Inference about Time-dependent Prognostic Accuracy Measures in the Presence of Competing Risks

R. Dey¹, G. Sebastiani² and P. Saha-Chaudhuri¹†

Abstract

Background: Evaluating a candidate marker or developing a model for predicting risk of future conditions is one of the major goals in medicine. However, model development and assessment for a time-to-event outcome may be complicated in the presence of competing risks. In this manuscript, we propose a local and a global estimators of cause-specific AUC for right-censored survival times in the presence of competing risks.

Methods: The local estimator - cause-specific weighted mean rank (cWMR) - is a local average of time-specific observed cause-specific AUCs within a neighborhood of given time t. The global estimator - cause-specific fractional polynomials (cFPL) - is based on modelling the cause-specific AUC as a function of t through fractional polynomials.

Results: We investigated the performance of the proposed cWMR and cFPL estimators through simulation studies and real-life data analysis. The estimators perform well in small samples, have minimal bias and appropriate coverage.

Conclusions: The local estimator cWMR and the global estimator cFPL will provide computationally efficient options for assessing the prognostic accuracy of markers for time-to-event outcome in the presence of competing risks in many practical settings.

Keywords: Competing Risks; Area under the ROC curve (AUC); Cause-specific AUC; Fractional Polynomials

Background

In modern evidence-based medicine, decisions on appropriate early medical intervention or the choice and timing of interventions frequently rely on prognostic markers that are informative for survival outcome. Such marker can be a single covariate or a risk score where the latter is estimated from a survival model of the association between covariates and survival time. For instance, a well-known Framingham risk score predicts the 10-year risk of cardiovascular disease [1]. This model often guides clinicians to identify those at high/low-risk for future cardiovascular disease. Similar risk models and markers exist for other diseases, e.g. Gail score for breast cancer [2], prostate specific antigen for prostate cancer, kidney donor risk index (KDRI) of donor for kidney transplant etc. In liver transplantation study (LT), the side-effects of medications put the organ recipients at risk for recurrent liver disease (liver fibrosis) which can lead to graft failure [3]. Liver biopsy is the gold standard for staging, managing and treating liver fibrosis, and useful in the prognosis of LT recipients. However, biopsy is often impractical due to its invasiveness and higher
cost. Serum fibrosis markers such as the fibrosis score 4 (FIB-4), and the nonalcoholic fatty liver disease fibrosis score (NAFLD) measured one year after LT have recently been shown to accurately stratify organ recipients who are at high risk from those who are at low risk of recurrent liver disease [3], [4]. Therefore, rather than performing biopsy on all patients, these markers could guide clinicians to start timely treatment for those high-risk patients while sparing low-risk patients from the side effects of biopsy and from the unnecessary costs. Before one adopts such markers into clinical practice, it is crucial to evaluate their prognostic accuracy i.e. whether markers correctly discriminate subjects who will subsequently experience the event of interest at or by time \( t \) from those who will not experience any event by \( t \). The goal of this paper is to develop a method to estimate time-dependent prognostic accuracy measure of such marker.

The evaluation of prognostic accuracy of marker becomes complicated in the presence of competing risks. Competing risks arise when a subject experiences a terminal event due to one of the multiple mutually exclusive causes. For example, post-liver transplant, patients may die due to adverse liver and/or transplant-related outcomes (e.g. graft failure) while other patients may die due to competing events (e.g. non-liver causes) before experiencing liver-related adverse outcomes. This leads to a competing risks situation because liver-related marker could predict graft failure or graft related death, but may not predict non-graft related events. [5]. Here, the scientific question is how well do FIB-4 or NAFLD discriminate between patients who progress to graft-related death and those who do not. Understanding whether FIB-4 or NAFLD is highly predictive for death due to graft failure but not others could potentially lead to more rational and cost-effective use of specific medications or treatment strategies. In order to facilitate the assessment of prognostic accuracy of marker, the goal of this manuscript is to develop methods to estimate time-dependent prognostic accuracy of a baseline marker after taking right-censoring and competing risks into account.

A number of statistical measures have been proposed to assess the prognostic accuracy of a marker. The Receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) [6] are the most popular measures of a binary classifier system. ROC curve is a graphical illustration of the diagnostic ability of a binary classifier system as its discrimination threshold is varied and AUC provides a global summary of the discriminatory capacity of the marker. For a survival outcome, the event status of a subject can change over time and the risk of developing the event conditional on marker value changes over the follow-up time. Therefore, the accuracy summaries for evaluating the performance of a marker must take this time-dependence into account. Time-dependent ROC methods for survival time [7],[8] classify the subjects as cases or controls depending on their survival status at or by time \( t \) and compare their observed status with a predicted risk at some or all times. The incident cases/dynamic controls classification arises [8] when scientific interest focuses on correct classification of subjects who are still in risk at time \( t \). The incident (I) cases are those subjects who had an event at \( t \) and the dynamic (D) controls are those who survived through \( t \). Among the other definitions of time-dependent ROC, cumulative (C) case and dynamic control definition pair is most common. In this classification, cumulative cases are defined as patients having
Dey et al.

Page 3 of 18

an event within a certain time range, say \([0, t]\). The I/D version of AUC focuses on incident cases and is better suited for characterizing the trajectory of AUC over time, while C/D AUC does not characterize the evolution of accuracy over time. Another appealing property of I/D measure is that the evaluation of the model at a certain time point \(t\) only focuses on the riskset at the time \(t\) and therefore prior events and performance of the marker does not influence or distort the marker accuracy at \(t\). As prognostic models aim to predict future, this property is very appealing when evaluating dynamic prognostic models.

In this article, we focus on I/D definition of AUC(\(t\)), which is the time-dependent area under the I/D ROC curve at \(t\) and was introduced in [8]. There has been extensive research in estimating the I/D AUC(\(t\)) in the case of single cause of failure. In [9], a locally weighted mean rank (WMR) smoothing based on the intermediate concordance measure was proposed. A fractional polynomials (FPL) estimator based on modeling AUC(\(t\)) as a function of time was proposed in [10]. To evaluate the methods, the authors in [11] compared different approaches (i.e. [8], [12], [6]) of estimating the concordance index based on the AUC(\(t\)) by simulation studies. However, the estimation of the AUC(\(t\)) at different follow-up time points was not investigated in that study.

Though competing risks is an important issue in many practical clinical settings, there is limited research on estimating the I/D AUC(\(t\)) in the presence of competing risks. When there is only a single cause of failure, the I/D definition stratifies only the subjects in the riskset. In the presence of competing risks, I/D definition further stratifies the subjects who are still at risk at time \(t\) into a single control group and cause-specific case groups depending on the cause of failure. Estimation of time-dependent measures under competing risks were discussed in [13] and [14]. In [13] a time-dependent I/D ROC curve was estimated using a Cox model for the cause-specific hazards and riskset reweighting of the marker distribution. This approach is semi-parametric, indirect, and computationally intensive. First, it requires correct specification of a conditional hazard regression model linking the marker to the event time. This approach provides biased estimate when the monotonicity of association between marker and event time is violated. Second, in order to obtain the AUC curve from the ROC, numerical integration of the ROC curve is required. Furthermore, within an interval around each unique event time, a Cox model is assumed and the parameters of the sequence of Cox models are estimated for each neighborhood around the unique event times. Therefore, their approach is also time-consuming. The goal of this manuscript is to propose non-parametric, direct, intuitive and scalable methods to estimate the I/D cause-specific AUC(\(t\)) of baseline marker, accounting for censoring and competing events. The major advantage of our proposed method over the existing semi-parametric estimator will be that it requires no specification of a conditional hazard model linking the marker to the event time and hence robust to model misspecification. In addition, the inference of this measure can be developed under minimal assumptions.

