DYNAMIC PREDICTION MODELS FOR DATA WITH COMPETING RISKS

by

Qing Liu

B.S. Biological Sciences, Shanghai Jiao Tong University, China,

2007

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Qing Liu

It was defended on

December 1st, 2014

and approved by

Chung-Chou H. Chang, Ph.D., Professor
Departments of Medicine and Biostatistics
School of Medicine and Graduate School of Public Health, University of Pittsburgh

Yu Cheng, Ph.D., Associate Professor, Department of Statistics
Kenneth P. Dietrich School of Arts and Sciences, University of Pittsburgh

Joseph P. Costantino, Dr.P.H., Professor, Department of Biostatistics
Graduate School of Public Health, University of Pittsburgh

Mary Ganguli, M.D., M.P.H., Professor
Departments of Psychiatry, Epidemiology, and Neurology
School of Medicine, University of Pittsburgh

Gong Tang, Ph.D., Associate Professor, Department of Biostatistics
Graduate School of Public Health, University of Pittsburgh

Dissertation Director: Chung-Chou H. Chang, Ph.D., Professor
Departments of Medicine and Biostatistics
School of Medicine and Graduate School of Public Health, University of Pittsburgh

DYNAMIC PREDICTION MODELS FOR DATA WITH COMPETING RISKS

Qing Liu, PhD

University of Pittsburgh, 2014

ABSTRACT: Prediction of cause-specific cumulative incidence function (CIF) is of primary interest to clinical researchers when conducting statistical analysis involving competing risks. The predicted CIFs need to be dynamically updated by incorporating the time-dependent information measured during follow-up. However, dynamic prediction of the conditional CIFs requires simultaneously updating the overall survival and the CIF while adjusting for the time-dependent covariates and the time-varying covariate effects which is complex and challenging. In this study, we extended the landmark Cox models to data with competing risks, because the landmark Cox models provide a simple way to incorporate various types of time-dependent information for data without competing risks. The resulting new models are called landmark proportional subdistribution hazards (PSH) models. In this study, we first investigated the properties of the Fine-Gray model under non-PSH and proposed a robust risk prediction procedure which is not sensitive to the PSH assumption. Then, we developed a landmark PSH model and a more comprehensive landmark PSH supermodel. The performance of our models was assessed via simulations and through analysis of data from a multicenter clinical trial for breast cancer patients. As compared with other dynamic predictive models, our proposed models exhibited three advantages. First, our models are robust against violations of the PSH assumption and can directly predict the conditional CIFs bypassing the estimation of overall survival and greatly simplify the prediction procedure. Second, our landmark PSH supermodel enables users to make predictions with a set of landmark points in one step. Third, the proposed models can simply incorporate various

types of time-varying information. Finally, our models are not computationally intensive and can be easily implemented with existing statistical software.

Public Health Significance: Prognostic models for predicting the absolute risk of a patient in having a disease are very useful in performing risk stratification and making treatment decisions. Since the patient's prognosis can change over time, it is necessary to update the risk prediction accordingly. The dynamic prediction models developed in this study can provide more accurate prognoses over the course of disease progression and will be helpful to physicians in adopting personalized treatment regimes.

Keywords: Competing risks; cumulative incidence function; dynamic prediction; landmark analysis; proportional subdistribution hazards; time-dependent variables; time-varying covariate effects.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

For data with competing risks, although much attention was given to identify the prognostic or risk factors on cause-specific hazard rate and on the cause-specific failure probabilities, clinicians are showing great interest in predicting absolute risk of a cause-specific failure in the presence of competing risks.

In quantifying the likelihood of failure from a specific cause of interest, the complement of the Kaplan-Meier estimator where competing events are treated as censored is inappropriate and not interpretable when the main event and competing events are correlated. In contrast, the cumulative incidence function (CIF, also referred to as sub-distribution) is more proper in describing a cause-specific failure probability with no assumptions of dependencies between competing events [21]. The CIF is defined as the cumulative probability that the event of interest occurs before a given time t in a framework where a subject is exposed to multiple causes of failure [13].

Predicting a cause-specific CIF based on a patient's prognostic information collected at the time of diagnosis or at the start of treatment is essential in risk stratification and in decision-making process. To predict CIFs, the most commonly used regression procedure is the proportional sub-distribution hazards (PSH) model proposed by Fine and Gray (1999) [8] which is easy to implement and can yield a simple form of the estimated CIF after adjusting for multiple discrete and continuous covariates. In practice, however, the proportionality assumption of the PSH model is often violated, especially for clinical studies with long-term follow up. When it is used on non-proportional sub-distribution hazards, the standard PSH model will lead to biased estimates of the CIFs. In the first part of this dissertation, we propose a simple risk prediction procedure that adopts the PSH model yet relaxes the

PSH assumption. The procedure can yield easy and accurate predictions of CIFs without modeling the potentially time-varying covariate effects.

With the progression of a disease, the patient's prognosis may change as time elapse from the initial diagnosis. Risk prediction is not only needed at baseline but also at later time points during follow-up. For a patient who has not yet experienced any event at a certain time point, clinicians may predict the cumulative incidence for the occurrence of event of interest within the next w-years. To perform dynamic prediction for competing risks data is to dynamically predict the conditional cause-specific CIF based on the patient's disease history and all the information available at the specific time point during follow-up. For data containing no competing risks, van Houwelingen (2007) [31] proposed landmark dynamic prediction models to predict an additional w-years conditional survival probability for a patient who is still alive at certain time points during follow-up. The proposed landmark models can incorporate various types of time-dependent information, including the potential time-varying covariate effects, intermediate clinical events, and longitudinally measured biomarkers simultaneously through a simple prediction model and the implementation is not computationally intensive. In the context of dynamic prediction, the landmark models are more advantageous and straightforward as compared with multi-state models and joint modeling approach [32, 33]. In the second part of this dissertation, we extend the landmark method to data with competing risks and propose landmark proportional sub-distribution hazards model and supermodel for dynamic prediction of the conditional CIFs for data containing competing risks.

2.0 ROBUST PREDICTION OF CUMULATIVE INCIDENCE FUNCTION UNDER NON-PROPORTIONAL SUBDISTRIBUTION HAZARDS

2.1 INTRODUCTION

For data containing competing risk events, the cumulative incidence function (CIF) is a proper summary statistic describing a subject's absolute risk of failure from a specific cause of interest with no assumptions of dependencies between competing events. In clinical applications, prognostic models that predict CIF from a patient's clinical and genomic information collected at the time of diagnosis are very useful to the physicians when performing risk stratifications and making decisions on treatments. For example, with accurate predictions of the CIF of locoreginal recurrence (LRR) oncologists can optimize radiation therapy for the breast cancer patients. The main purpose of our study is to introduce a new approach for predicting a subject-specific CIF directly.

Similar to the Kaplan-Meier estimator, the widely used nonparametric estimator of the CIF has its limitations in that only a few discrete and categorized continuous covariates can be included. In practice, the risk of LRR often depends on a number of histological and clinical factors, such as tumor size, nodal status, surgical margin status, histologic subtype, and vascular invasion. When physicians tailor therapy for personalized oncology interventions, it is required that regression modeling procedures incorporate both discrete and continuous covariates simultaneously. One approach used is to model the cause-specific hazards for all causes separately then combine them to estimate the CIF of the cause of interest. The validity of this approach depends on all cause-specific hazards being modeled correctly which is difficult to achieve in practice.

