicates of $\{X, \varepsilon, \mathbf{Z}\}$.

Let τ denote a certain time point of interest, which can be prespecified according to scientific interest. For example, 1- and 5-year relapse and overall survival rates are often of interest in cancer studies. One may also examine multiple τ 's for more comprehensive insights. At a specific τ , a subject's status falls into one of the following categories: $\mathbb{I}(T \leq \tau, \varepsilon = 1)$, ..., $\mathbb{I}(T \leq \tau, \varepsilon = K)$, or $\mathbb{I}(T > \tau)$. For the purpose of disease prognosis, it is often important to predict which category a patient will fall into, based on a set of baseline covariates $\mathbf{Z}$. The probabilities of these categories are given by the multinomial probability vector $\boldsymbol{\mu}(\tau; \mathbf{Z}) \equiv \{F_1(\tau; \mathbf{Z}), F_2(\tau; \mathbf{Z}), \dots, F_K(\tau; \mathbf{Z}), S(\tau; \mathbf{Z})\}$, where $F_l(\tau; \mathbf{Z}) = \mathbb{P}(T \leq \tau, \varepsilon = l | \mathbf{Z})$ is the Cause l CIF, $l = 1, 2, \dots, K$, and $S(\tau; \mathbf{Z}) \equiv 1 - \sum_{l=1}^K F_l(\tau; \mathbf{Z})$ gives the probability that the subject remains event-free at τ . The vector $\boldsymbol{\mu}(\tau; \mathbf{Z})$ facilitates direct interpretation and is clinically important in many disease settings.

For estimation and prediction of $\boldsymbol{\mu}(\tau; \mathbf{Z})$, one may adopt a sensible parametric or semiparametric regression model for competing risks data. Let θ_l denote a set of unknown parameters for Cause l failure, and let $\hat{\theta}_l$ be its estimated counterpart. Estimated parameters from a competing risks regression model often allow one to construct a predictor of $F_l(\tau; \mathbf{Z})$ for any given $\mathbf{Z}$, and we denote this predictor by $m_l(\tau; \mathbf{Z}, \hat{\theta}_l)$. For example, the Fine & Gray model (Fine & Gray, 1999) assumes that

$$\lambda_l(t; \mathbf{Z}) = \lambda_{l0}(t) \exp(\mathbf{Z}_l^T \boldsymbol{\beta}_l), \quad (1)$$

where $\mathbf{Z}_l$ is a subset of $\mathbf{Z}$, or $\mathbf{Z}$ itself, that enters the regression model for the Cause l failure, $\lambda_l(t; \mathbf{Z})$ is the conditional subdistribution hazard function, $\lambda_{l0}(t)$ is an unspecified baseline subdistribution hazard function, and $\boldsymbol{\beta}_l$ is the unknown vector of regression coefficients. Let $\Lambda_{l0}(\tau) = \int_0^\tau \lambda_{l0}(s) ds$ and $\boldsymbol{\theta}_l = \{\boldsymbol{\beta}_l^T, \Lambda_{l0}(\tau)\}^T$. Under model (1), the Cause l CIF, $F_l(\tau; \mathbf{Z})$, can be estimated by $m_l(\tau; \mathbf{Z}, \hat{\theta}_l)$, where $m_l(\tau; \mathbf{Z}, \boldsymbol{\theta}_l) \equiv 1 - \exp\{-\Lambda_{l0}(\tau) \exp(\mathbf{Z}_l^T \boldsymbol{\beta}_l)\}$. For the simplicity of notations, we omit the subscript l in $\mathbf{Z}_l$ below, with the understanding that different subsets of $\mathbf{Z}$ can enter the regression models for different failure types. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_K^T)^T$ and $m_{K+1}(\cdot; \mathbf{Z}, \boldsymbol{\theta}) = 1 - \sum_{l=1}^K m_l(\cdot; \mathbf{Z}, \boldsymbol{\theta}_l)$. The Fine & Gray model entails an estimator of $\boldsymbol{\mu}(\tau; \mathbf{Z})$ in the form of $\mathbf{m}(\tau; \mathbf{Z}, \hat{\boldsymbol{\theta}}) = \{m_1(\tau; \mathbf{Z}, \hat{\theta}_1), \dots, m_K(\tau; \mathbf{Z}, \hat{\theta}_K), m_{K+1}(\tau; \mathbf{Z}, \hat{\boldsymbol{\theta}})\}$.

Other competing risks regression models, such as the direct binomial regression model (Scheike, Zhang & Gerds, 2008) and parametric regression models (Jeong & Fine, 2006), also provide estimators of $\boldsymbol{\mu}(\tau; \mathbf{Z})$ based on the model parameters. For the purpose of illustration, we focus on the $\mathbf{m}(\tau; \mathbf{Z}, \hat{\boldsymbol{\theta}})$ from the Fine & Gray model here and in the sequel, but our methods below also accommodate the predictors from other competing risks regression models.

2.2. Evaluating Competing Risks Model Using the PDI

Our goal is to evaluate the performance of a model-based predictor $\mathbf{m}(\tau; \mathbf{Z}, \hat{\boldsymbol{\theta}})$ in predicting the τ -year outcome, $\mathbb{I}(T \leq \tau, \varepsilon = 1)$, ..., $\mathbb{I}(T \leq \tau, \varepsilon = K)$, and $\mathbb{I}(T > \tau)$. For broader applicability, we allow the regression model that underlies $\mathbf{m}(\tau; \mathbf{Z}, \hat{\boldsymbol{\theta}})$ to be subject to model

misspecification, by viewing the model as a working model instead of a true model. Despite the potential model misspecification, it is realistic to assume that $\hat{\theta}$ converges in probability to a deterministic value, denoted by θ_0 , when sample size increases, and that $\mathbf{m}(\tau; \mathbf{Z}, \hat{\theta})$ converges to a deterministic function $\mathbf{m}(\tau; \mathbf{Z}, \theta_0)$ uniformly in $\mathbf{Z}$. In the sequel, we always treat the regression model as a working model to accommodate potential model misspecification, though all methods and theories hold for the special case where the working model is a true model. We shall verify these under a working Fine & Gray model, and these assumptions are expected to be satisfied under most existing parametric and semi-parametric competing risks regression models.

To quantify the performance of the model-based predictors $m_l(\tau; \mathbf{Z}, \hat{\theta}_l)$ and $\mathbf{m}(\tau; \mathbf{Z}, \hat{\theta})$, we extend the PDI for multinomial outcomes (Van Calster et al., 2012) to the setting of competing risks. We first consider a cause-specific PDI, which targets the ability of a prediction tool to differentiate patients with Cause l failure by time τ from those with other outcomes. We define a τ -year cause-specific PDI as

$$\text{PDI}_l(\tau; \theta) = \mathbb{E} \left[\prod_{j=1, j \neq l}^{K+1} I \left\{ m_l(\tau; \mathbf{Z}_{i_l}, \theta) > m_l(\tau; \mathbf{Z}_{i_j}, \theta) \right\} \right] \Big| \mathcal{A}, \quad (2)$$

where $\mathcal{A} = \left\{ T_{i_1,1}^* \leq \tau, \dots, T_{i_K,K}^* \leq \tau, T_{i_{K+1}}^* > \tau \right\}$, and $T_{i,l}^* = I(\varepsilon_i = l) \cdot T_i + I(\varepsilon_i \neq l) \cdot \infty$ is an improper event time variable which has a nonzero probability mass at infinity. The indicator $I(T_{i,l}^* \leq \tau)$ is equivalent to $I(T_i \leq \tau, \varepsilon_i = l)$. Subjects $i_1, \dots, i_{K+1}$ are randomly selected, each from one of the following outcome categories: $\{T \leq \tau, \varepsilon = 1\}, \dots, \{T \leq \tau, \varepsilon = K\}$, and $\{T > \tau\}$, where subject i_j is the one that experiences the Cause l failure by time τ . For a given θ , $\text{PDI}_l(\tau; \theta)$ is the probability that, among $K+1$ randomly selected subjects $i_1, \dots, i_{K+1}$ from distinct outcome groups, the subject who actually experiences the Cause l failure (i.e., subject i_l) has the largest model-based Cause l risk prediction. A PDI_l value of $(K+1)^{-1}$ corresponds to a random predictor for Cause l . If the PDI_l value is below $(K+1)^{-1}$, the model performance is worse than random guessing. A value above $(K+1)^{-1}$ suggests that the predictive model has a positive capability to discriminate the Cause l case among patients with different outcomes, where a value of 1 corresponds to perfect predictiveness.

In some situations, one may encounter ties in the predicted risks. Following Van Calster et al. (2012), we divide $\text{PDI}_l(\tau; \theta)$ by the number of subjects who have the same predicted CIF as that of subject i_l . This is equal to the probability of selecting subject i_l if we randomly select from subjects that share the same predicted CIF as i_l . Formally speaking, the modified $\text{PDI}_l(\tau; \theta)$ in the presence of ties is defined as

$$\left[\frac{\prod_{j=1, j \neq l}^{K+1} I \left\{ m_l(\tau; \mathbf{Z}_{i_l}, \theta) \geq m_l(\tau; \mathbf{Z}_{i_j}, \theta) \right\}}{1 + \sum_{j=1, j \neq l}^{K+1} I \left\{ m_l(\tau; \mathbf{Z}_{i_l}, \theta) = m_l(\tau; \mathbf{Z}_{i_j}, \theta) \right\}} \right] \Big| \mathcal{A}.$$

Observe that

$$\sum_{k=1}^{K+1} \frac{\prod_{j=1, j \neq k}^{K+1} I \left\{ m_l(\tau; \mathbf{Z}_{i_k}, \boldsymbol{\theta}) \geq m_l(\tau; \mathbf{Z}_{i_j}, \boldsymbol{\theta}) \right\}}{1 + \sum_{j=1, j \neq k}^{K+1} I \left\{ m_l(\tau; \mathbf{Z}_{i_k}, \boldsymbol{\theta}) = m_l(\tau; \mathbf{Z}_{i_j}, \boldsymbol{\theta}) \right\}} = 1,$$

with probability 1, and when the prognostic model is noninformative,

$$\mathbb{E} \left[\frac{\prod_{j=1, j \neq k}^{K+1} I \left\{ m_l(\tau; \mathbf{Z}_{i_k}, \boldsymbol{\theta}) \geq m_l(\tau; \mathbf{Z}_{i_j}, \boldsymbol{\theta}) \right\}}{1 + \sum_{j=1, j \neq k}^{K+1} I \left\{ m_l(\tau; \mathbf{Z}_{i_k}, \boldsymbol{\theta}) = m_l(\tau; \mathbf{Z}_{i_j}, \boldsymbol{\theta}) \right\}} \right]_{\mathcal{A}}$$

is constant over k . Thus, the null value of PDI_l with ties is still $(K+1)^{-1}$.

To quantify the overall discriminatory performance of a prognostic model in predicting the τ -year outcome, the overall PDI is defined as

$$\text{PDI}(\tau; \boldsymbol{\theta}) = (K+1)^{-1} \sum_{l=1}^{K+1} \text{PDI}_l(\tau; \boldsymbol{\theta}), \quad (3)$$

which combines the cause-specific PDI's by assigning equal weights to each outcome category. Alternative forms of the overall PDI can be defined, for example, by weighing the outcome categories using the inverse of their prevalence (Van Calster et al., 2012). Without loss of generality, we shall focus on the definition in Equation (3) of this article.