The rest of the article is organized as follows. In subsection 2.1, we provide the notation. In subsections 2.3 and 2.4, we introduce a local and a global estimators of cause-specific I/D AUC(\(t\)) that are direct, flexible and non-parametric. We report simulation results to illustrate our methodology in section 3.1. We report a real data
example in section 3.2, and put a discussion in section 4. We end with conclusion in section 5.

Methods

Notation

Let \( M \) denote the (baseline) marker that could potentially be used in predicting the survival time in the presence of competing risks where a subject can fail due to \( J \) mutually exclusive causes. Note that \( M \) can be a single covariate or it could be a risk score that may be calculated from a survival model (e.g. proportional hazard model). Let the implicit event times for each of the \( J \) causes be \( \{ T^{(1)}, \ldots, T^{(J)} \} \).

In the presence of competing risks, only the time to the occurrence of the first event is potentially observable. Thus, without censoring, one observes \( T = \min(T^{(1)}, \ldots, T^{(J)}) \) and the failure indicator \( \delta \), which takes value \( j \) if \( T = T^{(j)} \) (\( j = 1, \ldots, J \)). Define the observed event time, \( Z = \min(T, C) \) where \( C \) is the censoring time. A censored observation has \( Z = C \) and this is recorded by \( \delta = 0 \). Suppose that there are \( n \) subjects in the study. Let \( 1(\cdot) \) denote the indicator function. Let \( R_i(t) = 1\{Z_i \geq t\} \) denote the at-risk indicator for the \( i \)-th individual at time \( t \) (\( i = 1, 2, \ldots, n \)). Let \( R_t = \{ i : R_i(t) = 1 \} \) denote the subjects that are in the riskset at time \( t \). Among the subjects in \( R_t \), the subjects who had an event from cause \( j \) at \( t \) are the \( j \)-th cause-specific incident (I) cases: \( R_{1t}^{(j)} = \{ i : T_i = t, \delta_i = j \} \). The subjects who did not have an event by \( t \) are the dynamic (D) controls: \( R_{0t} = \{ i : T_i > t \} \).

Let \( n_t \) be the size of \( R_{0t} \) i.e. \( n_t = |R_{0t}| \) and \( d_t^{(j)} \) be the size of \( R_{1t}^{(j)} \), \( d_t^{(j)} = |R_{1t}^{(j)}| \).

If there is a single cause of failure (\( J = 1 \)), the time-dependent incident/dynamic area under the ROC curve at time \( t \), I/D AUC\((t)\), is defined as

\[
\Pr(M_i > M_k | T_i = t, T_k > t).
\]

This is the time-dependent probability at time \( t \) that for a randomly selected case-control pair \((i, k)\) the marker value for the incident case is higher than the marker value for the control.

Time-dependent I/D Cause-Specific AUC\((t)\)

In the presence of competing risks, the I case and D control definition can be extended as cause-specific incident cases and dynamic controls as follows:

- \( j \)-th cause-specific case: \( T = t, \delta = j; \ j = 1, 2, \ldots, J \).
- Control: \( T > t \).

The I/D AUC\((t)\) in (1) can be redefined as the \( j \)-th I/D cause-specific AUC\((t)\) and define as

\[
\text{AUC}^{(j)}_t = \Pr(M_i > M_k | T_i = t, \delta_i = j, T_k > t),
\]

where \( j = 1, 2, \ldots, J \). Below we propose a local and a global estimators for the time-dependent cause-specific AUC\((t)\) curve.
Weighted Mean Rank Estimator of I/D Cause-Specific AUC\(_{(t)}\) (cWMR)

A non-parametric estimator of I/D AUC\(_{(t)}\) was proposed in [9] using a nearest neighbor method. This estimator is a local average of time-specific observed AUCs. While the method do not address the issue of more than one cause of failure, natural modification of the approach allows estimation of accuracy in the presence of multiple causes of failure. To illustrate our proposed method, let A\(_{(j)}(t)\) denote the proportion of \((i, k)\) pairs where subject \(k\) has a lower marker value compared with that of subject \(i\) who experienced failure due to cause \(j\), provided subject \(k\) has longer survival than subject \(i\) and define as

\[
A^{(j)}(t) = \frac{1}{d^{(j)}_t \times n_t} \sum_{i \in R^{(j)}_t} \sum_{k \in R^{(j)}_t} \{ M_i > M_k \mid T_i = t, \delta_i = j, T_k > t \}.
\]

Note that, A\(_{(j)}(t)\) can be considered as an estimator of AUC\(_{(j)}\) in (2). However, typically failure time is measured in continuous scale and it is reasonable to assume that at a given time the likelihood of failures due to multiple causes is negligible. Therefore, there are only few cases experiencing the event of interest (e.g. \(j\)-th cause) at \(t\) and often \(d^{(j)}_t = 1\). When \(d^{(j)}_t = 1\), A\(_{(j)}(t)\) becomes

\[
A^{(j)}(t) = \frac{1}{n_t} \sum_{k \in R^{(j)}_t} \{ M_i > M_k \mid T_i = t, \delta_i = j, T_k > t \}
\]

the rank of the \(j\)-th cause-specific case marker value among the control markers at time \(t\). In extreme situations, this quantity could jump from 1 to 0 and back between adjacent time points. Hence, the estimation of AUC\(_{(j)}\) based on A\(_{(j)}(t)\) requires some degree of smoothing. In this situation, the information within a neighborhood around \(t\), \(N_t^{(j)}(h_n) = \{ t_k : |t - t_k| < h_n, \delta = j \}\) can be used to estimate marker concordance at \(t\). Let cWMR\(_{t}^{(j)}\) be an \(j\)-th cause-specific weighted mean rank estimator of AUC\(_{(j)}\) and define as

\[
cWMR_{t}^{(j)} = \frac{1}{|N_t^{(j)}(h_n)|} \sum_{t_k \in N_t^{(j)}(h_n)} A^{(j)}(t_k), \tag{3}
\]

where \(|N_t^{(j)}(h_n)|\) is the size of \(N_t^{(j)}(h_n)\). The optimal bandwidth \((h_n)\) balances bias and variance of cWMR\(_{t}^{(j)}\). Therefore, to select the bandwidth, we followed the leave-one-out cross validation approach of [9] and adapted it to account for competing risks. In addition, we also propose a variance estimator for cWMR\(_{t}^{(j)}\) based on the assumption of bivariate Normality of the case and control marker pairs [9]. An additional file shows this in more detail [see Additional file 1].

**Asymptotic Properties of cWMR\(_{t}^{(j)}\):**

In order to evaluate the asymptotic behaviours of cWMR\(_{t}^{(j)}\), we restrict our attention to a neighborhood around \(t\) and cause \(j\) i.e. \(N_t^{(j)}(h_n) = \{ t_k : |t - t_k| < h_n, \delta = j \}\) with size \(|N_t^{(j)}(h_n)| = m_t^{(j)}\). Other causes can be treated in a similar way. Let \(B_t\) denote the number of subjects at the start of the neighborhood i.e. at time \(t - h_n\).

**Theorem 1:** Let \(n - (B_t + m_t^{(j)})\) denote the number of subjects surviving past \(t + h_n\). If \(n - (B_t + m_t^{(j)}) \to \infty\) as \(n \to \infty\) and \(h_n \to 0\) then cWMR\(_{t}^{(j)}\) converges to a Normal distribution with mean AUC\(_{(j)} + b_n(t)\) and asymptotic variance
\[ V_n^{(j)} = \text{Var}[cWMR_t^{(j)}], \text{ where, } b_n(t) \text{ denotes the bias and the variance is} \]

\[
V_n^{(j)} = \frac{1}{[n_t^{(j)}]^2} \left\{ \sum_{t(i) \in N_t^{(j)}(h_n) \atop t(i) \neq t(k)} \text{var}[A^{(j)}(t(i))] + \sum_{t(i) \neq t(k)} \text{cov}[A^{(j)}(t(i)), A^{(j)}(t(k))] \right\}.
\]

The proof of this theorem is provided in additional file 1.