To date, several regression procedures were used to model the cumulative incidence probabilities directly, including the Fine-Gray model which is also a Cox-like proportional subdistribution hazards model [8], the pseudovalue approach [15], and the direct binomial regression model [26]. The Fine-Gray model has become the most commonly used regression procedure, since it provides a simple explicit form of estimation for the CIF which makes the coefficients of covariates easier to interpret.

However, the proportional subdistribution hazards (PSH) assumption of the Fine-Gray model may not hold for all covariates, especially in studies with long-term follow-up. For example, some studies have shown that the effect of histologic grade shows a diminishing trend on overall survival among patients with breast cancer. Ignoring nonproportionality could introduce bias in prediction of CIF and this could lead to a misleading conclusion. To handle nonproportional hazards, the most straightforward approach is to fit a Fine-Gray model by adding time-covariate interaction terms; however, an additional assumption for the functional form of time is required. Sun et al. (2006) [30] proposed a more flexible and general additive-multiplicative subdistribution hazards model that can be used to estimate fixed covariate effects parametrically and to estimate time-varying covariate effects nonparametrically. Although this model is superior in flexibility and generalizability, the implementation of predicting the CIF is complicated. Alternatively, Zhou et al. (2011) [37] proposed a stratified Fine-Gray model over a set of discrete factors where the PSH assumption is not satisfied. This method categorizes a continuous variable with time-varying effect and it cannot be used to predict a CIF if data is highly stratified. Note that in each of the aforementioned methods a goodness-of-fit test has to be performed in advance to identify covariates that violate the PSH assumption.

For data not containing competing risks, Struthers and Kalbfleisch (1986) [29] and Xu and O'Quigley (2000) [35] showed that the maximum partial likelihood estimator for a Cox proportional hazards model converges to a limiting value which is a weighted average of the true time-varying covariate effect. Van Houwelingen (2007) [31] elucidated that a simple Cox model with additional administrative censoring at a certain time point of interest (the horizon time, t_{hor}) can provide an approximation for the predicted survival probabilities at

t_{hor} if the time-varying effect does not vary much over time, if the effect of covariate is small, and if the length of the follow-up period is limited. This valid prediction model was named the "stopped Cox model", meaning that the model stopped at the horizon time [33].

In this study, we introduced a new approach for predicting a subject-specific CIF directly using a method that robust to the PSH assumption. We extended the stopped Cox model approach to the Fine-Gray model for competing risks data, therefore our model is named "stopped proportional subdistribution hazards model" (or a "stopped PSH model"). When applying our models, a researcher can accurately predict a CIF without constructing complex procedures to estimate time-varying effects when the PSH assumption is violated. In order to eliminate the impact of heavy random censoring before the horizon time t_{hor}, we further modified the stopped PSH model by adding a weight to reduce bias.

In Section 2.2, we review the Fine-Gray PSH model and discuss its potential issues when the PSH assumption is violated. We then present the proposed stopped and weighted stopped PSH models. In Section 2.3, we assess the performance in prediction for the proposed models and compare the performance with existing methods through simulations. In Section 2.4, we apply the proposed models to predict the CIF of LRR given a set of prognostic factors from a breast cancer treatment trial. Discussion is provided in Section 2.5.

2.2 METHODS

2.2.1 Notations and PSH Model

Let T and C be the failure and censoring time, respectively; $\varepsilon \in \{1 \dots k\}$ be the cause of failure; and $\boldsymbol{Z}$ be a p-dimensional vector of time-fixed covariates. Here, we assume C_i is independent of T_i and $\boldsymbol{Z}_i$, and refer to it as random censoring. For right censored data, we observe an independently and identically distributed quadruplet of $\{X_i = T_i \wedge C_i, \Delta_i = I(T_i \leq C_i), \Delta_i \varepsilon_i, \boldsymbol{Z}_i\}$ for subject i $(i = 1 \dots n)$. Subdistribution or CIF for failure from cause 1 is defined as $F_1(t; \boldsymbol{Z}) = \Pr(T \leq t, \varepsilon = 1 | \boldsymbol{Z})$. Our objective is to predict the CIF of a subject given his or her covariate values.

Fine and Gray (1999) [8] proposed a proportional hazards regression model for the subdistribution $F_1(t; \boldsymbol{Z})$. This PSH model takes the form

$$\lambda_1(t; \boldsymbol{Z}) = \lambda_{10}(t) \exp(\boldsymbol{Z}^T \boldsymbol{\beta}),$$

where the baseline subdistribution hazards $\lambda_{10}(t) = -d \log\{1 - F_1(t; \boldsymbol{Z} = \underline{0})\}/dt$ is an unspecified, nonnegative function; and where $\boldsymbol{\beta}$ is a vector of unknown regression parameters. Thus, the cumulative incidence function can be calculated as

$$F_1(t; \boldsymbol{Z}) = 1 - \exp\left\{ - \int_0^t \lambda_{10}(u) \exp(\boldsymbol{Z}^T \boldsymbol{\beta})\, du \right\}.$$

The regression coefficients $\boldsymbol{\beta}$ are estimated through a partial likelihood approach with modified risk sets defined as $R(T_i) = \{j : (T_j \geq T_i) \cup (T_j \leq T_i \cap \varepsilon_j \neq 1)\}$ for the ith individual. $R(T_i)$ includes all individuals who have not failed from the cause of interest by time T_i. When random right censoring is present, the inverse probability of censoring weighting (IPCW) technique [25] is applied to obtain an unbiased partial likelihood estimator $\hat{\boldsymbol{\beta}}_{PL}$ via a weighted score function $\boldsymbol{U}(\boldsymbol{\beta})$ given by

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\infty} \left\{ \boldsymbol{Z}_i - \frac{\sum_j \omega_j(t) Y_j(t) \boldsymbol{Z}_j \exp(\boldsymbol{Z}_j^T \boldsymbol{\beta})}{\sum_j \omega_j(t) Y_j(t) \exp(\boldsymbol{Z}_j^T \boldsymbol{\beta})} \right\} \omega_i(t)\, dN_i(t), \tag{2.1}$$

where $N_i(t) = I(T_i \leq t, \varepsilon_i = 1)$, $Y_i(t) = I(T_i \geq t) + I(T_i < t, \varepsilon_i \neq 1)$ and $\omega_i(t) = I(C_i \geq T_i \wedge t)\hat{G}(t)/\hat{G}(X_i \wedge t)$ in which $\hat{G}(t)$ is the Kaplan-Meier estimator of the censoring survival distribution $G(t) = \Pr(C \geq t)$. The baseline cumulative subdistribution hazards $\Lambda_{10}(t) = \int_0^t \lambda_{10}(u)\, du$ can be estimated using a modified version of the Breslow estimator,

$$\hat{\Lambda}_{10}(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{1}{\frac{1}{n} \sum_j \omega_j(u) Y_j(u) \exp(\boldsymbol{Z}_j^T \hat{\boldsymbol{\beta}}_{PL})} \omega_i(u)\, dN_i(u). \tag{2.2}$$

Therefore, the predicted CIF can be calculated as $\hat{F}_1(t; \boldsymbol{z}) = 1 - \exp\left\{ - \exp(\boldsymbol{z}^T \hat{\boldsymbol{\beta}}_{PL}) \hat{\Lambda}_{10}(t) \right\}$ for an individual with covariates $\boldsymbol{Z} = \boldsymbol{z}$ at time t. It has been shown that $\hat{\boldsymbol{\beta}}_{PL}$ is consistent for $\boldsymbol{\beta}$; $\hat{\Lambda}_{10}(t)$ uniformly converges in probability to the true baseline cumulative subdistribution hazard value $\Lambda_{10}(t)$ on the interval $[0, \tau)$, where $\tau = \sup\{t : \Pr(X \geq t) > 0\}$; and $\hat{F}_1(t; z)$ uniformly converges to $F_1(t; z)$ on the same interval [8, 12].