2.3. Estimation

In most cases, we do not know the limiting value $\boldsymbol{\theta}_0$ in advance. This requires us to take an additional step to estimate $\boldsymbol{\theta}_0$ under the working model. We adopt a cross-validation (CV) scheme to mitigate the risk of overestimating model performance. Specifically, we split the observed data into two parts, namely $\mathcal{D}_L$ and $\mathcal{D}_V$, which are used to estimate model parameters and PDI, respectively. For the i -th individual, let $v_i = (v_{i1}, v_{i2})$ be the binary allocation variable, where v_{i1} indicates whether observation i is used to estimate $\boldsymbol{\theta}_0$ under the working model, and v_{i2} indicates whether subject i is used to compute the PDI based on $\hat{\boldsymbol{\theta}}$. Without loss of generality, we consider twofold CV, where we sample v_{i1} from i.i.d. Bernoulli distributions with probability 0.5 and set $v_{i2} = 1 - v_{i1}$.

For the purpose of illustration, we introduce the estimation procedure for $\text{PDI}_l(\tau, \boldsymbol{\theta})$ with $K = 2$ and $l = 1$. Denote $I_{ijk} = I(X_i = \tau, i \neq j = 1, X_j = \tau, j \neq k = 2, X_k > \tau)$ and $c_{ijk}(\boldsymbol{\theta}) = I\{m_1(\tau, \mathbf{Z}_i, \boldsymbol{\theta}) > \max\{m_1(\tau, \mathbf{Z}_j, \boldsymbol{\theta}), m_1(\tau, \mathbf{Z}_k, \boldsymbol{\theta})\}\}$. When ties are present, one may let

$$c_{ijk}(\boldsymbol{\theta}) = \frac{I\{m_1(\tau; \mathbf{Z}_i, \boldsymbol{\theta}) \geq \max\{m_1(\tau; \mathbf{Z}_j, \boldsymbol{\theta}), m_1(\tau; \mathbf{Z}_k, \boldsymbol{\theta})\}\}}{1 + I\{m_1(\tau; \mathbf{Z}_i, \boldsymbol{\theta}) = m_1(\tau; \mathbf{Z}_j, \boldsymbol{\theta})\} + I\{m_1(\tau; \mathbf{Z}_i, \boldsymbol{\theta}) = m_1(\tau; \mathbf{Z}_k, \boldsymbol{\theta})\}}.$$

Next, let $\hat{G}_V(x | \mathbf{Z})$ be the estimator of $G(x|\mathbf{Z}) = P(C \leq x|\mathbf{Z})$, constructed using the observations in $\mathcal{D}_V$. When censoring times are covariate dependent, one may obtain $\hat{G}_V(\cdot | \mathbf{Z})$ using an appropriate regression model, such as a Cox proportional hazards model. Covariate-dependent censoring sometimes arises in medical research. For example, in clinical research, certain treatment groups may have a higher censoring rate than other groups. In these situations, factor(s) that affect the censoring distribution need to be taken into account in the estimation of $G(\cdot|\mathbf{Z})$. Otherwise, one can simply adopt the Kaplan–Meier estimator when C is independent of both (T, ϵ) and $\mathbf{Z}$. Using the technique of inverse probability of censoring weighting (Robins, Rotnitzky & Zhao, 1995), we propose to estimate $\text{PDI}_1(\tau, \theta_0)$ as

$$\widehat{\text{PDI}}_1^{(1)}(\tau; \hat{\theta}) = \frac{\sum_{ijk} C_{ijk}(\hat{\theta}) I_{ijk} v_{i2} v_{j2} v_{k2} / \hat{G}_V(X_i | \mathbf{Z}_i) \hat{G}_V(X_j | \mathbf{Z}_j) \hat{G}_V(\tau | \mathbf{Z}_k)}{\sum_{ijk} I_{ijk} v_{i2} v_{j2} v_{k2} / \hat{G}_V(X_i | \mathbf{Z}_i) \hat{G}_V(X_j | \mathbf{Z}_j) \hat{G}_V(\tau | \mathbf{Z}_k)}. \quad (4)$$

To make effective use of the data, we next trade the role of $\mathcal{D}_L$ and $\mathcal{D}_V$, then implement the procedure above again to obtain $\widehat{\text{PDI}}_1^{(2)}(\tau, \hat{\theta})$. Our CV type estimator is defined as

$$\widehat{\text{PDI}}_1(\tau; \hat{\theta}) = \left\{ \widehat{\text{PDI}}_1^{(1)}(\tau; \hat{\theta}) + \widehat{\text{PDI}}_1^{(2)}(\tau; \hat{\theta}) \right\} / 2.$$

In practice, to improve the finite sample performance, one may repeat the CV procedure for $\mathcal{B}$ times (e.g., $\mathcal{B} = 3$), and the estimator is defined as the average over all the repetitions. The choice of $\mathcal{B}$ has no effect on the large sample properties of $\widehat{\text{PDI}}_1(\tau)$, but conducting repetitions can provide some improvement in finite samples.

2.4. Fast Computation of PDI

Brute force computations of $\widehat{\text{PDI}}_1(\tau; \hat{\theta}_l)$ and $\widehat{\text{PDI}}_1(\tau; \hat{\theta})$ based directly on the summation form in Equation (4) would result in a computation complexity of $O(n^{K+1})$, which is unfavourable when the dataset size is moderate to large. The computation cost becomes even more burdensome when one wishes to repeat the cross validation multiple times and/or conduct resampling. To tackle this, we design a computation algorithm that reduces the computation cost drastically to $O(n \log n)$ in time. The proposed algorithm is detailed in Algorithm 1 below, where we let $K = 2$ and $l = 1$ for the purpose of illustration. We categorize subjects into three groups, $\mathbb{I}_1 = \{i: X_i \leq \tau, \Delta_i \epsilon_i = 1, v_{i2} = 1\}$, $\mathbb{I}_2 = \{i: X_i \leq \tau, \Delta_i \epsilon_i = 2, v_{i2} = 1\}$, and $\mathbb{I}_3 = \{i: X_i > \tau, v_{i2} = 1\}$, and use $n_i, i = 1, 2, 3$ to denote the number of subjects in each group. We also denote $\hat{\mathbf{G}} = \{\hat{G}_V(X_1 | \mathbf{Z}_1), \dots, \hat{G}_V(X_n | \mathbf{Z}_n)\}^T$, $\hat{\mathbf{G}}_\tau = \{\hat{G}_V(\tau | \mathbf{Z}_1), \dots, \hat{G}_V(\tau | \mathbf{Z}_n)\}^T$ and $\hat{\mathbf{m}}_1 = \{\hat{m}_1(\tau; \mathbf{Z}_1, \hat{\theta}), \dots, \hat{m}_1(\tau; \mathbf{Z}_n, \hat{\theta})\}^T$. For a vector $\mathbf{a}$, let $\mathbf{a}[i]$ denote its i_{th} element.

To see how we can reformulate the computation of PDI, we first rewrite the triple summation in the numerator of (4) as

$$\sum_{ijk} C_{ijk}(\hat{\theta}) I_{ijk} v_{i2} v_{j2} v_{k2} / \hat{G}_V(X_i | \mathbf{Z}_i) \hat{G}_V(X_j | \mathbf{Z}_j) \hat{G}_V(\tau | \mathbf{Z}_k) = \sum_{i \in \mathbb{I}_1} \sum_{j \in \mathbb{I}_2} \sum_{k \in \mathbb{I}_3} C_{ijk}(\hat{\theta}) / \hat{G}_V(X_i | \mathbf{Z}_i) \hat{G}_V(X_j | \mathbf{Z}_j) \hat{G}_V(\tau | \mathbf{Z}_k).$$

Next, we observe that,

$$\sum_{j \in \mathbb{I}_2} \sum_{k \in \mathbb{I}_3} C_{ijk}(\hat{\theta}) / \hat{G}_V(X_i | \mathbf{Z}_i) \hat{G}_V(X_j | \mathbf{Z}_j) \hat{G}_V(\tau | \mathbf{Z}_k) = \hat{G}_V(X_i | \mathbf{Z}_i)^{-1} \cdot \sum_{j \in \mathbb{I}_2, \hat{\mathbf{m}}_1[j] < \hat{\mathbf{m}}_1[i]} \hat{G}_V(X_j | \mathbf{Z}_j)^{-1} \cdot \sum_{k \in \mathbb{I}_3, \hat{\mathbf{m}}_1[k] < \hat{\mathbf{m}}_1[i]} \hat{G}_V(\tau | \mathbf{Z}_k)^{-1}.$$

Thus for any fixed i , instead of doing a double summation, we can first sort out the elements from $\mathbb{I}_2$ and $\mathbb{I}_3$ and then conduct two single summations. Finally, after sorting $\mathbb{I}_1$ in descending order according to the corresponding values in $\hat{\mathbf{m}}_1$, the sets $\{j: j \in \mathbb{I}_2, \hat{\mathbf{m}}_1[j] < \hat{\mathbf{m}}_1[i]\}$ and $\{k: k \in \mathbb{I}_3, \hat{\mathbf{m}}_1[k] < \hat{\mathbf{m}}_1[i]\}$ are also descending in the number of elements as we loop over $i \in \mathbb{I}_1$. If $\mathbb{I}_2$ and $\mathbb{I}_3$ are also sorted in descending order, we could easily update the two sets in computation by examining the first few elements and excluding those that exceed $\hat{\mathbf{m}}_1[i]$.

Algorithm 1 facilitates fast computation of PDI. For example, when the sample size is 100,000, we can compute $\widehat{\text{PDI}}_1(\tau; \hat{\theta}_1)$ in around 0.1s, on a desktop computer with an Intel i5–4590 CPU and 16.0GB RAM. This indicates that our proposed statistic can be scaled up well to handle large datasets. The proposed algorithm can be easily extended to accommodate data with an arbitrary number of causes of competing risks, with the same computational complexity.

Algorithm 1

Computation of $\widehat{\text{PDI}}_1(\tau; \hat{\theta}_1)$

```

1:  if  $n_1 \cdot n_2 \cdot n_3 = 0$  then
2:      Return NA
3:  else
4:      Sort  $\mathbb{I}_1, \mathbb{I}_2, \mathbb{I}_3$  in descending order according to the corresponding values in  $\hat{\mathbf{m}}_1$ . Denote the corresponding
      index vectors as  $\mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3$ .
5:      Set  $i = 1, j = 1, k = 1, count = 0, S_1 = \sum_{i \in \mathbb{I}_1} \hat{\mathbf{G}}[i]^{-1}, S_2 = \sum_{i \in \mathbb{I}_2} \hat{\mathbf{G}}[i]^{-1}, S_3 = \sum_{i \in \mathbb{I}_3} \hat{\mathbf{G}}[i]^{-1}$ .
      Let  $denom = S_1 \cdot S_2 \cdot S_3$ .
6:      while  $i \leq n_1$  do
7:          while  $\hat{\mathbf{m}}_1[\mathbf{I}_2, j] > \hat{\mathbf{m}}_1[\mathbf{I}_1, i]$  and  $j < n_2$  do

```



```

8:       $S_2 = S_2 - \widehat{\mathbf{G}}[\mathbf{I}_{2,j}]^{-1}$ 
9:       $j = j + 1$ 
10:     while  $\widehat{\mathbf{m}}_1[\mathbf{I}_{3,k}] > \widehat{\mathbf{m}}_1[\mathbf{I}_{1,i}]$  and  $k < n_3$  do
11:          $S_3 = S_3 - \widehat{\mathbf{G}}_\tau[\mathbf{I}_{3,j}]^{-1}$ 
12:          $k = k + 1$ 
13:     if  $j > n_2$  or  $k > n_3$  then
14:         break
15:      $count = count + \widehat{\mathbf{G}}[\mathbf{I}_{1,i}]^{-1} \cdot S_2 \cdot S_3$ 
16:      $i = i + 1$ 
17: Return  $count/denom$ 

```

2.5. Theoretical Properties

We assume the following conditions for the consistency and asymptotic normality of the proposed estimators.