**Estimation of variance of cWMR**

The variance of cWMR, \( V_n^{(j)} \), has components \( \text{var}[A^{(j)}(t(i))] \) and \( \text{cov}[A^{(j)}(t(i)), A^{(j)}(t(k))] \).

In order to compute these components, we propose the following estimators in the spirit of variance calculation of AUC proposed in [15] and [9]

\[
\text{var}[A^{(j)}(t)] = \left( \frac{n_t - 1}{n_t} \right) \left\{ P_1^{(j)}(t) - [P_0^{(j)}(t)]^2 \right\} + \frac{1}{n_t} \left\{ P_0^{(j)}(t)(1 - P_0^{(j)}(t)) \right\}
\]

\[
\text{cov}[A^{(j)}(t), A^{(j)}(s)] = \frac{1}{n_t} \left\{ \{ P_2^{(j)}(t, s) - P_2^{(j)}(t_0, s) \} + \{ P_3^{(j)}(t, s) - P_3^{(j)}(t_0, s) \} \right\}.
\]

We estimate \( P_0^{(j)}(\cdot), P_1^{(j)}(\cdot) \) etc. using a Normal approximation for the \( j \)-th cause-specific case and control markers after a rank-based Z-score transformation and then empirically estimating the parameters of the approximating normal distributions. The detail is provided in additional file 1.

**Fractional Polynomial Estimator of I/D Cause-Specific AUC(t) (cFPL)**

The authors in [10] proposed a method for modelling I/D AUC(t) defined in equation (1) in the case of single event type. Their method directly models AUC(t) as a function of the event time \( t \) through a flexible fractional polynomials model proposed in [16]. We have extended it in the presence of competing risks as follows.

The AUC_{t}^{(j)} after transformation with link function \( \eta \) can be specified as a parametric function of \( t \) using fractional polynomials of degree \( L \):

\[
\eta(\text{AUC}_{t}^{(j)}(\beta)) = \beta_{j0} + \sum_{l=1}^{L} \beta_{jl} t^{(p_l)},
\]

where for \( l = 1, 2, ..., L \)

\[
t^{(p_l)} = \begin{cases} 
  t^{p_l} & \text{if } p_l \neq 0 \\
  \ln(t) & \text{if } p_l = 0,
\end{cases}
\]

with \( p_1 \leq p_2 \leq \cdots \leq p_L \) real-valued powers. As suggested in [10], we consider the power set \( p_1, \ldots, p_L \) in \{-2,-1,0.5,0,0.5,1,2\}, which is flexible enough to accommodate most applications. The set of regression parameters \( \beta = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{j7}) \) is then estimated by optimizing a likelihood function. A logit function considered for \( \eta(\cdot) \) similar to [10]. Let there are \( h_j \) number of failures due to cause \( j \) in the study. For each event time \( \{t_{(H)} \& \delta_H = j\} \), there are two types of random variables

\[
n_1(t_{(H)}) = \sum_k 1\{k : M_H > M_k | Z_H = t_{(H)}, \delta_H = j, k \in \mathbb{R}_{0(t_H)}\},
\]
and

\[ n_2(t_{(H)}) = \sum_k 1\{ k : M_H \leq M_k | Z_H = t_{(H)}, \delta_H = j, k \in \mathbb{R}_{0(t_{(H)})} \}, \]

where \( n_1(t_{(H)}) \) and \( n_2(t_{(H)}) \) are the number of concordant and number of discordant pairs, respectively. Note that, conditional on riskset \( \mathbb{R}_{0(t_{(H)})} \), the count \( n_1(t_{(H)}) \) follows a Binomial distribution with probability \( \text{AUC}_{t_{(H)}}^{(j)}(\beta_j) \). The partial-likelihood for \( j \)-th cause of failure is

\[ L(\beta_j) = \prod_{H=1}^{h_j} \{ \text{AUC}_{t_{(H)}}^{(j)}(\beta_j) \}^{n_1(t_{(H)})} \{ 1 - \text{AUC}_{t_{(H)}}^{(j)}(\beta_j) \}^{n_2(t_{(H)})} \]

Maximizing this partial-likelihood yields the parameter estimate \( \hat{\beta}_j \) of \( \beta_j \). Then, by using (4), we obtain \( \text{AUC}_{t_{(H)}}^{(j)}(\hat{\beta}_j) \) estimate as a smooth function of time \( t \) and \( \hat{\beta}_j \). For estimation, we evaluate the score equations that correspond to the proposed likelihood. The proposed likelihood is constructed in the spirit of the [10], however, unike [10], we have multiple causes of failure. We can write the log-likelihood in the following way

\[ l(\beta_j) = \sum_{i=1}^{n} \sum_{k=1}^{n} \delta_{ij} 1\{ Z_k > Z_i \} \{ 1\{ M_i > M_k \} \log(\text{AUC}_{Z_i}^{(j)}(\beta_j)) + 1\{ M_i \leq M_k \} \log(1 - \text{AUC}_{Z_i}^{(j)}(\beta_j)) \}, \]

where for \( j = 1, 2, \ldots, J \)

\[ \delta_{ij} = \begin{cases} 1, & \delta_i = j \\ 0, & \text{otherwise}. \end{cases} \]

We estimate \( \hat{\beta}_j \) of \( \beta_j \) as a solution of score vector \( U(\beta_j) = 0 \). The elements of \( U(\beta_j) \) are

\[ \frac{d}{d\beta_{j0}} l(\beta_j) = \sum_{i=1}^{n} \sum_{k=1}^{n} \delta_{ij} 1\{ Z_k > Z_i \} \{ 1\{ M_i > M_k \} \frac{d}{d\beta_{j0}} \frac{\text{AUC}_{Z_i}^{(j)}(\beta_j)}{\text{AUC}_{Z_i}^{(j)}(\beta_j)} - 1\{ M_i \leq M_k \} \frac{d}{d\beta_{j0}} \frac{\text{AUC}_{Z_i}^{(j)}(\beta_j)}{1 - \text{AUC}_{Z_i}^{(j)}(\beta_j)} \}, \]

where

\[ \frac{d}{d\beta_{j0}} \text{AUC}_{Z_i}^{(j)}(\beta_j) = \text{AUC}_{Z_i}^{(j)}(\beta_j) \{ 1 - \text{AUC}_{Z_i}^{(j)}(\beta_j) \}. \]

Again, for \( l = 1, 2, \ldots, 7 \), we get

\[ \frac{d}{d\beta_{jl}} l(\beta_j) = \sum_{i=1}^{n} \sum_{k=1}^{n} \delta_{ij} 1\{ Z_k > Z_i \} \{ 1\{ M_i > M_k \} \frac{d}{d\beta_{jl}} \frac{\text{AUC}_{Z_i}^{(j)}(\beta_j)}{\text{AUC}_{Z_i}^{(j)}(\beta_j)} - 1\{ M_i \leq M_k \} \frac{d}{d\beta_{jl}} \frac{\text{AUC}_{Z_i}^{(j)}(\beta_j)}{1 - \text{AUC}_{Z_i}^{(j)}(\beta_j)} \}, \]

where

\[ \frac{d}{d\beta_{jl}} \text{AUC}_{Z_i}^{(j)}(\beta_j) = Z_i^{(p_l)} \times \text{AUC}_{Z_i}^{(j)}(\beta_j) \{ 1 - \text{AUC}_{Z_i}^{(j)}(\beta_j) \}. \]
In order to make inferences about the proposed estimators of the cause-specific parameters, the major challenge lies in the fact that the proposed partial-likelihood cannot be treated as a regular likelihood function. Specifically, the asymptotic variance of the estimators is not the inverse of the negative second derivative of the partial-likelihood. We propose a sandwich variance estimator for the proposed global cause-specific AUC estimator below.