The PSH assumption would be violated if a subdistribution hazards ratio changes over time. Although it is straightforward to incorporate time-varying coefficients $\boldsymbol{\beta}(t)$ into a PSH model

$$\lambda_1(t; \boldsymbol{Z}) = \lambda_{10}(t) \exp\{\boldsymbol{Z}^T \boldsymbol{\beta}(t)\}, \tag{2.3}$$

we will need to estimate the time-varying functional form $\boldsymbol{\beta}(t)$ in order to predict the CIF. This could make prediction complicated, especially when there are more than one covariate with time-varying effects. By extending the idea of stopped Cox model [33] to data with competing risks, we investigate the prediction performance of the Fine-Gray PSH model with additional administrative censoring at the prediction horizon time t_{hor}, when the PSH assumption is not satisfied.

2.2.2 Stopped PSH Model

Struthers and Kalbfleisch (1986) [29] have investigated the properties of $\hat{\boldsymbol{\beta}}_{PL}$ of a Cox proportional hazards (PH) model when the true model is actually accelerated failure time model. Xu and O'Quigley (2000) [35] showed that under nonproportional hazards, the $\hat{\boldsymbol{\beta}}_{PL}$ is consistent for a weighted average of true time-varying effects $\boldsymbol{\beta}(t)$ over time. We can also derive similar properties of $\hat{\boldsymbol{\beta}}_{PL}$ for the PSH model when the proportionality assumption is violated. In the following sections, we suppose the true model is given by model (2.3). For complete data and censoring-complete data where the censoring time is always observed, the PSH model is inherited from the usual Cox PH model [8]; so that the properties derived from the Cox model can be directly generalized to the PSH model. Thus, we mainly focus on the situation where random right censoring is present.

Define $\boldsymbol{S}^{(r)}(\boldsymbol{\beta}, t) = \frac{1}{n}\sum_{i=1}^{n} \omega_i(t) Y_i(t) \boldsymbol{Z}_i^{\otimes r} \exp(\boldsymbol{Z}_i^T \boldsymbol{\beta})$ and $\boldsymbol{s}^{(r)}(\boldsymbol{\beta}, t) = E\boldsymbol{S}^{(r)}(\boldsymbol{\beta}, t)$, for $r = 0, 1, 2$, where the expectations are taken with respect to the true distribution of $(T, C, \boldsymbol{Z})$. Suppose the Andersen-Gill-type conditions (see Appendix A) hold throughout the paper, we have the following theorem.

THEOREM 1. *Under random right censoring, the partial likelihood estimator* $\hat{\boldsymbol{\beta}}_{PL}$ *from the proportional subdistributional hazards model is a consistent estimator of* $\boldsymbol{\beta}^*$*, where* $\boldsymbol{\beta}^*$ *is the solution to the equation*

$$\int_0^\infty \left\{ \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}(t),t)}{\boldsymbol{s}^{(0)}(\boldsymbol{\beta}(t),t)} - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta},t)}{\boldsymbol{s}^{(0)}(\boldsymbol{\beta},t)} \right\} \boldsymbol{s}^{(0)}(\boldsymbol{\beta}(t),t)\lambda_{10}(t)\,dt = 0. \tag{2.4}$$

As described in Fine and Gray (1999) [8], an improper failure time T^*, defined as $T \times I(\varepsilon = 1) + \infty \times \{1 - I(\varepsilon = 1)\}$, has a distribution function as $F_1(t)$ for $t \leq \infty$ and a point mass at $t = \infty$. The subdistribution hazard $\lambda_1(t)$ is actually the hazard function for T^*. We can treat the PSH model for event time T of cause 1 as the Cox PH model for the improper failure time T^*. Under random censoring, we have $\boldsymbol{S}^{(r)}(\boldsymbol{\beta},t) = \boldsymbol{S}_*^{(r)}(\boldsymbol{\beta},t)$ and $\boldsymbol{s}^{(r)}(\boldsymbol{\beta},t) = \boldsymbol{s}_*^{(r)}(\boldsymbol{\beta},t)$, for $r = 0,1,2$, where $\boldsymbol{S}_*^{(r)}(\boldsymbol{\beta},t) = \frac{1}{n}\sum_{i=1}^n Y_i^*(t)\boldsymbol{Z}_i^{\otimes r}\exp(\boldsymbol{Z}_i^T\boldsymbol{\beta})$, $\boldsymbol{s}_*^{(r)}(\boldsymbol{\beta},t) = E\boldsymbol{S}_*^{(r)}(\boldsymbol{\beta},t)$, $Y^*(t) = I(X^* \geq t)$, and $X^* = T^* \wedge C$. Hence, in terms of the improper failure time T^*, Theorem 1 is identical to Theorem 2.1 in Struthers and Kalbfleisch (1986) [29]. Proof of Theorem 1 can be seen in Struthers and Kalbfleisch (1986) [29].

To get an interpretable form of $\boldsymbol{\beta}^*$, we need to make some transformation for equation (2.4). As discussed in Xu and O'Quigley (2000) [35], $\boldsymbol{S}_*^{(1)}(\boldsymbol{\beta}(t),t)/\boldsymbol{S}_*^{(0)}(\boldsymbol{\beta}(t),t)$ can be thought of as a conditional expectation of $\boldsymbol{Z}$ taken with respect to the empirical distribution $\frac{Y_i^*(t)\exp\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(t)\}}{\sum_{j=1}^n Y_j^*(t)\exp\{\boldsymbol{Z}_j^T\boldsymbol{\beta}(t)\}}$. Under the Andersen-Gill-type conditions (see Appendix A), $\boldsymbol{S}_*^{(1)}(\boldsymbol{\beta}(t),t)/\boldsymbol{S}_*^{(0)}(\boldsymbol{\beta}(t),t)$ converges in probability to $\boldsymbol{s}_*^{(1)}(\boldsymbol{\beta}(t),t)/\boldsymbol{s}_*^{(0)}(\boldsymbol{\beta}(t),t)$. Then, we have

$$\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}(t),t)}{\boldsymbol{s}^{(0)}(\boldsymbol{\beta}(t),t)} = \frac{\boldsymbol{s}_*^{(1)}(\boldsymbol{\beta}(t),t)}{\boldsymbol{s}_*^{(0)}(\boldsymbol{\beta}(t),t)} = E(\boldsymbol{Z}|T^* = t), \tag{2.5}$$

$$\frac{\partial}{\partial\boldsymbol{\beta}}\left(\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta},t)}{\boldsymbol{s}^{(0)}(\boldsymbol{\beta},t)}\right)\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}(t)} = \frac{\partial}{\partial\boldsymbol{\beta}}\left(\frac{\boldsymbol{s}_*^{(1)}(\boldsymbol{\beta},t)}{\boldsymbol{s}_*^{(0)}(\boldsymbol{\beta},t)}\right)\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}(t)} = \mathrm{Var}(\boldsymbol{Z}|T^* = t). \tag{2.6}$$