(C1) (i) Let n_k , $k = 1, 2, \dots, K+1$ denote the sample size in the k_{th} outcome group, where $\sum_{k=1}^{K+1} n_k = n$. There exists a constant $0 < L_c < 1$, such that

$\lim_{n \rightarrow \infty} \min_{k=1}^{K+1} n_k/n > L_c$. (ii) The estimated parameters $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$ when n increases, where $\boldsymbol{\theta}_0$ is a deterministic vector, and $\sup_{\mathbf{z}} \|\mathbf{m}(\tau; \mathbf{z}, \widehat{\boldsymbol{\theta}}) - \mathbf{m}(\tau; \mathbf{z}, \boldsymbol{\theta}_0)\| \xrightarrow{P} 0$.

(C2) $|m(\tau; \mathbf{Z}_1, \boldsymbol{\theta}_0) - m(\tau; \mathbf{Z}_2, \boldsymbol{\theta}_0)| \leq L_0 \|\mathbf{Z}_1 - \mathbf{Z}_2\|$ for some constant L_0 .

(C3) $\mathbb{P}\{m(\tau; \mathbf{Z}_1, \boldsymbol{\theta}_0) = m(\tau; \mathbf{Z}_2, \boldsymbol{\theta}_0) \mid \mathbf{Z}_1 \neq \mathbf{Z}_2\} = 0$.

(C4) (i) $\widehat{\boldsymbol{\theta}}_l$ satisfies $\sqrt{n}(\widehat{\boldsymbol{\theta}}_l - \boldsymbol{\theta}_{l0}) = n^{-1/2} \sum_{i=1}^n 2v_{il} \boldsymbol{\theta}_{il} + o_p(1)$, where each $\boldsymbol{\theta}_{il}$ is an i.i.d. influence function of mean 0. (ii) Let $m'_l(\tau; \mathbf{z}, \boldsymbol{\theta}_l) = \frac{\partial}{\partial \boldsymbol{\theta}_l} m(\tau; \mathbf{z}, \boldsymbol{\theta}_l)$, then $m'_l(\tau; \mathbf{z}, \boldsymbol{\theta}_{l0})$ is uniformly bounded in $\mathbf{z}$.

Condition (C1) requires that the sample sizes in each outcome categories are of the same order as the total sample size n and the consistency of the estimated model parameters. We shall verify the consistency condition under a working Fine & Gray model. Condition (C2) requires that the model-based predictor is Lipschitz continuous in $\mathbf{Z}$, implying that two subjects with similar covariate values would have similar predicted probabilities. Condition (C3) requires that two subjects with different covariate values do not have identical predicted probabilities, which would be satisfied if the elements in $\boldsymbol{\beta}_{l0}$ are not exactly zero. Condition (C4) requires the asymptotic normality of the estimated parameters, which is satisfied by commonly used competing risks models, such as the Fine & Gray model and the direct binomial model. A direct implication of (C4) is that

$$\sqrt{n}\{m_l(\tau; \mathbf{z}, \hat{\boldsymbol{\theta}}_l) - m_l(\tau; \mathbf{z}, \boldsymbol{\theta}_{l0})\} = n^{-1/2} \sum_{i=1}^n 2v_{il} \psi_{il}(\mathbf{z}) + o_p(1),$$

with $\psi_{il}(\mathbf{z}) = m'_l(\tau; \mathbf{z}, \boldsymbol{\theta}_{l0}) \boldsymbol{\vartheta}_{il}$.

Theorem 1. Under conditions (C1)–(C3),

$$|\widehat{PDI}(\tau; \hat{\boldsymbol{\theta}}) - PDI_l(\tau; \boldsymbol{\theta}_0)| \xrightarrow{P} 0.$$

The consistency of $\widehat{PDI}_l(\tau; \hat{\boldsymbol{\theta}})$ can be established from the uniform consistency of U-processes and uniform convergence of $m_l(\tau; \cdot, \hat{\boldsymbol{\theta}})$. Under condition (C4) on the asymptotic distribution $\hat{\boldsymbol{\theta}}$, the asymptotic normality of $\widehat{PDI}_l(\tau; \hat{\boldsymbol{\theta}})$ can also be established.

Theorem 2. Under conditions (C1)–(C4), $\sqrt{n}\{\widehat{PDI}_l(\tau; \hat{\boldsymbol{\theta}}) - PDI_l(\tau; \boldsymbol{\theta}_0)\}$ is asymptotically normal.

The asymptotic normality can be proven by the application of martingale representation of $\hat{G}_V$ and other algebraic manipulations, along with the central limit theorem for U -processes and a functional central limit theorem. The conditions required above are satisfied by a variety of models, such as the Fine & Gray model and the direct binomial model. Specifically, when the Fine & Gray model is correctly specified, its asymptotic behaviour is well studied (Fine & Gray, 1999), and it is easy to check that conditions (C1)–(C4) are readily satisfied. When the model is subject to misspecification, the desired properties can still be satisfied under mild conditions. Write $\lambda_{i\ell}(t)$ as the true subdistribution hazard function for Cause ℓ failure of subject i , and $Y_{i\ell}(t) = 1 - I(T_i < t, \delta_i = \ell)$. Denote

$$s_l^{(p)}(\boldsymbol{\beta}_l, u) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y_{il}(u) \mathbf{Z}_i^{\otimes p} \exp(\mathbf{Z}_i^T \boldsymbol{\beta}_l),$$

$$s_l^{(p)}(u) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y_{il}(u) \lambda_{il}(u) \mathbf{Z}_i^{\otimes p}.$$

Define $\boldsymbol{\beta}_0$ as the solution to the equation $h_\ell(\boldsymbol{\beta}_\ell) = 0$ with

$$h_\ell(\boldsymbol{\beta}_\ell) = \int_0^\infty s_l^{(1)}(u) G(u) du - \int_0^\infty \frac{s_l^{(1)}(\boldsymbol{\beta}_\ell, u)}{s_l^{(0)}(\boldsymbol{\beta}_\ell, u)} s_l^{(0)}(u) G(u) du.$$

The limiting cumulative subdistribution hazard is defined as $\Lambda_{l0}(\tau) = \int_0^\tau s_l^{(0)}(u) s_l^{(0)}(\beta_{l0}, u) du$.

Write $\theta_0 = (\theta_{l0}^T, \dots, \theta_{K0}^T)^T$, where $\theta_{l0} = \{\beta_{l0}^T, \Lambda_{l0}(\tau)\}^T$. Then we have

Lemma 1. *Under regularity conditions (C5)–(C7) in the Appendix, $\|\hat{\theta} - \theta_0\| \xrightarrow{P} 0$ and $\|m_l(\tau; \cdot, \hat{\theta}) - m_l(\tau; \cdot, \theta_0)\| \xrightarrow{P} 0$. What's more, condition (C4) on asymptotic normality is satisfied under a working Fine & Gray model.*

2.6. Variance Estimation

To depict the precision of the point estimator, it is often important to conduct variance and interval estimation for an estimated prediction measure. It is noteworthy that the variability of the proposed $\widehat{\text{PDI}}_l(\tau; \hat{\theta}_l)$ can be attributed to both the U -statistic type summations, using the testing subset and the estimated $\hat{\theta}_l$ based on the training subset. However, the second source of variability poses challenges to direct variance estimation, due to the fact that $\hat{\theta}_l$ is involved in a nondifferentiable indicator function. This difficulty necessitates a resampling-based variance estimation procedure. The traditional bootstrap method is not preferred here, as it may lead to many ties and sparse counts in certain outcome groups. Following Uno et al. (2007), we design a perturbation-resampling scheme to consistently estimate the variance of $\widehat{\text{PDI}}_l(\tau; \hat{\theta})$.

Let $\{V_i\}_{i=1}^n$ be n independent copies of positive random variable V with unit mean and variance. To account for the variability in $\hat{\theta}$, we first propose the perturbed score function as

$$U_l^*(\beta) = \sum_{i=1}^n V_i \int_0^\infty v_{il} \left[\mathbf{Z}_i - \frac{\mathbf{S}_l^{(1*)}(\beta_l, u)}{\mathbf{S}_l^{(0*)}(\beta_l, u)} \right] w_i^*(u) dN_{il}(u),$$

where $N_{il}(t) = I(T_i \leq t, \varepsilon_i = l)$ and

$$\mathbf{S}_l^{(p*)}(\beta_l, u) \equiv \frac{1}{n\kappa_n} \sum_{i=1}^n V_i v_{il} w_i^*(u) Y_{il}(u) \mathbf{Z}_i^{\otimes p} \exp(\mathbf{Z}_i^T \beta_l),$$

with $\kappa_n = n^{-1} \sum_{i=1}^n v_{il}$. Here, for simplicity, we assume that C is independent of T and ε ,

and that $G(\cdot)$ is free of $\mathbf{Z}$. The perturbed inverse weight is

$w_i^*(t) = I(C_i \geq T_i \wedge t) G_L^*(t) / G_L^*(X_i \wedge t)$, and

$$G_L^*(t) = \prod_{s \leq t} \left[1 - \frac{\sum_{i=1}^n V_i v_{il} I(X_i = s, \Delta_i = 0)}{\sum_{i=1}^n V_i v_{il} \{1 - I(X_i \leq s, \Delta_i = 0)\}} \right] \quad (5)$$

is the perturbed KM estimator of $G(t)$. Denote the zero-crossing of $U_l^*(\beta)$ as $\hat{\beta}_l^*$. We also obtain a perturbed estimate of Λ_{l0} as

$$\hat{\Lambda}_{i0}^*(t) = \frac{1}{n\kappa} \sum_{i=1}^n \int_0^t \frac{V_i v_{i1} w_i^*(u)}{s_l^{(0*)}(\hat{\beta}_l^*, u)} dN_{il}(u).$$

Finally, the perturbed $\widehat{\text{PDI}}_l^{(1*)}$ is calculated as

$$\widehat{\text{PDI}}_l^{(1*)}(\tau; \hat{\theta}_l^*) = \frac{\sum_{ijk} V_i V_j V_k C_{ijk}(\hat{\theta}_l^*) I_{ijk} v_{i2} v_{j2} v_{k2} / G_V^*(X_i) G_V^*(X_j) G_V^*(\tau)}{\sum_{ijk} V_i V_j V_k I_{ijk} v_{i2} v_{j2} v_{k2} / G_V^*(X_i) G_V^*(X_j) G_V^*(\tau)},$$

where $\hat{\theta}_l^* = \{\hat{\beta}_l^{*T}, \hat{\Lambda}_{i0}^*(\tau)\}^T$, and $G_V^*(t)$ is obtained by replacing the v_{i1} in Equation (5) by v_{i2} .