Asymptotic properties of cFPL estimator

In this section, we describe the asymptotic properties of the model parameter estimators. We state some regularity conditions in additional file 1. We summarize the asymptotic behavior of the regression parameter estimator in the following theorem.

**Theorem 2:** Under the regularity conditions, \( \hat{\beta}_j \) converges almost surely to \( \beta_{0j} \), where \( \sqrt{n}(\hat{\beta}_j - \beta_{0j}) \) converges to a multivariate Normal distribution with mean vector \( 0 \) and covariance matrix \( \Sigma_j^{-1} \Sigma_{j2} \Sigma_j^{-1} \). Here,

\[
\Sigma_{j1} = \mathbb{E}\{-\frac{d}{d\beta_j} f_{ik}(\beta_j); \beta_{0j}\},
\]

\[
\Sigma_{j2} = 4 \text{ Cov}\{g_{ik}(\beta_j), g_{ik'}(\beta_j); \beta_{0j}\},
\]

where, \( g_{ik}(\beta_j) = \frac{(f_{ik}(\beta_j) + f_{ik'}(\beta_j))}{2} \) and

\[
f_{ik}(\beta_j) = \int_0^\tau \int_0^\tau 1\{t > s\}\{1\{M_k > M_i\} \frac{d}{d\beta_j} \frac{\text{AUC}^{(j)}(\beta_j)}{\text{AUC}^{(j)}(\beta_j)} \ - 1\{M_i < M_k\} \frac{d}{d\beta_j} \frac{\text{AUC}^{(j)}(\beta_j)}{1 - \text{AUC}^{(j)}(\beta_j)}\} dN_i^{(j)}(s) dN_k^{(j)}(t),
\]

where \( N_i^{(j)}(\tau) \) counts number of events due to \( j \)-th cause of failure occurring over \([0, \tau]\). Theorem 2 can be proven in the spirit of the proof in [10]. However, the main difference is that, unlike [10], our likelihood construction account multiple causes of failure. An additional file shows the proof in more detail [see Additional file 1].

The covariance can be consistently estimated by \( \hat{\Sigma}_{j1}^{-1} \hat{\Sigma}_{j2} \hat{\Sigma}_{j1}^{-1} \), where

\[
\hat{\Sigma}_{j1} = \frac{-1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \delta_{ij} 1\{Z_k > Z_i\} \frac{d}{d\beta_j} \{1\{M_i > M_k\} \frac{d}{d\beta_j} \frac{\text{AUC}^{(j)}(\hat{\beta}_j)}{\text{AUC}^{(j)}(\hat{\beta}_j)} \ - 1\{M_i < M_k\} \frac{d}{d\beta_j} \frac{\text{AUC}^{(j)}(\hat{\beta}_j)}{1 - \text{AUC}^{(j)}(\hat{\beta}_j)}\},
\]

and

\[
\hat{\Sigma}_{j2} = \frac{4}{n(n-1)(n-2)} \sum_{i=1}^n \sum_{k \neq k', k' \neq i} g_{ik}(\hat{\beta}_j) g_{ik'}(\hat{\beta}_j)^T.
\]

**Corollary 1:** Under the regularity conditions, \( \text{AUC}^{(j)}(\hat{\beta}_j) \) is a consistent estimator of \( \text{AUC}^{(j)}(\beta_j) \). Furthermore, it follows that \( \sqrt{n}(\text{AUC}^{(j)}(\hat{\beta}_j) - \text{AUC}^{(j)}(\beta_{0j})) \) converges to a Normal distribution with mean 0 and variance \( \frac{d}{d\beta_j} \eta^{-1}(A \beta_j^T)^2_{\beta_j=\beta_j} A \hat{\Sigma}_{j1}^{-1} \hat{\Sigma}_{j2} \hat{\Sigma}_{j1}^{-1} A^T \), where \( A = (1, t^{p_1}, t^{p_2}, \ldots, t^{p_7}) \).

**Results**

Simulation Study

Extensive simulation studies are conducted in order to compare the performance of the cWMR, cFPL and semi-parametric [13] estimators for estimating \( \text{AUC}^{(j)} \).
We assume two causes of failure (i.e. \( j = 1, 2 \)) and a baseline marker \( M \) that is correlated with event time of cause 1, \( T^{(1)} \) but not with event time of cause 2, \( T^{(2)} \). We consider several parametric combinations under two major scenarios. For each setting, we generate 500 dataset with a sample size of \( n = 500 \) and for each simulated dataset 200 bootstrap simulations are performed. For each simulation, we estimate \( \text{AUC}^{(j)} \) at predicted time \( \log(t) = -1.5, -1.2, \ldots, 0.6 \). We report the average of bootstrap mean estimate of \( \text{AUC}^{(j)} \), absolute relative bias (ARB) of the estimate, average of model based standard error estimate (SE), average of the 500 bootstrap standard errors (BSE) and the coverage probability of the 90% confidence intervals (CI) for the estimates.

**Scenario 1:** We assume \((\log(T^{(1)}), M)\) jointly follows a bivariate Normal distribution (BVN) with correlation \( \rho \) i.e. \((\log(T^{(1)}), M) \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)\), where \( \mu_1 \) and \( \sigma_1 \) are mean and standard deviation (SD) of \( \log(T^{(1)}) \) and \( \mu_2 \) and \( \sigma_2 \) are mean and SD of \( M \). We show the results for \( \mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1, \) and \( \rho = -0.7 \). We consider a negative correlation between the marker and the event time which implies that higher marker value is more indicative of poor survival outcome and hence it is indicative of shorter event time. We further assume \( \log(T^{(2)}) \sim N(0, 1) \) and \( \log(C) \sim N(0, 1) \), such that approximately 20% subjects are censored. Since \( T^{(2)} \) and \( M \) are independent, the I/D ROC curve for the competing cause of failure (i.e. cause 2) lies diagonally on the null ROC curve.

**Scenario 2:** We focus on a heterogeneous population where the marginal relationship between \( M \) and \( T^{(1)} \) is non-monotone, while \( M \) and \( T^{(2)} \) are independent. The heterogeneous population comprised two distinct subgroups \((G = 0 \text{ or } 1)\) and for \( G = 1 \), \( M \) is also independent of \( \log(T^{(1)}) \). The distribution of \((\log(T^{(1)}), M)\) follows a mixture of two BVNs. We show the results for two different parameter combinations:

\[
(a) \quad (\log(T^{(1)}), M) \sim \begin{cases} 
N_2(-1.5, -1.5, 1, 1, 0), & \text{G} = 1, \\
N_2(0, 2, 1, 1, -0.8), & \text{G} = 0.
\end{cases}
\]

\[
(b) \quad (\log(T^{(1)}), M) \sim \begin{cases} 
N_2(-1.5, 2, 1, 1, 0), & \text{G} = 1 \\
N_2(0, 0, 1, 1, -0.8), & \text{G} = 0.
\end{cases}
\]

We assume \( G \sim \text{Bernoulli}(0.2) \). Note that the semi-parametric approach is biased because of violation of monotonicity in this scenario, and therefore not estimated. The resulting \( \text{AUC}^{(j)} \) curves mimic the relationship in the LT data as will be demonstrated later.