Hence,

$$\frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}(t),t)}{\boldsymbol{s}^{(0)}(\boldsymbol{\beta}(t),t)} - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta},t)}{\boldsymbol{s}^{(0)}(\boldsymbol{\beta},t)} \approx \{\mathrm{Var}(\boldsymbol{Z}|T^* = t)\}^T\{\boldsymbol{\beta} - \boldsymbol{\beta}(t)\}. \tag{2.7}$$

Under random censoring,

$$\begin{aligned} \boldsymbol{s}^{(0)}(\boldsymbol{\beta}(t), t)\lambda_{10}(t) &= E[\omega(t)Y(t)\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}(t)\}]\lambda_{10}(t) \\ &= E[\omega(t)Y(t)\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}(t)\}\lambda_{10}(t)] \\ &= E\{\omega(t)Y(t)\}E\{\lambda_1(t|\boldsymbol{Z})|T \geq t\} \\ &= \{1 - F_1(t)\}G(t)\lambda_1(t), \end{aligned} \tag{2.8}$$

where $\lambda_1(t)$ is the marginal subdistribution hazard and $F_1(t)$ is the marginal subdistribution. Thus, the equation (2.4) is approximately given by

$$\int_0^\infty \{\text{Var}(\boldsymbol{Z}|T^* = t)\}^T\{\boldsymbol{\beta} - \boldsymbol{\beta}(t)\}\{1 - F_1(t)\}G(t)\lambda_1(t)\,dt = 0. \tag{2.9}$$

When we add additional administrative censoring at horizon time t_{hor} and take the first-order Taylor series approximation to the integrand of (2.9), the $\boldsymbol{\beta}^*$ becomes

$$\boldsymbol{\beta}^*_{hor} \approx \frac{\int_0^{t_{hor}}\{1 - F_1(t)\}G(t)\lambda_1(t)\{\text{Var}(\boldsymbol{Z}|T^* = t)\}^T\boldsymbol{\beta}(t)\,dt}{\int_0^{t_{hor}}\{1 - F_1(t)\}G(t)\lambda_1(t)\text{Var}(\boldsymbol{Z}|T^* = t)\,dt}, \tag{2.10}$$

which is similar to the approximation of $\boldsymbol{\beta}^*$ in Cox model shown by van Houwelingen (2007) [31]. If the cumulative incidence function $F_1(t)$ does not get too large; the censoring proportion is not too high before t_{hor} ($G(t_{hor}) \approx 1$); and the effect of a covariate is not too large and does not vary too much over time, $\text{Var}(\boldsymbol{Z}|T^* = t)$ could be approximated by a constant. Therefore, we have

$$\boldsymbol{\beta}^*_{hor} \approx \frac{\int_0^{t_{hor}}\lambda_{10}(t)\boldsymbol{\beta}(t)\,dt}{\int_0^{t_{hor}}\lambda_{10}(t)\,dt}, \tag{2.11}$$

which is a weighted average of $\boldsymbol{\beta}(t)$ over time. Approximation (2.11) is similar to formula (3.2) given in Xu and O'Quigley (2000) [35].

After applying the PSH model to the data up to t_{hor}, the cumulative subdistribution hazard can be estimated by

$$\hat{\Lambda}_1(t_{hor}; \boldsymbol{Z}) = \exp(\boldsymbol{Z}^T\hat{\boldsymbol{\beta}}_{PL})\hat{\Lambda}_{10}(t_{hor}),$$

where $\hat{\Lambda}_{10}(t_{hor})$ is given in (2.2). Following the derivations of the approximated limiting values of the Breslow estimated hazards for the Cox model [31], the Breslow-type subdistribution baseline hazard estimate $\hat{\lambda}_{10}(t) = d\hat{\Lambda}_{10}(t)/dt$ converges in probability to $\lambda^*_{10}(t)$, which is approximated by

$$\lambda^*_{10}(t) \approx \lambda_{10}(t) \exp[E(\boldsymbol{Z}|T^* = t)^T \{\boldsymbol{\beta}(t) - \boldsymbol{\beta}^*_{hor}\}]. \tag{2.12}$$

The derivation of the approximation (2.12) is given in the Appendix B.

The corresponding limiting value of $\hat{\Lambda}_1(t_{hor}; \boldsymbol{Z})$ is

$$\Lambda^*_1(t_{hor}; \boldsymbol{Z}) = \exp(\boldsymbol{Z}^T \boldsymbol{\beta}^*_{hor}) \int_0^{t_{hor}} \lambda^*_{10}(t)\, dt,$$

so that $\Lambda^*_1(t_{hor}; \boldsymbol{Z})$ has an approximation

$$\Lambda^*_1(t_{hor}; \boldsymbol{Z}) \approx \exp(\boldsymbol{Z}^T \boldsymbol{\beta}^*_{hor}) \int_0^{t_{hor}} \lambda_{10}(t) \exp[E(\boldsymbol{Z}|T^* = t)^T \{\boldsymbol{\beta}(t) - \boldsymbol{\beta}^*_{hor}\}]\, dt. \tag{2.13}$$

The true value $\Lambda_1(t_{hor}; \boldsymbol{Z}) = \int_0^{t_{hor}} \lambda_{10}(t) \exp\{\boldsymbol{Z}^T \boldsymbol{\beta}(t)\}\, dt$ can be written as the form

$$\Lambda_1(t_{hor}; \boldsymbol{Z}) = \exp(\boldsymbol{Z}^T \boldsymbol{\beta}^*_{hor}) \int_0^{t_{hor}} \lambda_{10}(t) \exp[\boldsymbol{Z}^T \{\boldsymbol{\beta}(t) - \boldsymbol{\beta}^*_{hor}\}]\, dt. \tag{2.14}$$

Under the conditions for yielding the approximation (2.11), the conditional distribution of $\boldsymbol{Z}$ given $T^* = t$ can be seen as stable over time, and hence, we have the following relationships

$$\hat{\Lambda}_1(t_{hor}; \boldsymbol{Z}) \xrightarrow{p} \Lambda^*_1(t_{hor}; \boldsymbol{Z}) \approx \Lambda_1(t_{hor}; \boldsymbol{Z}). \tag{2.15}$$

Therefore, even though the proportional hazards assumption does not hold, the PSH model with administrative censoring at t_{hor} provides approximately correct prediction of $F_1(t; \boldsymbol{Z})$ at $t = t_{hor}$ over the interval $[0, t_{hor}]$, if $1 - F_1(t)$ and $G(t)$ are close to 1 and the true covariate effects $\boldsymbol{\beta}(t)$ or the subdistribution hazards ratio does not vary too much over time. We name such a PSH model as the stopped PSH model, indicating that the PSH model stopped at the horizon time t_{hor}.