We use the same $(v_{i1}, v_{i2})_{i=1}^n$ used in estimation to derive the two perturbed estimates,

$\widehat{\text{PDI}}_l^{(1*)}(\tau; \hat{\theta}_l^*)$ and $\widehat{\text{PDI}}_l^{(2*)}(\tau; \hat{\theta}_l^*)$, and then average them to derive an overall $\widehat{\text{PDI}}_l^*(\tau; \hat{\theta}_l^*)$.

When $\mathcal{B}$ sets of $\{(v_{i1}, v_{i2})_{i=1}^n\}_{b=1}^{\mathcal{B}}$ are used for repeated CV in estimation, we also obtain $\mathcal{B}$ replicates of $\widehat{\text{PDI}}_l^*(\tau; \hat{\theta}_l^*)$ and then take the average. The effect of such repetition on variance estimation is asymptotically ignorable but beneficial in finite samples.

The rationale behind this perturbation scheme is that the conditional distribution of $\sqrt{n}\{\widehat{\text{PDI}}_l^*(\tau; \hat{\theta}_l^*) - \text{PDI}_l(\tau; \hat{\theta}_l)\}$, given the observed data asymptotically follows a zero-mean normal distribution, which well approximates the unconditional distribution of $\sqrt{n}\{\widehat{\text{PDI}}_l(\tau; \hat{\theta}_l) - \text{PDI}_l(\tau; \theta_{l0})\}$. Therefore, one can repeat the perturbation to obtain $\mathcal{J}$ replicates of $\{\widehat{\text{PDI}}_l^{*b}(\tau; \hat{\theta}_l^{*b})\}_{b=1}^{\mathcal{J}}$, and the empirical variance covariance matrix of these perturbed estimates serves as a consistent variance estimator.

3. SIMULATION STUDIES

Simulations were conducted to demonstrate the performance of the proposed estimators for $\text{PDI}(\tau, \theta_0)$ and $\text{PDI}(\tau, \theta_0)$.

3.1. Simulation Settings

Our data generation scheme was similar to that in Fine & Gray(1999). We first generated covariate vectors $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3})$, where Z_{ij} followed the standard normal distribution, $j = 1, 2, 3$. The CIF of Cause 1 failure was

$$P(T_i \leq t, e_i = 1 | \mathbf{Z}_i) = 1 - [1 - \gamma\{1 - \exp(-t)\}]^{\exp(Z_{i1}\beta_{11} + Z_{i2}\beta_{12})},$$

which followed the Fine & Gray model in Equation (1). The probability of ever experiencing a Cause 2 failure was $P(e_i = 2 | \mathbf{Z}_i) = 1 - P(T_i \rightarrow \infty, e_i = 1 | \mathbf{Z}_i)$, and T_i given $\mathbf{Z}_i$ and $e_i = 2$

followed the exponential distribution with rate $\exp(Z_{i1}\beta_{21} + Z_{i2}\beta_{22})$. Note that the Cause 2 failure did not follow the Fine & Gray model. This is because

$$F_2(t; \mathbf{Z}) = P(T \leq t \mid \varepsilon = 2, \mathbf{Z})P(\varepsilon = 2 \mid \mathbf{Z}) \\ = \left[1 - \exp\{-t \cdot \exp(\beta_2^T \mathbf{Z})\} \right] \cdot \left[1 - (1 - \gamma)^{\exp(\beta_1^T \mathbf{Z})} \right],$$

and the corresponding subdistributional hazard $\lambda_2(t; \mathbf{Z}) = -d \log\{1 - F_2(t; \mathbf{Z})\}/dt$ does not satisfy the proportional hazard assumption required by the Fine & Gray model. However, the model can still be adopted as a working prediction tool for building $\mathbf{m}(\tau; \mathbf{Z}, \hat{\boldsymbol{\theta}})$. Independent censoring times were generated from the Uniform[0, U_c] distribution.

The true parameter values for $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})$ were $(0.5, 0.5, -0.5, 0.5)^T$. We considered two scenarios with different choices of γ . We set $\gamma = 0.48$ in Scenario 1, such that Cause 1 and Cause 2 failures were approximately balanced. In Scenario 2, we let $\gamma = 0.3$, and the resulting Cause 1 and Cause 2 failure rates were 33% and 67%, respectively, before independent censoring. We chose U_c to achieve different levels of independent censoring. In Scenario 1, we set $U_{c1} = 7.1$ and $U_{c2} = 3.2$ to get 15% and 30% censoring, respectively. In Scenario 2, we set $U_{c1} = 7.5$ and $U_{c2} = 3.3$ for the same censoring rates. For the purpose of illustration, we set $\tau = 1.4$ in our simulation studies, which renders realistic percentages for different outcome categories (see Table S1 in the Supplementary Material).

For each scenario, we estimated $\mathbf{m}(\tau; \mathbf{Z}, \hat{\boldsymbol{\theta}})$ using two different working prediction models. The first model involved covariates Z_{i1} and Z_{i2} , the two covariates used in the data generation procedure. The second model included Z_{i1} and Z_{i3} as covariates, where Z_{i3} carried no information about the outcome. We refer to these two models as $\mathcal{M}_A$ and $\mathcal{M}_B$. The true values of $\text{PDI}(\tau, \boldsymbol{\theta}_0)$ under these two models were approximated by averaging across 500 empirical versions of definition (2), each based on a dataset of sample size 40,000.

We implemented the proposed methods with sample sizes $n = 400, 600, 800$, and $1,200$, and conducted 2,000 simulations under each scenario. The CV procedure was repeated three times for each simulation. The standard error estimates were obtained using the proposed perturbation procedure with $\mathcal{J} = 400$, and the weights V_i were chosen to be unit exponential. Generally speaking, any positive random variable with mean 1 and variance 1 can be used as weights, in the sense that they all lead to asymptotically consistent variance estimation. The unit exponential distribution satisfies the conditions and becomes a natural choice. In unreported simulation studies, we found that weights following other distributions with unit mean and variance led to very similar performances in variance estimation and coverage rates.

Additionally, we conducted simulations to verify the performance of our methods under covariate-dependent censoring. The data generation scheme for (T_i, ε_i) remained the same as that in Scenario 1. We generated the covariate dependent censoring from the Weibull distribution, with $S_C(t; \mathbf{Z}) = \exp\{-(\lambda t)^{0.25} \exp(\mathbf{Z}_2 \boldsymbol{\nu}_2 + \mathbf{Z}_3 \boldsymbol{\nu}_3)\}$, where $\boldsymbol{\nu}_2 = \boldsymbol{\nu}_3 = 0.2$. We

considered two choices of λ by setting $\lambda = 0.13$ or $\lambda = 1$, which led to 16% and 32% censoring, respectively.

3.2. Simulation Results

Table 1 reports simulation results for the cause-specific PDI, $\text{PDI}_l(\tau, \theta_0)$, where we let $l = 1$ for the purpose of illustration. Table 2 summarizes the results for the overall PDI. Results for Cause 2 and event-free PDIs are provided in Tables S2 and S3 in the Supplementary Material. For both the cause-specific PDI and the overall PDI, the true values under both model $\mathcal{M}_A$ and $\mathcal{M}_B$ were markedly higher than the null value of $1/3$. The true PDIs for model $\mathcal{M}_A$ tended to be higher than their counterparts for model $\mathcal{M}_B$. This suggests that $\mathcal{M}_A$ could achieve better predictive discrimination for the τ -year outcome, by incorporating the information in Z_2 .

From Tables 1 and 2, we observed that estimators for both Cause 1 and overall PDI had negligible bias, and that the biases shrunk with the increase of sample size. The empirical standard deviations (SDs) approximately followed the $n^{-1/2}$ rate. The small SDs suggest that the PDIs were estimated with high accuracy under moderate sample sizes. The performances under 30% censoring were only slightly worse than those under 15%. The average standard errors (SEs) given by the perturbation methods were very close to the empirical SDs, and the coverage probabilities closely matched the nominal level of 95%, where the CIs were obtained using point estimates ± 1.96 SE.

The results under covariate-dependent censoring are provided in Table S4 in the Supplementary Material. The proposed PDI estimators remained unbiased, with SDs that were similar to their counterparts under covariate-independent censoring. The perturbation method still gave coverage rates that were close to the nominal level. Overall, these results suggest that the proposed method can flexibly accommodate the scenario of covariate-dependent censoring.

We also compared the performance between the proposed estimator and the original PDI for uncensored data, which was implemented in R package *mcca* (Li, Gao & D'Agostino, 2019; Gao & Li, 2019), under censoring rates of 0%, 15%, and 30%. The results are summarized in Table S5 in the Supplementary Material. As expected, we observed identical performance between the two methods under 0% censoring. In the presence of censoring, the bias of the original estimator by the R package *mcca* was not substantial in magnitude but was often as large as or larger than the corresponding empirical standard deviation.

4. DATA ANALYSIS

We applied our method to a study of MGUS (Kyle et al., 2002), which contained 1,384 patients diagnosed with MGUS at the Mayo Clinic between 1960 and 1994. The competing risks outcomes of interest involved time until progression to a PCM and time to death without PCM (Therneau, Crowson & Atkinson, 2019). Baseline covariates included patients' age at diagnosis, sex, haemoglobin level (Hgb), creatinine level (Creat), and size of the monoclonal serum spike (MSPIKE). The median and maximum follow-up time were 7

and 35 years, respectively. After excluding 46 (3.3%) patients with missing covariate information, we focused on $n = 1,338$ patients in this analysis. Of these, 112 (8.4%) patients developed PCM, 838 (62.6%) passed away without PCM, and 388 (29.0%) were censored. The event times were recorded in months, and we chose to inspect the 10-year outcome, by setting $\tau = 120$ months. We adopted the Fine & Gray model as our *working* prediction model. The proposed methods were implemented with twofold CV with ten repetitions, and the perturbation procedure was carried out with $\mathcal{J} = 400$. The estimated cause-specific and overall PDIs were summarized in Table 3, and the corresponding regression coefficients were provided in Table S6 in the Supplementary Material.

We first examined the models with only one covariate (models I–V, Table 3). For the PCM endpoint, MSPIKE was the only covariate that led to a cause-specific PDI that was significantly higher than the null value of $1/3$. The estimated $\widehat{\text{PDI}}_1(\tau, \hat{\theta}_I) = 0.514$ (95% CI: 0.440–0.588), suggesting the ability of MSPIKE to discriminate cases of PCM from those with outcomes of death without PCM or free of any events. Interestingly, despite age being significantly associated with the cumulative incidence of PCM, it showed little value in predictive discrimination for this endpoint, with an estimated cause-specific PDI that was worse than $1/3$. This seemingly counter-intuitive result can be explained in Figure S1 (left column) in the Supplemental Material. We observed that the median age in the PCM group was in between those of the event-free group and the death group. Thus, neither higher nor younger age helped identify patients with the PCM outcome, and the discriminative value of age for this outcome category was dubious. Next, the *working* Fine & Gray model may not be optimal for this specific model, as it assumes a linear effect of age that is constant across all time points. Thus, the low PDI may be attributable to both the low discriminative value of age for the PCM outcome and the potential misspecification of this Fine & Gray model.