In Tables 1-3, we summarize the simulation results for scenarios 1-2 respectively. Table 1 demonstrates that the estimates of \( \text{AUC}^{(j)} \) is less biased when derived using cWMR and cFPL compared with the semi-parametric method. For instance, for cause 1, the ARB in the estimate of \( \text{AUC}^{(j)} \) for the cWMR is 0.36% corresponding to predicted time \( \log(t) = -1.5 \). However, for the semi-parametric approach this is 2.16%. For large predicted time (i.e. \( \log(t) = 0.9 \)) both the cFPL and the semi-parametric estimates show large ARB. The bootstrap standard errors for both cWMR and cFPL methods are close to their corresponding model-based standard error estimates. For cWMR, the coverage probabilities of estimated \( \text{AUC}^{(j)} \) based
on 90% estimated confidence intervals are very close to the nominal value of 0.9 across all predicted times. When we have sufficient data around the given predicted times, say \(-1.5 \leq \log(t) \leq 0.6\), the coverage probabilities of estimated AUC\(_t^{(j)}\) using cFPL are very close to the nominal value. However, the coverage probability is much lower than the nominal level when riskset size is 4 at given time \(\log(t) = 0.9\). This is perhaps the issue of oversmoothing. It could be avoided by choosing the large predicted times as the 90th percentile of the observed survival time points [10]. In table 2 and table 3, we compare the cWMR and cFPL estimates as the semi-parametric estimator is known to be highly biased under scenario 2. Note that, in table 2, the AUC value for cause 1 increases steadily between \(-1.5 \leq \log(t) \leq 0\) and then start decreasing. This kind of non-monotone pattern may be due to violation of monotone relationship between marker and event time. The estimates of AUC\(_t^{(j)}\) for both cWMR and cFPL are close to their corresponding true values when the predicted time is small. In these comparisons cFPL performs slightly better than cWMR. The cFPL method yields substantially greater variances for large values of \(\log(t)\) compared to the cWMR. For both methods, the estimated coverage probabilities are very close to the nominal coverage probability of 90% except for edges.

Overall, our results demonstrate that both cWMR and cFPL appear to perform well compared to semi-parametric approach in terms of bias and standard errors. In addition, in scenario 2 where the monotonicity between marker and event time is violated, the semi-parametric approach is known to be biased. However, both our proposed methods perform adequately well.

Application

We demonstrate the proposed methods for estimating AUC\(_t^{(j)}\) using LT data from a retrospective study conducted at the McGill University Health Center [3]. The LT study included 547 patients who underwent LT between 1991 and 2012 and who met the criteria: patient with graft survival > 12 months; serum fibrosis biomarkers including FIB-4 and NAFLD score available at 1 year after LT; and a minimum follow-up of 1 year. The study found that serum fibrosis markers performed well in predicting death and graft loss in LT recipients. According to the authors, this is the first study to establish the prognostic value of the fibrosis markers in a large cohort of LT recipients over a long-term follow-up period. We further analyzed a subset of the subjects \((n = 423)\) after excluding subjects with missing outcome and/or marker values. During the study period, 64 patients who underwent LT died due to graft-related causes (e.g. graft failure). However, 62 patients died of causes that are unrelated to their transplantation (e.g. sepsis, cardiovascular disease, renal, respiratory failure etc.). Different causes of death led to a competing risks situation. The research objective is to evaluate the performance of FIB-4 as a marker to discriminate between subjects who died due to graft related causes and those who died of non-graft related causes after the LT. The top two subplots in Figure 1 show the estimated AUC obtained using cWMR and cFPL methods for graft-related, non-graft related and all-cause death (considering both graft and non-graft related death as events of interest). Irrespective of the methods, the AUC curve for all-cause mortality is biased compared to the curves for graft and no-graft related deaths. The magnitude of bias is downward compared to graft related deaths and the bias
is upward compared to non-graft related deaths. It indicates that consideration of all-cause mortality as an event of interest instead of competing risks will result in biased accuracy estimates. Therefore, it leads us to analyse the LT data using the methodology proposed here in the presence of competing risks. Figure 1 also illustrates estimated $AUC_{t}^{(j)}$ curves with 95% CI for graft and non-graft related causes using cWMR and cFPL methods. In order to estimate $AUC_{t}^{(j)}$ using cWMR, the choice of neighborhood was 1.35 years back and forward. This bandwidth of 1.35 years is obtained after minimizing the IMSE. The estimated $AUC_{t}^{(j)}$ curves of the FIB-4 for the graft related death sustains a high predictive value (above 0.65 for most of the study duration) irrespective of the estimation procedures. For example, over the first 3 years of follow-up the estimated $AUC_{t}^{(j)}$ under cFPL approach ranges between 0.89 to 0.56. This implies that on any day, $t$, during the first 3 years of follow-up, the probability that a subject after LT who dies due to graft related causes on day $t$ having a FIB-4 value greater than a subject who survives beyond day $t$ is at least 0.56. Overall, $AUC_{t}^{(j)}$ curve estimated using cFPL is a smooth function over predicted time while estimated $AUC_{t}^{(j)}$ curve of cWMR is less smooth. We could not find any definitive reasons for the cause-specific AUC($t$) for graft related cause increases between years 3.3 and 8. However, in reference to results from simulation scenario 2, we observed that if there is any latent (or unobserved) heterogeneity in the data, the AUC curve shows non-monotone trend over time. On the other hand, these two curves for non-graft related death are almost flat around the horizontal line at AUC($t$) = 0.5. Furthermore, 95% CI’s of the estimates of AUC contain the null value of 0.5 which implies that FIB-4 is non-informative as a prognostic marker for non-graft related events. Therefore, FIB-4 as a baseline marker does not discriminate patients with non-graft related deaths after LT, which is expected. In addition, under cFPL method the CIs of the estimates of $AUC_{t}^{(j)}$ curves at the tails of the study period are relatively wider/narrower than that of under cWMR method. The CIs of the estimates using cFPL are wider in both small and large predicted times because cFPL may have oversmoothing issue especially towards the start and end of the study. Finally, our analysis indicates a better discrimination by FIB-4 for graft related deaths than non-graft related deaths after LT for most of the study duration.

**Discussion**

Measures of calibration and discrimination are integral parts to evaluate prognostic accuracy of a marker. Calibration indices provide information on how close the predicted risks are to the observed risks while discrimination indices measure whether markers correctly discriminate subjects who will subsequently experience the event of interest at or by time $t$ from those who will not experience any event by $t$. Calibration measures e.g. an expected Brier score for competing risks and corresponding estimator were provided in [17]. It measures the closeness of the observed event status and the model predicted event probabilities in the presence of competing risks. Here, we primarily focus on the discrimination accuracy of marker. The main goal of this manuscript is to estimate cause-specific AUC($t$) of a baseline marker in the presence of competing risks. During such estimation, analysts often censor subjects when a competing event occurs. For instance, the outcome in LT study is
time-to-death attributable to LT-related causes, an analyst may consider a subject as censored once that subject dies of causes unrelated to LT. Because subjects who died of non LT-related causes are not at risk of dying due to LT, censoring these competing risks events (informative censoring) may lead to distorted risk estimates [5] and subsequently biased accuracy estimators. Alternatively, some may consider a composite event where deaths attributable to LT and non-LT deaths are merged together as any adverse events. In the application section, we mimic this to our LT data to show the drawback of using a composite endpoint to demonstrate the importance of considering competing risks in prognostic accuracy estimation. It is evident from our results that simply considering a composite event instead of competing events introduces bias in accuracy estimation. For competing risks analysis, the influence of covariate can be evaluated in relation to cause-specific hazard or the sub-distribution hazard of different causes of failure. For estimating cause-specific hazard, when a subject experience any event, they are removed from the subsequent riskset. In contrast, for estimating sub-distribution hazard [18], a subject who experiences a competing event is not removed from the riskset at that time, but rather is censored at the end of the follow-up.