So far, we have shown that under non-proportional hazards, the stopped PSH model performs well on predicting the CIF. We also investigated its properties when the PSH assumption holds. For complete data, the stopped PSH model at t_{hor} is exactly the PSH model for data with administrative censoring, so that the model properties have been shown

in Section 3.2 of Fine and Gray (1999) [8]. For incomplete data with random right censoring, the stopped PSH model also applies the IPCW approach to adjust for the censoring effect before t_{hor}. When $t \leq t_{hor}$, the weight function $\omega(t)$ in stopped PSH model is the same as it calculated via the Fine-Gray PSH model. To predict the $\hat{\Lambda}_1(t; \boldsymbol{Z})$ at time t_{hor},

$$\hat{\Lambda}_1(t_{hor}; \boldsymbol{Z}) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{t_{hor}} \frac{\exp(\boldsymbol{Z}^T \hat{\boldsymbol{\beta}}_{PL})}{\frac{1}{n} \sum_j \omega_j(u) Y_j(u) \exp(\boldsymbol{Z}_j^T \hat{\boldsymbol{\beta}}_{PL})} \omega_i(u)\, dN_i(u), \tag{2.16}$$

we only need the information up to t_{hor}, because for subjects $\{j : X_j \geq t_{hor}\}$, the risk set at t_{hor} will not change regardless of whether they were censored, had competing events, or were still alive. Hence, under proportional hazards, according to the properties of the PSH model given in Fine and Gray (1999) [8], the stopped PSH model also gives reliable prediction of $F_1(t_{hor}; \boldsymbol{Z})$ if t_{hor} is less than $\tau = \sup\{t : \Pr(X \geq t) > 0\}$. However, as shown in equation (2.16), the $\hat{\Lambda}_1(t_{hor}; \boldsymbol{Z})$ depends on the partial likelihood estimator $\hat{\boldsymbol{\beta}}_{PL}$; so that the $\hat{F}_1(t_{hor}; \boldsymbol{Z})$ from the stopped PSH model may be less efficient compared to that from the Fine-Gray PSH model, since less events are used when we calculate the $\hat{\boldsymbol{\beta}}_{PL}$ in the stopped PSH model.

2.2.3 Weighted Stopped PSH Model

As discussed in Section 2.2.2, the limiting value $\boldsymbol{\beta}^*_{hor}$ of the partial likelihood estimator $\hat{\boldsymbol{\beta}}_{PL}$ in the stopped PSH model depends on the unknown censoring distribution $G(t)$ through $\boldsymbol{s}^{(0)}(\boldsymbol{\beta}(t), t)$. If there is a heavy censoring before t_{hor}, the approximation of $\boldsymbol{\beta}^*_{hor}$ shown in (2.11) will not be valid, and the limiting value of the $\hat{\Lambda}_1(t_{hor}; \boldsymbol{Z})$ will not be approximately equal to the true value. Thus, $\hat{F}_1(t_{hor}; \boldsymbol{Z})$ will lose accuracy when t_{hor} is far away from the origin. Therefore, the stopped PSH model needs to be modified in order to correct the heavy random censoring before t_{hor}.

In the Cox model, Xu and O'Quigley (2000) [35] proposed a weighted partial likelihood estimator whose limiting value is independent of censoring. We generalize their method to the PSH model and derive a reweighted score equation as the form

$$\boldsymbol{U}_w(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\infty} W(t) \left\{ \boldsymbol{Z}_i - \frac{\sum_j \omega_j(t) Y_j(t) \boldsymbol{Z}_j \exp(\boldsymbol{Z}_j^T \boldsymbol{\beta})}{\sum_j \omega_j(t) Y_j(t) \exp(\boldsymbol{Z}_j^T \boldsymbol{\beta})} \right\} \omega_i(t)\, dN_i(t), \tag{2.17}$$

where $W(t) = \hat{S}(t)/\sum_{i=1}^{n} Y_i'(t)$, $\hat{S}(t)$ is the left continuous version of the Kaplan-Meier estimator of the overall survival $S(t)$ and $Y_i'(t) = I(X_i \geq t)$. The solution to equation (2.17) is denoted as $\hat{\boldsymbol{\beta}}_w$. If there is no ties, $W(X_i)$ can be viewed as the increment of the nonparametric estimate of the marginal cumulative incidence function $F_1(t)$ at an observed main event time X_i. It can be verified that $W(t)G(t)$ is approximately equal to a constant c. Under random censoring, now we have

$$\begin{aligned} \boldsymbol{s}^{(0)}(\boldsymbol{\beta}(t), t)\lambda_{10}(t) &= E[W(t)\omega(t)Y(t)\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}(t)\}\lambda_{10}(t)] \\ &= c\{1 - F_1(t)\}\lambda_1(t). \end{aligned} \tag{2.18}$$

Thus, the influence of censoring has been removed from $\boldsymbol{s}^{(0)}(\boldsymbol{\beta}(t), t)$. Following Theorem 3.2 in Lin (1991) [16] and Theorem 2 in Xu and O'Quigley (2000) [35], we have

THEOREM 2. *Under non-proportional PSH model (2.3), the weighted estimator $\hat{\boldsymbol{\beta}}_w$ converges in probability to $\tilde{\boldsymbol{\beta}}$, where $\tilde{\boldsymbol{\beta}}$ is the unique solution to the equation*

$$\int_0^\infty \left\{ \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}(t), t)}{\boldsymbol{s}^{(0)}(\boldsymbol{\beta}(t), t)} - \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)}{\boldsymbol{s}^{(0)}(\boldsymbol{\beta}, t)} \right\} \{1 - F_1(t)\}\lambda_1(t)\, dt = 0, \tag{2.19}$$

if $\int_0^\infty \upsilon(\tilde{\boldsymbol{\beta}}, t)\{1 - F_1(t)\}\lambda_1(t)\, dt > 0$, where $\upsilon(\tilde{\boldsymbol{\beta}}, t) = \frac{\boldsymbol{s}^{(2)}(\tilde{\boldsymbol{\beta}}, t)}{\boldsymbol{s}^{(0)}(\tilde{\boldsymbol{\beta}}, t)} - \left\{ \frac{\boldsymbol{s}^{(1)}(\tilde{\boldsymbol{\beta}}, t)}{\boldsymbol{s}^{(0)}(\tilde{\boldsymbol{\beta}}, t)} \right\}^{\otimes 2}$

If we rewrite equation (2.19) in terms of T^*, the proof of Theorem 2 could follow the arguments in Xu (1996) [34]. It is clear that the censoring distribution $G(t)$ is not involved in equation (2.19); therefore, $G(t)$ is not involved in the form of solution $\tilde{\boldsymbol{\beta}}$.

Similarly, in the presence of administrative censoring at t_{hor}, applying (2.7) to equation (2.19), then the first-order Taylor series approximation of $\tilde{\boldsymbol{\beta}}$ becomes

$$\tilde{\boldsymbol{\beta}}_{hor} \approx \frac{\int_0^{t_{hor}} \{1 - F_1(t)\}\lambda_1(t)\{\mathrm{Var}(\boldsymbol{Z}|T^* = t)\}^T \boldsymbol{\beta}(t)\, dt}{\int_0^{t_{hor}} \{1 - F_1(t)\}\lambda_1(t)\mathrm{Var}(\boldsymbol{Z}|T^* = t)\, dt}. \tag{2.20}$$

Following the same derivation in Section 2.2.2 , it can be easily shown that the limiting value $\tilde{\Lambda}_1(t_{hor}; \boldsymbol{Z})$ of $\hat{\Lambda}_1(t_{hor}; \boldsymbol{Z})$ calculated by $\hat{\boldsymbol{\beta}}_w$ is approximately equal to the true value $\Lambda_1(t_{hor}; \boldsymbol{Z})$ if the variation of $\beta(t)$ over time is not too large and $1 - F_1(t_{hor})$ is not too far away from 1, where $\tilde{\Lambda}_1(t_{hor}; \boldsymbol{Z})$ takes the same approximation form as $\Lambda_1^*(t_{hor}; \boldsymbol{Z})$ does, which was shown in (2.13), only replacing $\boldsymbol{\beta}^*$ with $\tilde{\boldsymbol{\beta}}$.