For death without malignancy, age, Creat, and Hgb showed significant predictive discrimination, where age had the highest $\widehat{\text{PDI}}_2(\tau, \hat{\theta}_I)$ of 0.573 (95% CI: 0.522–0.624).

Generally speaking, all covariates but sex showed some ability to distinguish subjects with different 10-year outcomes, with age at diagnosis possessing the highest discriminative ability, with an overall $\text{PDI}(\tau, \hat{\theta})$ of 0.472.

Next, we examined the performance of multivariable prediction models (models VI–IX, Table 3). When we incorporated additional covariates, we observed slight improvements for the cause-specific PDIs for death without PCM and event-free. For example, the $\widehat{\text{PDI}}_2(\tau, \hat{\theta}_I)$ for death without PCM improved from 0.573 under a univariable model with age alone (model I) to 0.607 under a multivariable model that included all five predictors (model IX). For the PCM endpoint, however, adding additional covariates, such as age, into the model with MSPIKE alone did not lead to any benefit in discriminating PCM cases (Table 3), despite age being significantly associated with the cumulative incidence of PCM in the multivariable model.

Our results agreed well with those using the integrated discrimination improvement (IDI) and improvement in the AUC (IAUC) for comparing nested competing risks regression models (Shi, Cheng & Li, 2014). For the PCM outcome, for example, including age into the

univariable model with MSPIKE led to an IDI of -0.001 (95% CI: -0.012 to 0.017) and an IAUC of -0.008 (95% CI: -0.116 to 0.270), confirming our results that age offered little predictive benefit for the PCM outcome. Adding other predictors into the model with MSPIKE alone also corresponded to IDIs and IAUCs whose 95% CIs covered 0. For death without PCM, the nested comparison between the model with all five predictors vs. the model with age alone led to an IDI of 0.035 (95% CI: 0.021 – 0.058) and an IAUC of 0.166 (95% CI: 0.096 – 0.235), confirming some benefits of the multivariable model vs. the age alone model for the death outcome.

In terms of the overall prediction performance for all three outcome categories, we observed that combining age and MSPIKE (model VI) led to a marked improvement in $\text{PDI}(\tau, \hat{\theta})$ of 0.072 when compared to the univariable models. Adding the remaining three risk factors into the model only led to slight improvements in overall prediction, and the model that included all five factors had an overall PDI of 0.552 (95% CI: 0.509 – 0.595). The estimated cumulative incidence curves under this model are displayed in Figure S2 in the Supplementary Material.

5. DISCUSSION

In this article, we propose an estimator of the PDI to assess the predictive performance of a competing risks regression model in discriminating patients with different τ -year outcomes. The proposed estimator can be consistently estimated following a repeated CV scheme, which makes effective use of the observed data. Moreover, we propose an efficient computation algorithm that can be scaled to handle large datasets and facilitates fast variance estimation through the perturbation-resampling procedure. R scripts for implementing the proposed methods are available at <https://github.com/rli1010/CompetingRisksPDI>. For illustration purposes, we adopt the Fine & Gray model as the working prediction model, without assuming that the regression model is correctly specified. Our methods can easily accommodate prediction tools that stem from other CIF regression models.

In this article, we have focused on the prevalence-independent version of the PDI (Van Calster et al., 2012; Li et al., 2018). Our estimation procedure can be easily extended to estimate the prevalence-dependent version of PDI using the sample prevalence. However, the prevalence-dependent version may become less meaningful in studies where the sample prevalence is not representative. Its interpretation is also not as straightforward as its prevalence-independent counterpart. Therefore, we would recommend the prevalence-independent definition of PDI for general applications.

We have focused on evaluating the prognostic performance of prediction rules that need to be estimated from the same dataset. This is often the case in practical scenarios, where only one dataset is available for the development and validation of a prediction rule. We have adopted repeated twofold CV to avoid optimism in the estimated prediction performance. In situations where the goal is to validate an existing predictor $\mathbf{m}(\tau; \mathbf{Z}, \hat{\theta})$ using an external validation dataset, one should only apply the PDI estimator and the perturbed estimator once by setting $v_{\mathcal{L}} \equiv 1$ for all subjects. In this way, the estimation and inference procedures will treat $\hat{\theta}$ as given.

Besides the PDI, other metrics exist that assess the predictive discrimination of a marker or model for polytomous outcomes. These include the pairwise C-index, the M-index (Hand & Till, 2001), volume under the surface (Mossman, 1999), and hypervolume under the manifold (HUM) (Scurfield, 1996). Extensions of those metrics to the setting of competing risks are of interest, but beyond the scope of the current work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors thank the editor, associate editor, and two anonymous reviewers for their helpful comments and suggestions, which have led to substantial improvements of this article. This work was partially supported by grants from the National Institute of Health (R01DK117209, R01CA193878, R03NS108136) and the Andrew Sabin Family Fellowship.

APPENDIX

Without loss of generality, we prove our theorem with $K = 2$ and consider the Kaplan–Meier method for $\hat{G}(\cdot)$.

Proof of Theorem 1

Let $\mathcal{H}$ be all the functions $f: \mathcal{Z} \rightarrow [0, 1]$ that satisfy condition (C2) and endow $\mathcal{H}$ with the metric $d(f, g) = \max_{\mathbf{z} \in \mathcal{Z}} |f(\mathbf{z}) - g(\mathbf{z})|$. By the Arzelà-Ascoli theorem, $\mathcal{H}$ is compact. Define $\mathcal{H}_0$ as the subset of $\mathcal{H}$ whose elements satisfy condition (C3), then $\mathcal{H}_0$ is a dense subset of $\mathcal{H}$. We also write

$$\hat{D}_l(h) = \frac{1}{n(n-1)(n-2)} \sum_{ijk} v_{i2} v_{j2} v_{k2} \frac{C_l\{h(\mathbf{Z}_i), h(\mathbf{Z}_j), h(\mathbf{Z}_k)\} I_{ijk}}{\hat{G}_V(X_i) \hat{G}_V(X_j) \hat{G}_V(\tau)},$$

and $D_l(h) = \mathbb{E}[C_l\{h(\mathbf{Z}_i), h(\mathbf{Z}_j), h(\mathbf{Z}_k)\} I(T_i \leq \tau, \varepsilon_i = 1, T_j \leq \tau, \varepsilon_j = 2, T_k > \tau)]/8$, where $C_l(x_1, x_2, x_3) = I\{\max(x_1, x_2, x_3) = x_l\} I\{\sum_{i=1}^3 I(x_i = x_l)\}$.

Note that the summand in $\hat{D}_l(h)$ is almost surely continuous in $(\mathcal{H}_0, d)$. By the consistency property of U -statistics, the uniform law of large numbers (Newey, 1991), and the uniform consistency of $\hat{G}_V(\cdot)$ (Kalbfleisch & Prentice, 2011), it can be shown that

$\sup_{h \in \mathcal{H}_0} |\hat{D}_l(h) - D_l(h)| = o_p(1)$. Combining this with the uniform consistency of $m_l(\tau; \cdot, \hat{\theta})$,

we have $\hat{D}_l\{m_l(\tau; \cdot, \hat{\theta})\} \xrightarrow{P} D_l\{m_l(\tau; \cdot, \theta_0)\}$.

Define $\hat{D}_0 = \{n(n-1)(n-2)\}^{-1} \sum_{ijk} v_{i2} v_{j2} v_{k2} I_{ijk} \{\hat{G}_V(X_i) \hat{G}_V(X_j) \hat{G}_V(\tau)\}$, and

$D_0 = \mathbb{E}[I(T_i \leq \tau, \varepsilon_i = 1, T_j \leq \tau, \varepsilon_j = 2, T_k > \tau)]/8$. Using the asymptotic properties of U

statistics and uniform consistency of $\hat{G}_V$, we can show that $\hat{D}_0 \xrightarrow{P} D_0$. By Slutsky's lemma, we have $\widehat{\text{PDI}}_l^{(1)}(\tau; \hat{\theta}) = \hat{D}_l \{m_l(\tau; \cdot, \hat{\theta})\} / \hat{D}_0 \xrightarrow{P} \text{PDI}_l(\tau; \theta_0) = D_l \{m_l(\tau; \cdot, \theta_0)\} / D_0$. Similarly, $\widehat{\text{PDI}}_l^{(2)}(\tau; \hat{\theta}) \xrightarrow{P} \text{PDI}_l(\tau; \theta_0)$. Thus we have $\widehat{\text{PDI}}_l(\tau; \hat{\theta}) = \{\widehat{\text{PDI}}_l^{(1)}(\tau; \hat{\theta}) + \widehat{\text{PDI}}_l^{(2)}(\tau; \hat{\theta})\} / 2 \xrightarrow{P} \text{PDI}_l(\tau; \theta_0)$.

Proof of Theorem 2

We define $I_i = \mathbb{I}(X_i - \tau, -\varepsilon_i = 1)$, $I_j = \mathbb{I}(X_j - \tau, -\varepsilon_j = 2)$, and $I_k = \mathbb{I}(X_k > \tau)$. Condition (C4) implies that $n^{1/2} [D_l \{m_l(\tau; \cdot, \hat{\theta})\} - D_l \{m_l(\tau; \cdot, \theta_0)\}] = n^{-1/2} \sum_{i=1}^n 2v_{i1} \Gamma(\psi_{il}) + o_p(1)$, where Γ is the Hadamard derivative of $D(\cdot)$ at $m(\tau, \cdot, \theta_0)$. Through the application of martingale representation of $\hat{G}_V(\cdot)$ and other approximations, we have

$$n^{1/2}(\hat{D}_0 - D_0) = \frac{n^{1/2}}{n(n-1)(n-2)} \sum_{ijk} \left\{ \frac{v_i 2^v j 2^v k 2^v I_{ijk}}{G(X_i)G(X_j)G(\tau)} - D_0 \right\} + n^{-1/2} \sum_{i=1}^n \int_0^\tau h_1(u) dM_i^c(u) + o_p(1),$$

where $h_1(u) = \mathbb{E} \left[I_{ijk} \left\{ 1 + I(X_i \geq u) + I(X_j \geq u) \right\} / \left\{ 8G(X_i)G(X_j)G(\tau)\pi(u) \right\} \right]$. By the projection theorem of U -statistics, we further obtain