We propose a local and a global estimators of cause-specific AUC(t) for right-censored survival time outcome in the presence of competing risks. In [13], a semi-parametric approach based on Cox model was suggested for estimating cause-specific AUC(t). This approach provides bias estimate when the association between event time and marker is non-monotone. Also, this is not a one-step approach for estimating AUC trajectory over time and it requires longer computation time. In addition, their method lacks analytical development of the large sample properties for statistical inference. These motivate us to propose new estimators for estimating cause-specific AUC(t). As pointed out earlier, the observed proportion of controls ranked lower than the (cause-specific) case, generally leads to unstable estimates because it is based on a single 'case' subject who had an event of interest at the specific time. Hence, the estimation of cause-specific AUC that is based on observed proportions requires some degree of smoothing. Our proposed estimators implement the degree of smoothing in different ways. The local estimator - cause-specific weighted mean rank (cWMR) - is a local average of unsmoothed time specific observed cause-specific AUCs within a neighborhood of a given time t. cWMR is sensitive to neighborhood span. The width of the neighbourhood directly influenced the smoothness of the curves especially towards the end of the study when size of the riskset gets very small. We have considered cross-validation approach to choose optimal neighborhood span. Use of adaptive smoothing techniques, for example, fixing the number of neighbors instead of a fixed bandwidth may be useful. Instead of using local average of concordance within a neighborhood of time, we propose an alternative method based on global curve fitting approach, cFPL, which estimates the cause-specific AUC as a function of time through a flexible fractional polynomial function. It expresses the unsmoothed intermediate AUCs as a function of fractional polynomials of time and then estimates the coefficients of polynomials through a partial likelihood optimization. cFPL overcomes the issue of sensitivity to neighborhood span of cWMR. However, oversmoothing is an issue with cFPL in both small and large time points. In terms of computation time, cWMR is computationally efficient compared to cFPL. We also develop the large sample properties of
both estimators as well as their corresponding analytical variances. Our simulation study suggests that these two estimators perform very well compared to the existing semi-parametric method for measuring cause-specific AUC. The performance has been evaluated in terms of absolute relative bias and coverage probability. Between our two proposed approaches, both estimators show similar level of relative bias particularly for small predicted time in simulation studies. However, for large times the cFPL estimator shows large bias compared to the cWMR estimator. In addition, the coverage probability of cFPL estimator is very different from the nominal coverage probability particularly at the edges. In LT study, our goal is to evaluate the performance of FIB-4 as a baseline biomarker to discriminate between subjects who died due to graft related causes against those who died of non-graft related causes after the liver transplantation. Our analysis indicates a better discrimination of graft related deaths than non-graft related deaths after LT for most of the study duration. The estimation methods assume that the censoring time is independent of the survival time. It warrants additional research to allow covariate dependent censoring. Furthermore, in our settings, we do not consider time-varying marker. However, the proposed methodology is equally applicable to settings where time-varying marker exist.

Finally, between two estimators, which one is better: cWMR and cFPL? There is no general answer. The former is more adaptive to the local changes while the latter is good for an overall description. Our advice is to perform sensitivity analysis in which the choice of estimation methods vary. Also, for all methods the estimates at the higher time range are unstable, emphasizing the fact that one should have a sufficient number of events for estimation of the cause-specific AUC. This problem may be avoided by choosing a sufficiently wide neighborhood (e.g. fixing the number of neighbors) when using the cWMR. For cFPL, we can choose relevant short future time horizon over which we have sufficient cause-specific cases for numerically stable results.

Our study has some limitations. The first limitation is related to the bandwidth selection in cWMR approach. To implement the proposed cWMR estimator in practice, the appropriate bandwidths must be chosen. In this manuscript, we have used leave-one-out cross validation approach [9]. It would be interesting to perform sensitivity analysis by varying different bandwidth selectors. Next, our proposed approaches are not applicable when the data have missing information. However, following a typical imputation method, our approach could be applied to the imputed data. However, details of the inference and sensitivity to the imputation methods is yet to be explored. Furthermore, the variance calculation of cFPL estimator is computationally intensive. Future research to explore alternate methods for efficient computation may be worthwhile. Moreover, the estimation methods assume that the censoring time is independent of the survival time. Additional research to allow covariate dependent censoring is warranted. In addition, in our settings, we do not consider time-varying marker. However, the proposed methodology is equally applicable to settings where time-varying marker exist. This is to be explored in the future.
Conclusions
We developed estimation procedures of estimating time-dependent prognostic accuracy measures for a right-censored time-to-event outcome in the presence of competing risks. The proposed methods are non-parametric, direct and computationally simple that will overcome the shortcomings of the existing approach. The methods will provide computationally efficient options for assessing the prognostic accuracy of markers for time-to-event outcome.

Supplementary information
Supplementary information accompanies this paper in additional file 1.

Abbreviations
KDRI: Kidney donor risk index; LT: liver-transplantation; FIB-4: fibrosis score 4; NAFLD: nonalcoholic fatty liver disease fibrosis score; ROC: Receiver operating characteristic curve; AUC: Area under the ROC curve; I: Incident cases; D: Dynamic Controls; C: Cumulative cases; WMR: Weighted Mean Rank; FPL: Fractional Polynomial; cWMR: Weighted Mean Rank Estimator of I/D Cause-Specific AUC(t); cFPL: Fractional Polynomial Estimator of I/D Cause-Specific AUC(t); BVN: bivariate Normal distribution; ARB: absolute relative bias; BSE: bootstrap standard errors; CI: Confidence interval; SD: standard deviation.

Declarations
Ethics approval and consent to participate
The need for informed consent had been waived due to the retrospective nature of the study.

Consent to publish
Not Applicable

Availability of data and materials
The liver transplant dataset is not publicly available due to patient privacy restrictions, but could be made available from the corresponding author pending appropriate approval. The R codes for simulations and data analysis are available from the corresponding author upon request.

Competing interests
The authors declare that they have no competing interests.

Funding
We would like to acknowledge our funding agencies. PSC is supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2017-06100) and Fonds de la Recherche en Santé du Québec (Junior 1 Salary Award). GS is supported by a Junior 1 and 2 Salary Award from FRQS (27127 and 267806) and research salary from the Department of Medicine of McGill University.

Author’s contributions
This article is conceptualized by RD and PSC. RD carried out the methodological work, programming, data analysis and writing, under the supervision of PSC. GS provided the data. All authors reviewed the manuscript.

Acknowledgements
The authors thank Professor James A. Hanley at McGill University for many constructive inputs during write-up which lead to improvement of the manuscript.

Author details
1Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada. 2Division of Gastroenterology and Hepatology, McGill University Health Centre, Montreal, Canada. 3Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada.