Therefore, compared to the stopped PSH model, the weighted stopped PSH model not only provides an accurate predictive cumulative incidence at t_{hor} but also remedies the impact of the heavy censoring before t_{hor}.

2.2.4 Measure of Predictive Accuracy

To evaluate the predictive accuracy of the proposed procedures, we adapted the Brier score which is an estimate of the mean-squared prediction errors of the predicted event probabilities over the observed event status [11, 27]. For competing risks data, the expected Brier score at horizon time is defined as

$$\mathbf{BS}_{t_{hor}} = E\{I(T \leq t_{hor}, \varepsilon = 1) - F_1(t_{hor}; \boldsymbol{Z})\}^2.$$

In the presence of random censoring, a consistent estimator of $\mathbf{BS}_{t_{hor}}$ was proposed by Schoop et al. (2011) [27], which was applied the IPCW approach [24],

$$\widehat{\mathbf{BS}}_{t_{hor}} = \frac{1}{n}\sum_{i=1}^{n}\{I(X_i \leq t_{hor}, \Delta_i\varepsilon_i = 1) - \hat{F}_1(t_{hor}; \boldsymbol{Z}_i)\}^2 w_b(X_i),$$

for all times $t_{hor} \leq \sup\{t : G(t) > 0\}$, where

$$w_b(X_i) = \frac{I(X_i \leq t_{hor}, \Delta_i\varepsilon_i \neq 0)}{\hat{G}(X_i-)} + \frac{I(X_i > t_{hor})}{\hat{G}(t_{hor})}.$$

To quantify the improvement of predictive accuracy for the stopped PSH model compared to the Fine-Gray PSH model under nonproportional hazards, we utilized a relative increment of prediction errors (RIPE) by making the nonparametric estimates as a reference, i.e.,

$$\mathbf{RIPE}_{\text{S-PSH}} = \widehat{\mathbf{BS}}^{\text{S-PSH}}_{t_{hor}} / \widehat{\mathbf{BS}}^{\text{NP}}_{t_{hor}} - 1.$$

In addition, we also measure the global prediction errors by computing the integrated Brier scores,

$$\mathbf{IBS}_{t_{hor}} = \int_0^{t_{hor}} \mathbf{BS}_t \, dt,$$

which is estimated by using the composite trapezoidal rule [27, 33].

2.3 SIMULATION STUDIES

We evaluated the performance in prediction for the proposed procedures with simulated data under both proportional and nonproportional subdistribution hazards. We compared the Brier scores of the proposed methods to the nonparametric estimates and the PSH model. For simplicity, only two failure types were considered. Type 1 failure was the main event of interest, and type 2 failure indicated competing event.

For data under PSH, we generated the type 1 failure times from the subdistribution function

$$F_{1,PSH}(t;Z_i) = 1-(1-p[1-\exp\{-(\lambda t)^{\alpha}\}])^{\exp(Z_i\beta)},$$

which is a two-parameter Weibull mixture distribution. Z_i is a dichotomous covariate from Bernoulli(0.5). We let $(\alpha,\lambda,\beta) = (2.2, 0.18, -1)$. For data under non-PSH, we created two different sets of simulations. In the first set we let the rate parameter λ be dependent on the covariate Z_i; and the subdistribution of the main event became

$$F_{1,1^{st}nonPSH}(t;Z_i) = p(1-\exp[-\{\lambda_1\exp(Z_i\beta_1)t\}^{\alpha_1}]).$$

In the second set we generated data based on

$$F_{1,2^{nd}nonPSH}(t;Z_i) = 1-(1-p[1-\exp\{-(\lambda_2 t)^{\alpha_2}\}])^{\exp(Z_i\beta_{21}+Z_i\beta_{22}t)},$$

where the coefficient of Z_i is a linear function of time. Under both non-PSH settings, we considered Z_i as Bernoulli(0.5) variates and let $(\alpha_1,\lambda_1,\beta_1,\alpha_2,\lambda_2,\beta_{21},\beta_{22}) = (3.2, 0.18, -0.8, 2.2, 0.12, 0.5, 0.01)$. For the second non-PSH set, we also considered Z_i as a continuous variable from Normal$(0.5, 0.01)$ and let $(\alpha_2,\lambda_2,\beta_{21},\beta_{22}) = (1.2, 0.12, 0.5, 0.01)$. For all simulations, we let $p = 0.3$, which produced about 30% main events at $Z_i = 0$ when there was no censoring. We generated the type 2 failure times from an exponential distribution $\Pr(T_i \le t|\varepsilon_i = 2, Z_i) = 1-\exp\{-\exp(Z_i\beta_c t)\}$ by taking $\Pr(\varepsilon_i = 2|Z_i) = 1-\Pr(\varepsilon_i = 1|Z_i)$, where $\beta_c = 0.5$. Sample size of $n = 500$ was chosen and the data were simulated repeatedly; $N = 1000$ times. The censoring times were generated independently from a uniform distribution which resulted in about 20% censoring.

In the first non-PSH set, we compared the performances in prediction for CIFs obtained from the nonparametric method, the PSH model, the proposed stopped PSH model, and the proposed weighted stopped PSH model. In the second non-PSH set, we also fitted a PSH model with time-varying effects by adding an interaction term of covariate Z_i and a linear function of time. The resulted CIFs were compared with that from our proposed models. Figure 1 depicts the true and the estimated CIFs under different approaches. By comparing the performances between models, we found that the stopped PSH model was as good as the nonparametric method under both PSH and non-PSH scenarios in the absence of censoring, and that the stopped PSH model shows slight bias at the late follow-up years when there was a 20% censoring. The results showed that the bias was reduced if a weighted stopped PSH model was used. Note that the differences in the estimated CIFs obtained from different models diminish over time. This is because at the end of a long follow-up, the number of events is very small and not enough to capture the differences.

We also evaluated the prediction errors for each approach by estimating the Brier scores and used 3-fold cross-validation to correct for possible overfitting. Figure 2 shows the relative increments of prediction errors with corresponding empirical standard deviation at horizon time from 0 to 10 years. In the case of discrete covariate, we chose the nonparametric estimate as a reference. In the case of continuous covariate, we chose the estimate from the PSH model with time-covariate interactions as a reference. To examine the global prediction accuracy over short-term and long-term periods, we calculated the cross-validated integrated Brier score (with empirical standard deviation and the 95% confidence interval) at 5 and 10 years, respectively. The results are shown in Table 1. As expected, in both non-PSH sets the predictive accuracy of the stopped PSH model was almost the same as that obtained from the nonparametric method. Compare to the stopped PSH model, the weighted stopped PSH model reduced the prediction errors further in the presence of censoring. In the second non-PSH set, there were only minor differences in prediction errors among the stopped PSH model, the weighted stopped PSH model, and the PSH model with time-covariate interactions.

Table 1: Cross-validated (3-fold) estimates for the integrated Brier scores (IBS) at $t_{hor} = 5$ and 10, the corresponding empirical standard deviation (SD) and empirical 95% confidence interval.