$n^{1/2}(\hat{D}_0 - D_0) = n^{-1/2} \sum_{s=1}^n v_{s2} h_{2s} + n^{-1/2} \sum_{i=1}^n \int_0^\tau h_1(\mu) dM_i^c(\mu) + o_p(1)$, where

$$h_{2s} = \frac{1}{4} \left(\mathbb{E} \left[\frac{I(X_s \leq \tau, \Delta_s \varepsilon_s = 1) I_j I_k}{G(X_s)G(X_j)G(\tau)} \middle| X_s, \varepsilon_s, \Delta_s \right] + \mathbb{E} \left[\frac{I_i I(X_s \leq \tau, \Delta_s \varepsilon_s = 2) I_k}{G(X_i)G(X_s)G(\tau)} \middle| X_s, \varepsilon_s, \Delta_s \right] + \mathbb{E} \left[\frac{I_i I_j I(X_s > \tau)}{G(X_i)G(X_j)G(\tau)} \middle| X_s, \varepsilon_s, \Delta_s \right] - 3D_0 \right).$$

Let $W_l(h) = \sqrt{n} \{ \hat{D}_l(h) - D_l(h) \}$, a similar manipulation gives us

$W_l(h) = n^{-1/2} \sum_{i=1}^n \int_0^\tau h_{3l}(u, h) dM_i^c(u) + n^{-1/2} \sum_{s=1}^n v_{s2} h_{4s}(h) + o_p(1)$, where

$$h_{3l}(u, h) = \mathbb{E} \left[\frac{1 + I(X_i \geq u) + I(X_j \geq u)}{8G(X_i)G(X_j)G(\tau)\pi(u)} I_{ijk} C_l \left\{ h(\mathbf{Z}_i), h(\mathbf{Z}_j), h(\mathbf{Z}_k) \right\} \right],$$

and

$$h_{4s}(h) = \frac{1}{4} \left(\mathbb{E} \left[\frac{I(X_s \leq \tau, \Delta_s \varepsilon_s = 1) I_j I_k C_l \{h(\mathbf{Z}_s), h(\mathbf{Z}_j), h(\mathbf{Z}_k)\}}{G(X_s) G(X_j) G(\tau)} \middle| X_s, \mathbf{Z}_s, \varepsilon_s, \Delta_s \right] + \mathbb{E} \left[\frac{I_l I(X_s \leq \tau, \Delta_s \varepsilon_s = 2) I_k C_l \{h(\mathbf{Z}_i), h(\mathbf{Z}_s), h(\mathbf{Z}_k)\}}{G(X_i) G(X_s) G(\tau)} \middle| X_s, \mathbf{Z}_s, \varepsilon_s, \Delta_s \right] + \mathbb{E} \left[\frac{I_l I_j I(X_s > \tau) C_l \{h(\mathbf{Z}_i), h(\mathbf{Z}_j), h(\mathbf{Z}_s)\}}{G(X_i) G(X_j) G(\tau)} \middle| X_s, \mathbf{Z}_s, \varepsilon_s, \Delta_s \right] - 3D_f(h) \right).$$

It follows from the projection theorem of U -statistics and a functional central limit theorem (Pollard, 1990) that $W_\lambda(h)$ converges weakly to a zero mean Gaussian process in h , and thus $W_\lambda\{m_\lambda(\tau; \cdot, \hat{\theta})\}$ is asymptotically equivalent to $W_\lambda\{m_\lambda(\tau, \cdot, \theta_0)\}$. Simple algebraic manipulation and Slutsky's theorem gives us

$$\begin{aligned} \sqrt{n} \left\{ \widehat{\text{PDI}}_l^{(1)}(\tau; \hat{\theta}) - \text{PDI}_l(\tau; \theta_0) \right\} &= D_0^{-1} n^{-1/2} \sum_{i=1}^n \left\{ 2v_{i1} \Gamma(\psi_{il}) + \int_0^\tau h_{3l}(u, h) dM_i^c(u) + v_{i2} h_{4l}(h) \right\} + D_0^{-1} \\ &\cdot \text{PDI}_l(\tau; \theta_0) \cdot n^{-1/2} \sum_{i=1}^n \left\{ \int_0^\tau h_{1l}(u) dM_i^c(u) + v_{i2} h_{2l}(h) \right\} + o_p(1). \end{aligned}$$

Using the fact that $v_{i1} + v_{i2} = 1$ and $\widehat{\text{PDI}}_l(\tau; \hat{\theta}) = \left\{ \widehat{\text{PDI}}_l^{(1)}(\tau; \hat{\theta}) + \widehat{\text{PDI}}_l^{(2)}(\tau; \hat{\theta}) \right\} / 2$, we have

$$\begin{aligned} \sqrt{n} \left\{ \widehat{\text{PDI}}_l(\tau; \hat{\theta}) - \text{PDI}_l(\tau; \theta_0) \right\} &= D_0^{-1} \cdot \text{PDI}_l(\tau; \theta_0) \cdot n^{-1/2} \sum_{i=1}^n \left\{ \int_0^\tau h_{1l}(u) dM_i^c(u) + \frac{1}{2} h_{2l}(h) \right\} \\ &+ D_0^{-1} n^{-1/2} \sum_{i=1}^n \left\{ \Gamma(\psi_{il}) + \int_0^\tau h_{3l}(u, h) dM_i^c(u) + \frac{1}{2} h_{4l}(h) \right\} + o_p(1). \end{aligned}$$

Thus $\sqrt{n} \left\{ \widehat{\text{PDI}}_l(\tau; \hat{\theta}) - \text{PDI}_l(\tau; \theta_0) \right\}$ is asymptotically normal.

Proof of Lemma 1

Notations and regularity conditions

The score function for β_l can be written as

$$U_l(\beta_l) = \sum_{i=1}^n \int_0^\infty v_{i1} \left[\mathbf{Z}_i - \frac{S_l^{(1)}(\beta_l, u)}{S_l^{(0)}(\beta_l, u)} \right] w_i(u) dN_{il}(u),$$

where $w_i(u) = I(C_i \geq T_i \wedge u) \hat{G}_L(u) / \hat{G}_L(X_i \wedge u)$ and

$$S_l^{(p)}(\beta_l, u) = (n\kappa_n)^{-1} \sum_{i=1}^n v_{i1} w_i(u) Y_{il}(u) \mathbf{Z}_i^{\otimes p} \exp(\mathbf{Z}_i^T \beta_l).$$

Given that $\hat{\beta}_p, \Lambda_{l0}(t)$ can be estimated by the Breslow estimator,

$$\hat{\Lambda}_{l0}(t) = (n\kappa_n)^{-1} \sum_{i=1}^n \int_0^t v_{i1} w_i(u) \{s_l^{(0)}(\hat{\beta}_p, u)\}^{-1} dN_{il}(u), \text{ where } \kappa_n = n^{-1} \sum_{i=1}^n v_{i1}.$$

Denote $\tilde{s}_l^{(p)}(\beta, u) = (n\kappa_n)^{-1} \sum_{i=1}^n v_{i1} Y_{il}(u) I(C_i \geq u) \mathbf{Z}_i^{\otimes p} \exp(\mathbf{Z}_i^T \beta)$, and

$s_l^{(p)}(u) = (n\kappa_n)^{-1} \sum_{i=1}^n v_{i1} Y_{il}(u) \lambda_{il}(u) \mathbf{Z}_i^{\otimes p}$. We assume the following conditions for consistency and asymptotic normality of the misspecified Fine & Gray model.

$$(C5) \text{ There exists } \zeta > 0, \text{ such that } n^{-1/2} \sup_{i,t} \|Z_i\|_2 Y_{il}(t) I(\beta_{l0}^T Z_i > -\zeta \|Z_i\|_2) \xrightarrow{P} 0.$$

(C6) There exists a neighbourhood $\mathcal{B}_l$ of β_{l0} such that for each $t < \infty$ and for $j = 0, 1, 2$,

$$\sup_{u \in [0, t], \beta_l \in \mathcal{B}_l} \|\tilde{s}_l^{(j)}(\beta_l, u) - s_l^{(j)}(\beta_l, u) G(u)\| \xrightarrow{P} 0.$$

(C7) For each $t < \infty$, $s_l^{(j)}(\cdot, t)$, $j = 0, 1, 2$ are continuous functions of $\beta_l \in \mathcal{B}_l$,

uniformly in $u \in [0, t]$, $s_l^{(j)}(\cdot, \cdot)$, $j = 0, 1, 2$ are bounded on $\mathcal{B}_l \times [0, t]$; $s_l^{(0)}(\cdot, \cdot)$ is bounded away from zero on $\mathcal{B}_l \times [0, t]$, and the matrix

$$A_l(\beta) = \int_0^\infty \left[\frac{s_l^{(2)}(\beta, u)}{s_l^{(0)}(\beta, u)} - \left\{ \frac{s_l^{(1)}(\beta, u)}{s_l^{(0)}(\beta, u)} \right\}^{\otimes 2} \right] s_l^{(0)}(u) G(u) du$$

is positive definite on $\mathcal{B}_l$.

Consistency of $\hat{\beta}_l$

For any β_l let $\tilde{l}_l(\beta) = \sum_{i=1}^n \int_0^\infty v_{i1} [\beta_l^T \mathbf{Z}_i - \log \{s_l^{(0)}(\beta, u)\}] I(C_i \geq u) dN_{il}(u)$ be the censoring complete log partial likelihood, which was defined as Section 3.2 of Fine & Gray (1999). Following the proof of Lemma 3.1 in Andersen & Gill (1982), we can show that

$$n^{-1} \tilde{l}_l(\beta_l) \xrightarrow{P} H_l(\beta_l) \text{ for each } \beta_l \in \mathcal{B}_l, \text{ where}$$

$$H_l(\beta_l) = \beta_l^T \int_0^\infty s_l^{(1)}(u) G(u) du - \int_0^\infty \log \{s_l^{(0)}(\beta, u)\} s_l^{(0)}(u) G(u) du. \text{ By condition (C7), } H_l(\beta_l) \text{ is a concave function with a unique maximum at } \beta_l = \beta_{l0}. \text{ Let } \hat{\beta}_l \text{ be the maximizer of } n^{-1} \tilde{l}_l(\beta_l).$$

By Theorem II.1 in Andersen & Gill (1982), we see that $\hat{\beta}_l$ converges in probability to β_{l0} .

Under the presence of right censoring, the log partial likelihood is

$$l_l(\beta_l) = \sum_{i=1}^n \int_0^\infty v_{i1} [\beta_l^T \mathbf{Z}_i - \log \{ \sum_{i=1}^n w_i(u) Y_{il}(u) \mathbf{Z}_i^{\otimes p} \exp(\mathbf{Z}_i^T \beta_l) / n \}] w_i(u) dN_{il}(u). \text{ Notice that } w_i(u) dN_{il}(u) = \tilde{w}_i(u) dN_{il}(u) = I(C_i \geq u) dN_{il}(u), \text{ where } \tilde{w}_i(u) = I(C_i \geq T_i \wedge u) G(u) / G(X_i \wedge u). \text{ We can rewrite the log partial likelihood as}$$

$$l_l(\beta_l) = \sum_{i=1}^n \int_0^\infty v_{i1} [\beta_l^T \mathbf{Z}_i - \log \{s_l^{(0)}(\beta_l, u)\}] I(C_i \geq u) dN_{il}(u). \text{ Under the conditions (C5)–(C7),}$$

$n^{-1}\{\tilde{l}_l(\beta_l) - l_l(\beta_l)\} \xrightarrow{P} 0$ for each $\beta_l \in \mathcal{B}_l$. It follows from Theorem II.1 in Andersen & Gill (1982) that $\hat{\beta}_l \xrightarrow{P} \beta_{l0}$.