References
1. Wilson, P.W., D’Agostino, R.B., Levy, D., Belanger, A.M., Silbershatz, H., Kannel, W.B.: Prediction of coronary heart disease using risk factor categories. Circulation 97(18), 1837–1847 (1998)
2. Gail, M.H., Brinton, L.A., Byar, D.P., Corle, D.K., Green, S.B., Schairer, C., Mulvihill, J.J.: Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. JNCI: Journal of the National Cancer Institute 81(24), 1879–1886 (1989)
3. Bhat, M., Ghali, P., Rollet-Kurhajec, K.C., Bhat, A., Wong, P., Deschenes, M., Sebastiani, G.: Serum fibrosis biomarkers predict death and graft loss in liver transplantation recipients. Liver Transplantation 21(11), 1383–1394 (2015)
4. Vallet-Pichard, A., Mallet, V., Nalpas, B., Verkarre, V., Nalpas, A., Dhalluin-Venier, V., Fontaine, H., Pol, S.: Fib-4: an inexpensive and accurate marker of fibrosis in hcv infection. comparison with liver biopsy and fibrotest. Hepatology 46(1), 32–36 (2007)
5. Prentice, R.L., Kalbfleisch, J.D., Peterson Jr, A.V., Flookoy, N., Farewell, V.T., Breslow, N.E.: The analysis of failure times in the presence of competing risks. Biometrics, 541–554 (1978)
6. Harrell, F.E., Califf, R.M., Pryor, D.B., Lee, K.L., Rosati, R.A.: Evaluating the yield of medical tests. Jama 247(18), 2543–2546 (1982)
7. Heagerty, P.J., Lumley, T., Pepe, M.S.: Time-dependent roc curves for censored survival data and a diagnostic marker. Biometrics 56(2), 337–344 (2000)
8. Heagerty, P.J., Zheng, Y.: Survival model predictive accuracy and roc curves. Biometrics 61(1), 92–105 (2005)
9. Saha-Chaudhuri, P., Heagerty, P.: Non-parametric estimation of a time-dependent predictive accuracy curve. Biostatistics 14(1), 42–59 (2013)
10. Shen, W., Ning, J., Yuan, Y.: A direct method to evaluate the time-dependent predictive accuracy for biomarkers. Biometrics 71(2), 439–449 (2015)
11. Schmid, M., Potapov, S.: A comparison of estimators to evaluate the discriminatory power of time-to-event models. Statistics in medicine 31(23), 2588–2609 (2012)
12. Song, X., Zhou, X.-H.: A semiparametric approach for the covariate specific roc curve with survival outcome. Statistica Sinica, 947–965 (2008)
13. Saha, P., Heagerty, P.: Time-dependent predictive accuracy in the presence of competing risks. Biometrics 66(4), 999–1011 (2010)
14. Zheng, Y., Cai, T., Jin, Y., Feng, Z.: Evaluating prognostic accuracy of biomarkers under competing risk. Biometrics 68(2), 388–396 (2012)
15. Pepe, M.S.: The Statistical Evaluation of Medical Tests for Classification and Prediction. Medicine, Oxford (2003)
16. Royston, P., Altman, D.G.: Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. Journal of the Royal Statistical Society: Series C (Applied Statistics) 43(3), 429–453 (1994)
17. Schoop, R., Beyersmann, J., Schumacher, M., Binder, H.: Quantifying the predictive accuracy of time-to-event models in the presence of competing risks. Biometrical Journal 53(1), 88–112 (2011)
18. Fine, J.P., Gray, R.J.: A proportional hazards model for the subdistribution of a competing risk. Journal of the American statistical association 94(444), 496–509 (1999)

Figures

Figure 1  Estimates of incident/dynamic cause-specific AUC(t) curves using weighted mean rank (cWMR) and fractional polynomial (cFPL) for Liver Transplantation data. Plots (a) and (b) illustrate the incident/dynamic AUC(t) curve for all-cause mortality, incident/dynamic cause-specific AUC(t) curves for graft-related death and non-graft-related death. Plots (c) and (d) are estimated using cWMR and cFPL methods, respectively. Plots (e) and (f) illustrate the incident/dynamic cause-specific AUC(t) curves for graft-related death with pointwise 95% confidence intervals (CI) using cWMR and cFPL methods, respectively. Plots (c) and (f) illustrate the incident/dynamic cause-specific AUC(t) curves for non-graft-related death with pointwise 95% CIs using cWMR and cFPL methods, respectively.

Tables
Table 1: Simulation results for estimation of incident/dynamic cause specific AUC of a single marker in Scenario 1, comparing the methods cWMR, Semi-parametric and cFPL. Average of bootstrap mean (Mean), average of bootstrap standard errors (BSE), absolute relative bias (ARB), model-based standard error (SE), coverage probability (CP)(nominal level is 90 percentage) of AUC.

| Cause | Log(t) | ∑_{i=1}^{n} R_i(t) | AUC | Mean | ARB | BSE | SE | CP | Mean | ARB | BSE | CP | Mean | ARB | BSE | SE | CP |
|-------|--------|---------------------|-----|------|-----|-----|----|----|------|-----|-----|----|------|-----|-----|----|----|
| -1.5  | 407    | 0.833               |     | 0.830| 0.36 | 0.031| 0.030| 88.0|       | 0.815| 2.16 | 0.046| 90.0|       | 0.833| 0   | 0.026| 0.028| 93.1|
| -1.2  | 347    | 0.802               |     | 0.800| 0.25 | 0.030| 0.029| 89.0|       | 0.783| 2.37 | 0.050| 91.6|       | 0.802| 0   | 0.025| 0.027| 93.5|
| -0.9  | 272    | 0.771               |     | 0.771| 0.0  | 0.030| 0.029| 84.1|       | 0.753| 2.33 | 0.053| 92.6|       | 0.771| 0   | 0.026| 0.028| 92.1|
| -0.6  | 192    | 0.743               |     | 0.743| 0.0  | 0.032| 0.031| 87.3|       | 0.727| 2.15 | 0.055| 91.9|       | 0.742| 0.1 | 0.030| 0.031| 91.3|
| 1     | -0.3  | 119                 |     | 0.719| 0.41 | 0.038| 0.036| 88.6|       | 0.707| 1.25 | 0.057| 90.1|       | 0.716| 0   | 0.036| 0.038| 93.1|
|       | 0.0    | 63                  |     | 0.693| 0.43 | 0.048| 0.046| 86.9|       | 0.684| 1.23 | 0.058| 92.0|       | 0.694| 0.14| 0.053| 0.050| 92.8|
|       | 0.3    | 28                  |     | 0.672| 0.6  | 0.071| 0.064| 89.7|       | 0.670| 0.29 | 0.056| 92.2|       | 0.678| 0.89| 0.094| 0.084| 93.1|
|       | 0.6    | 10                  |     | 0.654| 0.76 | 0.126| 0.104| 86.0|       | 0.653| 0.15 | 0.060| 92.6|       | 0.653| 0.15| 0.211| 0.186| 87.8|
|       | 0.9    | 4                   |     | 0.639| 2.91 | 0.189| 0.169| 89.0|       | 0.578| 9.54 | 0.119| 82.1|       | 0.599| 6.10| 0.353| 0.586| 75.6|
| -1.5  | 407    | 0.5                 |     | 0.501| 0.2  | 0.051| 0.050| 87.3|       | 0.504| 0.8  | 0.082| 90.2|       | 0.5   | 0   | 0.045| 0.049| 92.1|
| -1.2  | 347    | 0.5                 |     | 0.501| 0.2  | 0.045| 0.044| 88.5|       | 0.503| 0.6  | 0.078| 91.2|       | 0.5   | 0   | 0.040| 0.042| 91.9|
| -0.9  | 272    | 0.5                 |     | 0.501| 0.2  | 0.041| 0.041| 88.8|       | 0.499| 0.4  | 0.075| 90.0|       | 0.499 | 0.2 | 0.037| 0.041| 92.1|
| -0.6  | 192    | 0.5                 |     | 0.500| 0.0  | 0.042| 0.041| 89.1|       | 0.501| 0.4  | 0.071| 86.8|       | 0.499 | 0.2 | 0.041| 0.041| 90.9|
| 2     | -0.3  | 119                 |     | 0.499| 0.2  | 0.047| 0.045| 90.5|       | 0.494| 1.2  | 0.069| 90.2|       | 0.498 | 0.4 | 0.047| 0.049| 93.2|
|       | 0.0    | 63                  |     | 0.499| 0.2  | 0.057| 0.054| 86.3|       | 0.496| 0.8  | 0.065| 91.4|       | 0.498 | 0.4 | 0.068| 0.061| 91.3|
|       | 0.3    | 28                  |     | 0.499| 0.2  | 0.080| 0.073| 87.3|       | 0.505| 1.0  | 0.063| 90.4|       | 0.496 | 0.8 | 0.120| 0.107| 92.5|
|       | 0.6    | 10                  |     | 0.500| 0.0  | 0.136| 0.111| 83.3|       | 0.496| 1.6  | 0.062| 89.2|       | 0.498 | 0.4 | 0.247| 0.223| 87.7|
|       | 0.9    | 4                   |     | 0.499| 0.2  | 0.209| 0.191| 87.1|       | 0.471| 5.8  | 0.103| 92.0|       | 0.495 | 1.0 | 0.376| 0.497| 62.8|
Table 2 Simulation results for estimation of incident/dynamic cause-specific AUC of a single marker in Scenario 2(a), comparing the methods- cWMR & cFPL. Average of bootstrap mean (Mean), average of bootstrap standard errors (SE), absolute relative bias (ARB(%)), model-based standard error (SE), and coverage probability (CP(%))-( nominal level —90%) of AUC.