(a) 1st non-PSH setting: $Z \sim$ Bernoulli (0.5)

		$t_{hor} = 5$			$t_{hor} = 10$		
Censoring		Ave.	SD	95% C.I.	Ave.	SD	95% C.I.
0	IBS_{NP}	10.744	1.795	–	72.350	5.521	–
	IBS_{PSH}	11.295	1.925	–	80.143	6.186	–
	$IBS_{S\text{-}PSH}$	10.743	1.795	–	72.368	5.527	–
	$RIBS_{PSH/NP}$	5.095	2.148	(4.961, 5.228)	10.807	3.112	(10.614, 11.000)
	$RIBS_{S\text{-}PSH/NP}$	-0.008	0.022	(-0.009, -0.007)	0.025	0.120	(0.017, 0.032)
20%	IBS_{NP}	10.784	1.965	–	72.459	6.341	–
	IBS_{PSH}	10.912	2.001	–	73.962	6.409	–
	$IBS_{S\text{-}PSH}$	10.785	1.965	–	72.595	6.385	–
	$IBS_{WS\text{-}PSH}$	10.781	1.964	–	72.492	6.353	–
	$RIBS_{PSH/NP}$	1.180	1.633	(1.079, 1.282)	2.098	1.827	(1.985, 2.212)
	$RIBS_{S\text{-}PSH/NP}$	0.003	0.185	(-0.008, 0.015)	0.186	0.633	(0.145, 0.225)
	$RIBS_{WS\text{-}PSH/NP}$	-0.032	0.192	(-0.044, -0.020)	0.046	0.517	(0.014, 0.078)

(b) 2nd non-PSH setting: $Z \sim$ Bernoulli (0.5)

		$t_{hor} = 5$			$t_{hor} = 10$		
Censoring		Ave.	SD	95% C.I.	Ave.	SD	95% C.I.
0	IBS_{NP}	17.532	2.585	–	98.202	6.703	–
	$IBS_{PSH.f(t)}$	17.577	2.590	–	98.380	6.724	–
	$IBS_{S\text{-}PSH}$	17.531	2.585	–	98.198	6.703	–
	$RIBS_{PSH.f(t)/NP}$	0.261	0.747	(0.215, 0.307)	0.180	0.263	(0.164, 0.197)
	$RIBS_{S\text{-}PSH/NP}$	-0.005	0.021	(-0.007, -0.004)	-0.004	0.062	(-0.008, 0.000)
20%	IBS_{NP}	17.548	2.785	–	98.505	7.602	–
	$IBS_{PSH.f(t)}$	17.516	2.777	–	98.548	7.614	–
	$IBS_{S\text{-}PSH}$	17.546	2.786	–	98.554	7.623	–
	$IBS_{WS\text{-}PSH}$	17.537	2.782	–	98.488	7.606	–
	$RIBS_{PSH.f(t)/NP}$	-0.179	0.505	(-0.210, -0.147)	0.044	0.596	(0.007, 0.081)
	$RIBS_{S\text{-}PSH/NP}$	-0.011	0.147	(-0.020, -0.002)	0.049	0.500	(0.018, 0.080)
	$RIBS_{WS\text{-}PSH/NP}$	-0.064	0.146	(-0.073, -0.055)	-0.017	0.461	(-0.046, 0.011)

Table 1 continued

(c) 2nd non-PSH setting: $Z \sim$ Normal (0.5, 0.01)

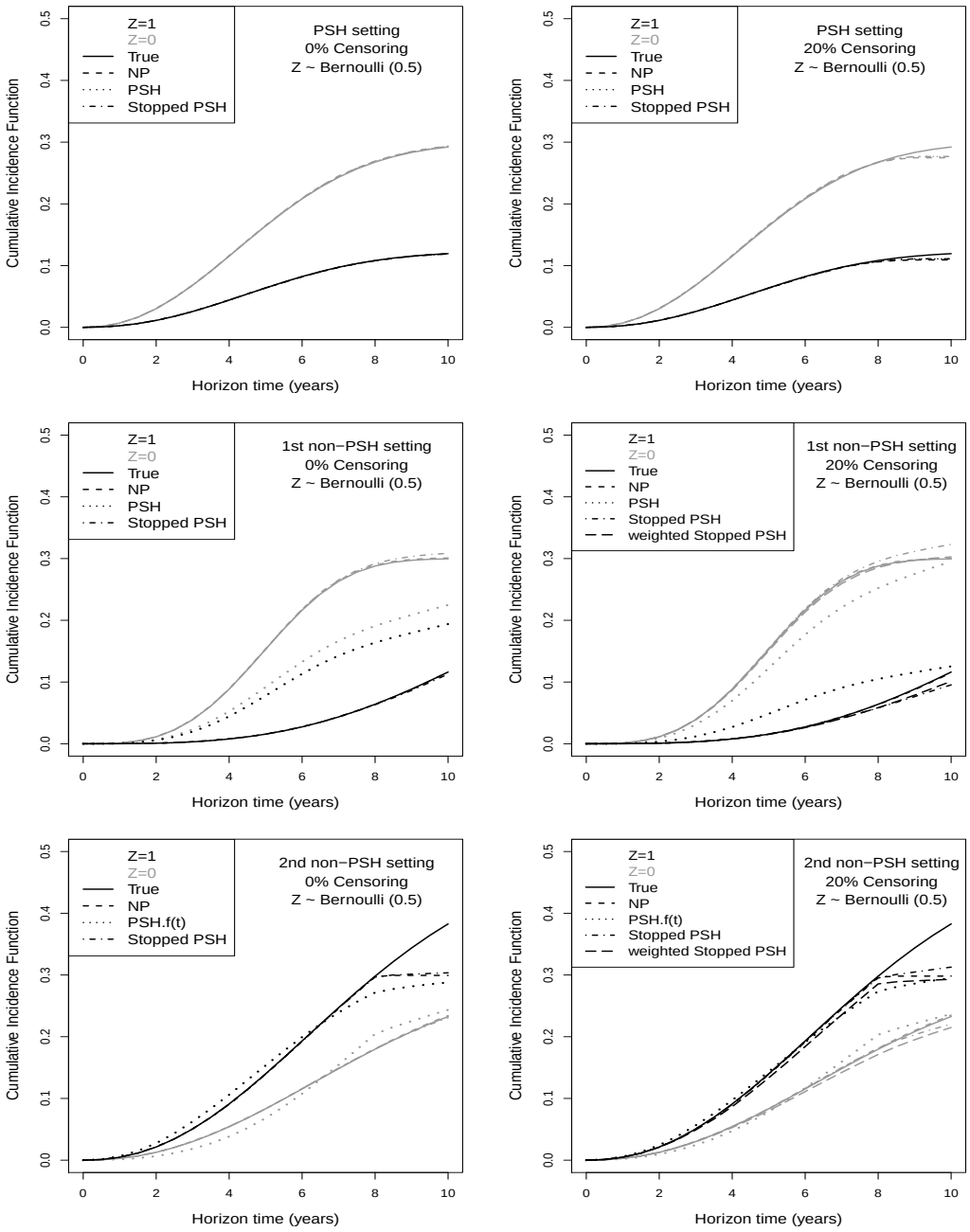
Censoring		$t_{hor} = 5$			$t_{hor} = 10$		
		Ave.	SD	95% C.I.	Ave.	SD	95% C.I.
0	$IBS_{PSH.f(t)}$	36.202	3.944	–	123.236	8.289	–
	$IBS_{S\text{-}PSH}$	36.238	3.947	–	123.301	8.294	–
	$RIBS_{S\text{-}PSH/PSH.f(t)}$	0.099	0.153	(0.089, 0.108)	0.053	0.059	(0.049, 0.057)
20%	$IBS_{PSH.f(t)}$	36.226	4.160	–	123.310	9.032	–
	$IBS_{S\text{-}PSH}$	36.259	4.163	–	123.352	9.031	–
	$IBS_{WS\text{-}PSH}$	36.250	4.162	–	123.352	9.046	–
	$RIBS_{S\text{-}PSH/PSH.f(t)}$	0.092	0.125	(0.084, 0.099)	0.035	0.150	(0.026, 0.044)
	$RIBS_{WS\text{-}PSH/PSH.f(t)}$	0.067	0.128	(0.059, 0.075)	0.074	0.151	(0.065, 0.084)