Asymptotic normality of $\hat{\beta}_l$

Taking the Taylor expansion of $U(\beta_l)$ around β_{l0} , we have

$n^{1/2}(\hat{\beta}_l - \beta_{l0}) = \hat{A}_l^{-1}(\tilde{\beta}_l)n^{-1/2}U(\beta_{l0})$, where $\tilde{\beta}_l$ lies on the line segment between $\hat{\beta}_l$ and β_{l0} , and

$$\hat{A}_l(\beta_l) = \frac{1}{n} \sum_{i=1}^n \int_0^\infty v_{il} \left[\frac{s_l^{(2)}(\beta_l, u)}{s_l^{(0)}(\beta_l, u)} - \left(\frac{s_l^{(1)}(\beta_l, u)}{s_l^{(0)}(\beta_l, u)} \right)^{\otimes 2} \right] w_l(u) dN_{il}(u).$$

The consistency of $\hat{A}_l(\tilde{\beta}_l)$ for $A_l(\beta_{l0})$ can be established by the same techniques used to prove Theorems 3.2 and 4.2 of Andersen & Gill (1982). A theoretical derivation similar to Lin & Wei (1989) gives $n^{-1/2}U(\beta_{l0}) = n^{-1/2} \sum_{i=1}^n 2v_{il}a_{il}(\beta_{l0}) + o_p(1)$, where $\bar{F}_l(t) = \mathbb{E}N_{il}(t)$, and

$$a_{il}(\beta_l) = \int_0^\infty \left[Z_{il}(t) - \frac{s_l^{(1)}(\beta_l, t)}{s_l^{(0)}(\beta_l, t)} \right] \tilde{w}_{il}(t) dN_{il}(t) - \int_0^\infty \frac{Y_{il}(t) \exp\{\beta_l^T Z_{il}(t)\} \tilde{w}_{il}(t)}{s_l^{(0)}(\beta_l, t)} \left[Z_{il}(t) - \frac{s_l^{(1)}(\beta_l, t)}{s_l^{(0)}(\beta_l, t)} \right] G(t) d\bar{F}_l(t).$$

It follows from Slutsky's lemma and central limit theorem that $n^{1/2}(\hat{\beta}_l - \beta_{l0})$ are asymptotically normal.

Consistency of $\hat{\Lambda}_{l0}(\tau; \hat{\beta}_l)$ and $m_l(\tau; \cdot, \hat{\theta})$

We start with the censoring-complete version of the Breslow estimator,

$\hat{\Lambda}_{l0}(\tau; \hat{\beta}_l) = (n\kappa_n)^{-1} \sum_{i=1}^n v_{il} \int_0^\tau I(C_i \geq u) \left\{ \tilde{s}_l^{(0)}(\hat{\beta}_l, u) \right\}^{-1} dN_{il}(u)$. Using a decomposition of $\hat{\Lambda}_{l0}(\tau; \hat{\beta}_l) - \Lambda_{l0}(\tau)$ similar to Equation (2.8) of Andersen & Gill (1982), we can show that $\hat{\Lambda}_{l0}(\tau; \hat{\beta}_l)$ is a consistent estimator for $\Lambda_{l0}(\tau)$. By Conditions (C5)-(C7) and the consistency of $\hat{\beta}$ and $\hat{\beta}_l$, we have $\tilde{s}_l^{(0)}(\hat{\beta}_l, u) - s_l^{(0)}(\beta_{l0}, u)G(u) = o_p(1)$ and $s_l^{(0)}(\hat{\beta}_l, u) - s_l^{(0)}(\beta_{l0}, u)G(u) = o_p(1)$, uniformly over $[0, \tau]$. Since $s_l^{(0)}(\cdot, \cdot)$ is bounded away from zero on $\mathcal{B}_l \times [0, \tau]$, we have

$$\hat{\Lambda}_{l0}(\tau; \hat{\beta}_l) - \hat{\Lambda}_{l0}(\tau; \hat{\beta}_l) = 2/n \sum_{i=1}^n v_{il} \int_0^\tau \left[\left\{ \tilde{s}_l^{(0)}(\hat{\beta}_l, u) \right\}^{-1} - \left\{ s_l^{(0)}(\hat{\beta}_l, u) \right\}^{-1} \right] I(C_i \geq u) dN_{il}(u) + o_p(1) = o_p(1)$$

Thus we have $|\hat{\Lambda}_{l0}(\tau; \hat{\beta}_l) - \Lambda_{l0}(\tau)| \xrightarrow{P} 0$. The uniform consistency of $m_l(\tau; \cdot, \hat{\theta})$ can be easily obtained from the consistency of $\hat{\theta}_l = \{\hat{\beta}_l^T, \hat{\Lambda}_{l0}(\tau; \hat{\beta}_l)\}^T$, and uniform boundedness of $dm_l(\tau; \mathbf{z}, \hat{\theta}_{lt})/dt$ over $\mathbf{z}$, where $\hat{\theta}_{lt} = (1-t)\theta_{l0} + t\hat{\theta}_l, t \in [0, 1]$.

Asymptotic normality of $\hat{\Lambda}_{l0}(\tau; \hat{\beta}_l)$ and $\hat{\theta}_l$

By taking a Taylor series of $\hat{\Lambda}_{l0}(\tau; \hat{\beta}_l)$ around β_{l0} and using algebraic manipulations, we obtain

$$\begin{aligned} & \frac{\sqrt{n}}{2} \{ \hat{\Lambda}_{l0}(\tau; \hat{\beta}_l) - \Lambda_{l0}(\tau) \} \\ &= Q_l(\beta_l^*, \tau)^T n^{1/2} (\hat{\beta}_l - \beta_{l0}) + n^{-1/2} \sum_{i=1}^n v_{i1} \int_0^\tau \frac{\tilde{w}_l(u)}{s_l^{(0)}(\beta_{l0}, u)} dM_{il}(u) \\ &+ n^{-1/2} \sum_{i=1}^n v_{i1} \int_0^\tau \left[\frac{\tilde{w}_l(u) \lambda_{il}(u) Y_{il}(u)}{G(u) s_l^{(0)}(\beta_{l0}, u)} - \frac{s_l^{(0)}(u)}{s_l^{(0)}(\beta_{l0}, u)} \right] du \\ &+ n^{-1/2} \sum_{i=1}^n v_{i1} \int_0^\tau \left[\frac{1}{s_l^{(0)}(\beta_{l0}, u)} - \frac{1}{G(u) s_l^{(0)}(\beta_{l0}, u)} \right] \tilde{w}_l(u) dN_{il}(u) + o_p(1), \end{aligned}$$

where $dM_{il}(u) = dN_{il}(u) - Y_{il}(u) \lambda_{il}(u) du$, β_l^* is on the line segment between $\hat{\beta}_l$ and β_{l0} , and

$$Q_l(\beta, t) = \frac{1}{n} \sum_{i=1}^n v_{i1} \int_0^t \frac{-s_l^{(1)}(\beta, u)}{\{s_l^{(0)}(\beta, u)\}^2} \tilde{w}_l(u) dN_{il}(u).$$

Denote $M_i^c(u) = I(X_i \leq u, \Delta_i = 0) - \int_0^u I(X_i \geq t) d\Lambda^c(t)$ as the martingale associated with the censoring process, where $\Lambda^c(t)$ is the cumulative hazard for the censoring variable. Write $q_l(\beta, t) = \int_0^t \left\{ -s_l^{(1)}(\beta, u)/s_l^{(0)}(\beta, u) \right\} d\Lambda_{l0}(u)$, $\pi(u) = \mathbb{E}\{I(X_i \geq u)\}$ and

$$v_l(u) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{Y_{il}(s) \exp(\mathbf{Z}_i^T \beta_{l0}) I(C_i \geq T_i \wedge s) I(X_i \leq u \leq s)}{\{s_l^{(0)}(\beta_{l0}, s)\}^2 G(X_i)} d\bar{F}_l(s).$$

Notice that $\mathbb{E}\{\tilde{w}_l(u) dM_{il}(u); \mathbf{Z}_i\} = G(u) \mathbb{E}\{dM_{il}(u); \mathbf{Z}_i\}$, and $2Q_l(\beta_l^*, \tau) \xrightarrow{P} q_l(\beta_{l0}, \tau)$. We get $n^{1/2} \{ \hat{\Lambda}_{l0}(\tau; \hat{\beta}_l) - \Lambda_{l0}(\tau) \} = n^{-1/2} \sum_{i=1}^n 2v_{i1} \eta_{il} + o_p(1)$, where

$$\begin{aligned} \eta_{il} &= \int_0^\tau \frac{\tilde{w}_l(u)}{s_l^{(0)}(\beta_{l0}, u)} dM_{il}(u) + \int_0^\tau \left[\frac{\tilde{w}_l(u) \lambda_{il}(u) Y_{il}(u)}{G(u) s_l^{(0)}(\beta_{l0}, u)} - \frac{s_l^{(0)}(u)}{s_l^{(0)}(\beta_{l0}, u)} \right] du + \int_0^\tau \frac{v_l(u)}{\pi(u)} dM_i^c(u) \\ &+ q_l(\beta_{l0}, \tau)^T A_l^{-1}(\beta_{l0}) a_{il}(\beta_{l0}) + \int_0^\tau \frac{1}{s_l^{(0)}(\beta_{l0}, u)} \left[1 - \frac{\tilde{w}_l(u) Y_{il}(u) \exp(\mathbf{Z}_i^T \beta_{l0})}{G(u) s_l^{(0)}(\beta_{l0}, u)} \right] d\bar{F}_l(u). \end{aligned}$$

It follows that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_I - \boldsymbol{\theta}_{I0}) = n^{-1/2} \sum_{i=1}^n 2v_{i1} \boldsymbol{\vartheta}_{iI} + o_p(1),$$

where $\boldsymbol{\vartheta}_{iI} = \{a_{iI}(\boldsymbol{\beta}_{I0})^T A_I^{-1}(\boldsymbol{\beta}_{I0}) \cdot \eta_{iI}\}^T$.