| Cause | Log(t) | AUC  | Mean  | ARB   | BSE   | SE    | CP(%) | Mean  | ARB   | BSE   | SE    | CP(%) |
|-------|--------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|       |        |      | cWMR  | cFPL  |       |       |       |       |       |       |       |       |
| -1.5  | 0.535  | 0.538| 0.561 | 0.061 | 0.052 | 83.6  | 0.534 | 0.187 | 0.047 | 0.051 | 92.2  |
| -1.2  | 0.600  | 0.598| 0.333 | 0.054 | 0.049 | 86.0  | 0.597 | 0.500 | 0.043 | 0.048 | 93.5  |
| -0.9  | 0.649  | 0.643| 0.924 | 0.049 | 0.047 | 90.3  | 0.645 | 0.616 | 0.042 | 0.043 | 93.5  |
| -0.6  | 0.680  | 0.674| 0.882 | 0.047 | 0.047 | 90.5  | 0.678 | 0.294 | 0.043 | 0.046 | 93.0  |
| 1     | -0.3   | 0.692| 0.690 | 0.289 | 0.049 | 0.050 | 92.7  | 0.693 | 0.144 | 0.052 | 0.05  | 89.4  |
|       | 0.0    | 0.697| 0.698 | 0.143 | 0.057 | 0.057 | 89.2  | 0.697 | 0.00  | 0.080 | 0.069 | 92.6  |
|       | 0.3    | 0.688| 0.697 | 1.308 | 0.080 | 0.073 | 87.0  | 0.702 | 2.034 | 0.139 | 0.128 | 92.6  |
|       | 0.6    | 0.681| 0.695 | 2.056 | 0.135 | 0.113 | 83.0  | 0.655 | 3.818 | 0.284 | 0.294 | 83.2  |
|       | -1.5   | 0.500| 0.0   | 0.049 | 0.052 | 88.5  | 0.5005| 0.1   | 0.048 | 0.052 | 90.7  |
|       | -1.2   | 0.5   | 0.4999| 0.02  | 0.043 | 0.046 | 87.5  | 0.4988| 0.24  | 0.042 | 0.044 | 92.7  |
|       | -0.9   | 0.5   | 0.4974| 0.52  | 0.040 | 0.043 | 88.0  | 0.4979| 0.42  | 0.040 | 0.044 | 92.3  |
|       | -0.6   | 0.5   | 0.4976| 0.48  | 0.041 | 0.044 | 90.0  | 0.4975| 0.50  | 0.044 | 0.044 | 91.8  |
| 2     | -0.3   | 0.5   | 0.4985| 0.30  | 0.046 | 0.049 | 89.7  | 0.4973| 0.54  | 0.052 | 0.054 | 94.0  |
|       | 0.0    | 0.5   | 0.4968| 0.64  | 0.057 | 0.049 | 89.9  | 0.4967| 0.66  | 0.079 | 0.068 | 92.1  |
|       | 0.3    | 0.5   | 0.4952| 0.96  | 0.080 | 0.080 | 87.2  | 0.4953| 0.94  | 0.143 | 0.128 | 94.2  |
|       | 0.6    | 0.5   | 0.5011| 0.22  | 0.137 | 0.124 | 83.4  | 0.4969| 0.62  | 0.269 | 0.289 | 87.8  |
Table 3 Simulation results for estimation of incident/dynamic cause-specific AUC of a single marker in Scenario 2(b), comparing the methods- cWMR & cFPL. Average of bootstrap mean (Mean), average of bootstrap standard errors (BSE), absolute relative bias (ARB(%)), model-based standard error (SE) and coverage probability (CP(%))-(nominal level —90%) of AUC.

| Cause | log(t) | AUC | Mean  | ARB  | BSE  | SE   | CP(%) | Mean  | ARB  | BSE  | SE   | CP(%) |
|-------|--------|-----|-------|------|------|------|-------|-------|------|------|------|-------|
| -1.5  | 0.843  |     | 0.843 | 0.0  | 0.025 | 92.0 | 0.843 | 0.0  | 0.017 | 0.018 | 93.3 |
| -1.2  | 0.827  |     | 0.830 | 0.31 | 0.026 | 94.0 | 0.829 | 0.245 | 0.018 | 0.02  | 92.3 |
| -0.9  | 0.807  |     | 0.809 | 0.26 | 0.028 | 93.8 | 0.810 | 0.042 | 0.020 | 0.021 | 89.9 |
| -0.6  | 0.784  |     | 0.786 | 0.25 | 0.031 | 94.8 | 0.788 | 0.451 | 0.025 | 0.026 | 93.6 |
| 1     | -0.3   | 0.760 | 0.765 | 0.62 | 0.036 | 94.0 | 0.761 | 0.162 | 0.033 | 0.033 | 92.9 |
|       | 0.0    | 0.737 | 0.747 | 1.41 | 0.048 | 90.8 | 0.737 | 0.016 | 0.054 | 0.046 | 92.7 |
|       | 0.3    | 0.716 | 0.727 | 1.60 | 0.073 | 88.0 | 0.728 | 1.66  | 0.096 | 0.085 | 89.7 |
|       | 0.6    | 0.696 | 0.710 | 2.07 | 0.131 | 83.9 | 0.692 | 0.46  | 0.218 | 0.179 | 84.6 |
| -1.5  | 0.5    | 0.5029 | 0.58 | 0.049 | 0.052 | 88.9 | 0.5028 | 0.56  | 0.048 | 0.0517 | 92.3 |
| -1.2  | 0.5    | 0.5017 | 0.34 | 0.043 | 0.046 | 90.0 | 0.5025 | 0.50  | 0.042 | 0.045 | 94.3 |
| -0.9  | 0.5    | 0.5004 | 0.08 | 0.040 | 0.043 | 89.9 | 0.5003 | 0.06  | 0.040 | 0.044 | 93.7 |
| -0.6  | 0.5    | 0.4987 | 0.26 | 0.042 | 0.044 | 88.7 | 0.4983 | 0.34  | 0.045 | 0.044 | 92.1 |
| 2     | -0.3   | 0.5    | 0.5005 | 0.10 | 0.046 | 0.049 | 93.1 | 0.4991 | 0.178 | 0.052 | 0.054 | 93.5 |
|       | 0.0    | 0.5    | 0.5006 | 0.12 | 0.057 | 0.06  | 90.5 | 0.5030 | 0.60  | 0.080 | 0.07  | 92.5 |
|       | 0.3    | 0.5    | 0.5043 | 0.86 | 0.082 | 0.081 | 89.3 | 0.5039 | 0.78  | 0.146 | 0.126 | 95.1 |
|       | 0.6    | 0.5    | 0.5042 | 0.84 | 0.142 | 0.124 | 87.5 | 0.4938 | 0.62  | 0.276 | 0.259 | 85.0 |