BIBLIOGRAPHY

- Andersen PK & Gill RD (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics*, 10, 1100–1120.
- Andersen PK, Klein JP, & Rosthøj S (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, 90, 15–27.
- Bellach A, Kosorok MR, Rüschemdorf L, & Fine JP (2019). Weighted npmlr for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 114, 259–270. [PubMed: 31073256]
- Blanche P, Dartigues J-F, & Jacqmin-Gadda H (2013). Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine*, 32, 5381–5397. [PubMed: 24027076]
- de Jong VM, Eijkemans MJ, van Calster B, Timmerman D, Moons KG, Steyerberg EW, & van Smeden M (2019). Sample size considerations and predictive performance of multinomial logistic prediction models. *Statistics in Medicine*, 38, 1601–1619. [PubMed: 30614028]
- Fine JP & Gray RJ (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94, 496–509.
- Gao M & Li J (2019). mcca: Multi-Category Classification Accuracy.
- Hand DJ & Till RJ (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*, 45, 171–186.
- Harrell FE, Califf RM, Pryor DB, Lee KL, & Rosati RA (1982). Evaluating the yield of medical tests. *Journal of the American Medical Association*, 247, 2543–2546. [PubMed: 7069920]
- Jeong J-H & Fine JP (2006). Parametric regression on cumulative incidence function. *Biostatistics*, 8, 184–196. [PubMed: 16636138]
- Kalbfleisch JD & Prentice RL (2011). *The Statistical Analysis of Failure Time Data*, 2nd ed., John Wiley & Sons.
- Klein JP & Andersen PK (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, 61, 223–229. [PubMed: 15737097]
- Kyle RA, Therneau TM, Rajkumar SV, Offord JR, Larson DR, Plevak MF, & Melton LJ (2002). A long-term study of prognosis in monoclonal gammopathy of undetermined significance. *New England Journal of Medicine*, 346, 564–569.
- Li J, Feng Q, Fine JP, Pencina MJ, & Van Calster B (2018). Nonparametric estimation and inference for polytomous discrimination index. *Statistical Methods in Medical Research*, 27, 3092–3103. [PubMed: 28178877]
- Li J, Gao M, & D'Agostino R (2019). Evaluating classification accuracy for modern learning approaches. *Statistics in Medicine*, 38, 2477–2503. [PubMed: 30701585]
- Lin DY & Wei LJ (1989). The robust inference for the cox proportional hazards model. *Journal of the American Statistical Association*, 84, 1074–1078.
- Mossman D (1999). Three-way ROCs. *Medical Decision Making*, 19, 78–89. [PubMed: 9917023]
- Newey WK (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica*, 59, 1161–1167.
- Pollard D (1990). *Empirical processes: Theory and applications*. NSF—CBMS Regional Conference Series in Probability and Statistics, Vol. 2, IMS.

- Robins JM, Rotnitzky A, & Zhao LP (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90, 106–121.
- Saha P & Heagerty P (2010). Time-dependent predictive accuracy in the presence of competing risks. *Biometrics*, 66, 999–1011. [PubMed: 20070296]
- Scheike TH, Zhang MJ, & Gerds TA (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, 95, 205–220.
- Schoop R, Beyersmann J, Schumacher M, & Binder H (2011). Quantifying the predictive accuracy of time-to-event models in the presence of competing risks. *Biometrical Journal*, 53, 88–112. [PubMed: 21259311]
- Scurfield BK (1996). Multiple-event forced-choice tasks in the theory of signal detectability. *Journal of Mathematical Psychology*, 40, 253–269. [PubMed: 8979976]
- Shi H, Cheng Y, & Li J (2014). Assessing diagnostic accuracy improvement for survival or competing-risk censored outcomes. *Canadian Journal of Statistics*, 42, 109–125.
- Therneau T, Crowson C, & Atkinson E (2019). Multi-state models and competing risks, <https://cran.rproject.org/web/packages/survival/vignettes/compete.pdf>.
- Uno H, Cai T, Pencina MJ, D'Agostino RB, & Wei L (2011). On the c -statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*, 30, 1105–1117. [PubMed: 21484848]
- Uno H, Cai T, Tian L, & Wei L (2007). Evaluating prediction rules for t -year survivors with censored regression models. *Journal of the American Statistical Association*, 102, 527–537.
- Van Calster B, Van Belle V, Vergouwe Y, Timmerman D, Van Huffel S, & Steyerberg EW (2012). Extending the c -statistic to nominal polytomous outcomes: The polytomous discrimination index. *Statistics in Medicine*, 31, 2610–2626. [PubMed: 22733650]
- Wolbers M, Blanche P, Koller MT, Witteman JC, & Gerds TA (2014). Concordance for prognostic models with competing risks. *Biostatistics*, 15, 526–539. [PubMed: 24493091]
- Wu C & Li L (2018). Quantifying and estimating the predictive accuracy for censored time-to-event data with competing risks. *Statistics in Medicine*, 37, 3106–3124. [PubMed: 29766537]
- Zheng Y, Cai T, Jin Y, & Feng Z (2012). Evaluating prognostic accuracy of biomarkers under competing risk. *Biometrics*, 68, 388–396. [PubMed: 22150576]

Table 1:

Simulation results for the cause-specific PDI¹

Scenario	Censoring	n	Model $\mathcal{M}_A$					Model $\mathcal{M}_B$						
			Truth	EST	BIAS	ESD	ASE	CP	Truth	EST	BIAS	ESD	ASE	CP
1	15%	400	0.584	0.580	-0.004	0.037	0.037	0.956	0.478	0.475	-0.004	0.038	0.039	0.953
		600	0.584	0.582	-0.002	0.030	0.030	0.951	0.478	0.476	-0.002	0.031	0.031	0.955
		800	0.584	0.581	-0.002	0.026	0.026	0.950	0.478	0.475	-0.003	0.027	0.027	0.950
		1,200	0.584	0.582	-0.002	0.021	0.021	0.949	0.478	0.476	-0.002	0.021	0.022	0.950
	30%	400	0.584	0.579	-0.005	0.041	0.041	0.954	0.478	0.473	-0.005	0.042	0.044	0.955
		600	0.584	0.581	-0.002	0.033	0.033	0.943	0.478	0.475	-0.003	0.035	0.035	0.947
		800	0.584	0.580	-0.003	0.028	0.029	0.950	0.478	0.475	-0.003	0.029	0.030	0.954
		1,200	0.584	0.582	-0.002	0.023	0.023	0.946	0.478	0.476	-0.002	0.024	0.024	0.957
	15%	400	0.564	0.560	-0.004	0.041	0.042	0.959	0.455	0.450	-0.005	0.043	0.044	0.948
		600	0.564	0.562	-0.002	0.033	0.034	0.951	0.455	0.451	-0.003	0.033	0.035	0.956
		800	0.564	0.561	-0.003	0.029	0.029	0.948	0.455	0.451	-0.003	0.029	0.030	0.954
		1,200	0.564	0.563	-0.001	0.023	0.024	0.952	0.455	0.452	-0.003	0.024	0.024	0.951
30%	400	0.564	0.557	-0.008	0.045	0.047	0.962	0.455	0.446	-0.008	0.048	0.049	0.948	
	600	0.564	0.562	-0.003	0.036	0.037	0.955	0.455	0.452	-0.003	0.038	0.039	0.948	
	800	0.564	0.560	-0.004	0.032	0.032	0.952	0.455	0.451	-0.004	0.032	0.033	0.959	
	1,200	0.564	0.561	-0.003	0.025	0.026	0.952	0.455	0.452	-0.002	0.027	0.027	0.945	

ASE, average of estimated standard errors over replications; CP, coverage probability of 95% confidence interval; ESD, empirical standard deviation of the estimates over replications; EST, empirical average of estimated PDI over replications.

Table 2:

Simulation results for the overall PDI

Scenario	Censoring	n	Model $\mathcal{M}_A$						Model $\mathcal{M}_B$					
			Truth	EST	BIAS	ESD	ASE	CP	Truth	EST	BIAS	ESD	ASE	CP
1	15%	400	0.551	0.546	-0.005	0.028	0.028	0.952	0.481	0.474	-0.006	0.025	0.026	0.956
		600	0.551	0.548	-0.003	0.022	0.023	0.959	0.481	0.477	-0.004	0.020	0.021	0.957
		800	0.551	0.548	-0.003	0.019	0.020	0.950	0.481	0.477	-0.004	0.017	0.018	0.950
	30%	1,200	0.551	0.549	-0.002	0.016	0.016	0.949	0.481	0.478	-0.003	0.014	0.015	0.947
		400	0.557	0.551	-0.007	0.031	0.032	0.947	0.482	0.474	-0.008	0.027	0.029	0.962
2		600	0.557	0.553	-0.004	0.025	0.026	0.952	0.482	0.476	-0.006	0.022	0.023	0.956
		800	0.557	0.554	-0.003	0.022	0.022	0.948	0.482	0.478	-0.004	0.019	0.020	0.960
		1,200	0.557	0.555	-0.002	0.018	0.018	0.951	0.482	0.479	-0.003	0.016	0.016	0.957
	15%	400	0.555	0.550	-0.006	0.028	0.029	0.951	0.475	0.469	-0.006	0.023	0.025	0.967
		600	0.555	0.552	-0.004	0.023	0.023	0.946	0.475	0.470	-0.005	0.019	0.020	0.962
	30%	800	0.555	0.552	-0.003	0.019	0.020	0.950	0.475	0.471	-0.004	0.015	0.017	0.961
		1,200	0.555	0.554	-0.001	0.016	0.016	0.945	0.475	0.473	-0.002	0.012	0.013	0.961
		400	0.560	0.553	-0.006	0.031	0.032	0.951	0.475	0.468	-0.007	0.025	0.028	0.970
		600	0.560	0.556	-0.003	0.025	0.026	0.953	0.475	0.471	-0.004	0.020	0.022	0.968
		800	0.560	0.556	-0.003	0.022	0.022	0.944	0.475	0.471	-0.003	0.017	0.019	0.963
		1,200	0.560	0.557	-0.003	0.017	0.018	0.953	0.475	0.472	-0.002	0.014	0.015	0.959

ASE, average of estimated standard errors over replications; CP, coverage probability of 95% confidence interval; ESD, empirical standard deviation of the estimates over replications; EST, empirical average of estimated PDI over replications.

Table 3:

Estimated cause-specific and overall PDIIs for MGUS data

Model	Covariates	PCM		Death w/o PCM		Event free		Overall	
		PDI	CI	PDI	CI	PDI	CI	PDI	CI
I	Age	0.271	(0.212, 0.330)	0.573	(0.522, 0.624)	0.570	(0.517, 0.623)	0.472	(0.448, 0.496)
II	Creat	0.367	(0.287, 0.447)	0.476	(0.433, 0.519)	0.363	(0.287, 0.439)	0.402	(0.365, 0.439)
III	Hgb	0.310	(0.200, 0.420)	0.431	(0.382, 0.480)	0.460	(0.411, 0.509)	0.400	(0.361, 0.439)
IV	MSPIKE	0.514	(0.440, 0.588)	0.412	(0.314, 0.510)	0.332	(0.240, 0.424)	0.419	(0.360, 0.478)
V	Sex	0.363	(0.269, 0.457)	0.380	(0.327, 0.433)	0.341	(0.300, 0.382)	0.361	(0.320, 0.402)
VI	Age + MSPIKE	0.475	(0.389, 0.561)	0.586	(0.535, 0.637)	0.572	(0.519, 0.625)	0.544	(0.505, 0.583)
VII	Age + Creat + MSPIKE	0.475	(0.385, 0.565)	0.601	(0.550, 0.652)	0.573	(0.518, 0.628)	0.550	(0.509, 0.591)
VIII	Age + Creat + Hgb + MSPIKE	0.475	(0.379, 0.571)	0.598	(0.545, 0.651)	0.581	(0.526, 0.636)	0.551	(0.510, 0.592)
IX	Age + Creat + Hgb + MSPIKE + Sex	0.470	(0.372, 0.568)	0.607	(0.552, 0.662)	0.580	(0.523, 0.637)	0.552	(0.509, 0.595)

Age, age at diagnosis; CI, estimated 95% confidence interval for PDI; Creat, creatinine level; Hgb, haemoglobin level; MSPIKE, size of the monoclonal serum spike.