NP: the nonparametric estimates; PSH: the PSH model; PSH.f(t): the PSH model with time-covariate interactions; S-PSH: the stopped PSH model; WS-PSH: the weighted stopped PSH model. RIBS is the relative increment of integrated Brier scores, i.e., $RIBS_{PSH/NP} = IBS_{PSH}/IBS_{NP}$ - 1. All entries are multiplyed by 100.

2.4 APPLICATION

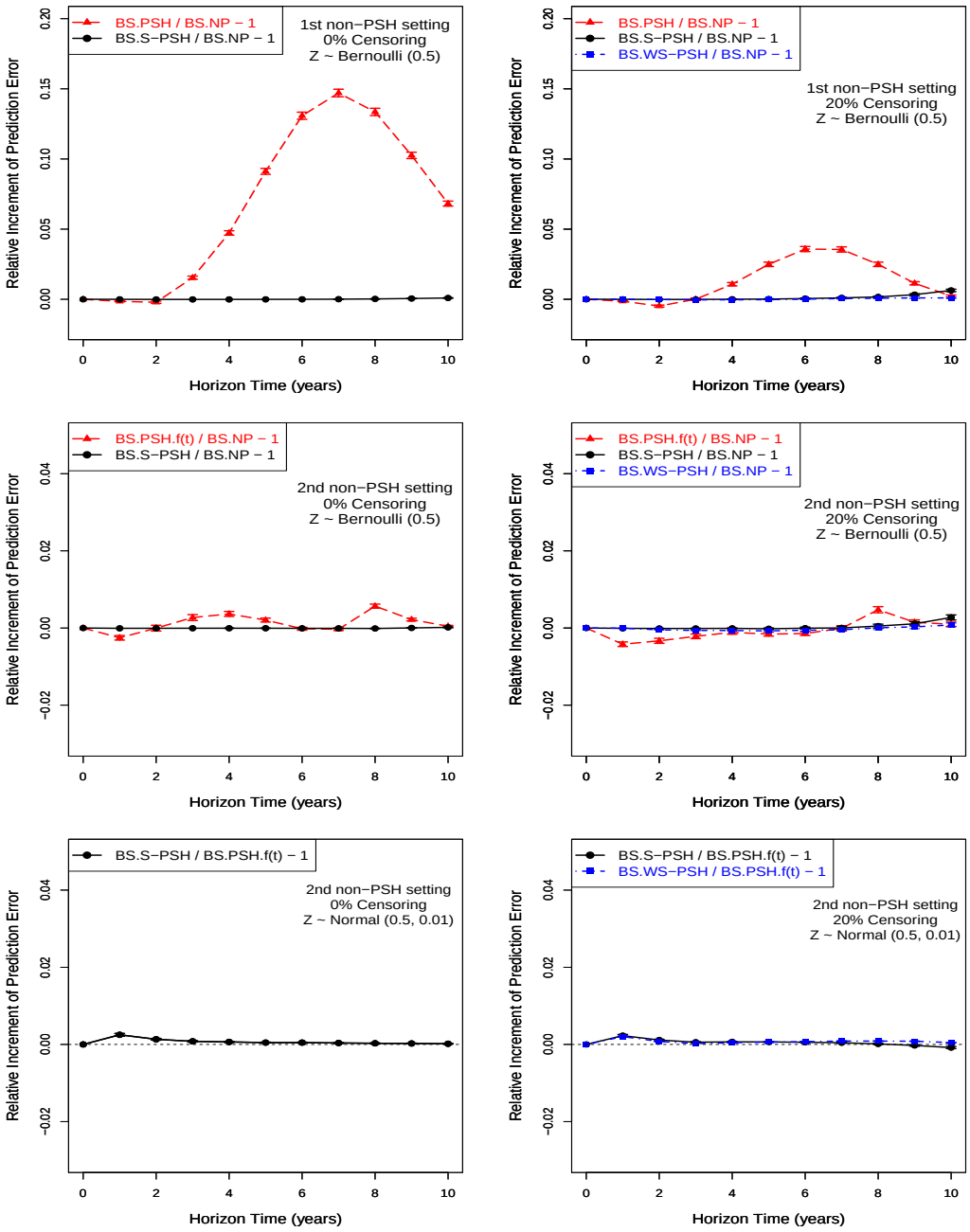
To illustrate the use of our proposed methods in predicting subject-specific CIF, we used data from the NSABP B-14 trial which was a multicenter phase III clinical trial for women with estrogen receptor positive and historically nodes-negative primary breast cancer [10]. In this trial, 2,892 patients were randomly assigned to receive 5 years of tamoxifen or placebo after surgeries. Among the 2,767 clinically eligible patients whom were followed up, 286 developed locoreginal recurrence (LRR), 1,155 had other events before LRR including distant recurrence, second primary cancers, and death; the remaining 1,326 were censored. The median follow-up time was 6.4 years. Our main interest in this application is to predict the cumulative probabilities of LRR given a set of prognostic factors including treatment (tamoxifen vs. placebo), surgery type (lumpectomy plus radiation therapy [L+XRT] vs. mastectomy), age at the time of surgery (< 50 vs. ≥ 50 years old), and tumor size ranging from 0 to 9.8 cm with a median of 2 cm.

We began our analysis by examining the PSH assumption for each prognostic factor from the plot of Shoenfeld-type residuals versus time. Figure 3 shows that all prognostic factors of interest violated the PSH assumption significantly. We also assessed the PSH assumption



Black curves: the $Z = 1$ group; gray curves: the $Z = 0$ group. True: underlying true CIFs. NP: nonparametric method; PSH: the PSH model; PSH.f(t): the PSH model with time-covariate interactions.

Figure 1: Predicted cumulative incidence functions (averaged over 1000 simulations) at a set of horizon times.



The prediction errors were cross-validated (3-fold) estimates for the Brier scores (BS). NP: nonparametric method; PSH: the PSH model; PSH.f(t): the PSH model with time-covariate interactions; S-PSH: the stopped PSH model; WS-PSH: the weighted stopped PSH model.

Figure 2: Relative increment of prediction errors (and their standard deviation) at horizon time from 0 to 10.

using the numeric goodness-of-fit test proposed by Zhou et al. (2013) [36]. The testing results suggested significant linear covariate effects for all prognostic factors and significant quadratic time-varying effects for treatment, surgery type, and tumor size (Table 2). By adding linear and quadratic time interaction terms to the Fine-Gray PSH model we substantially improved the model fit. This is depicted by the Shoenfeld-type residual plots in Figure 3.

We estimated the cumulative probabilities of LRR over time for patients with different characteristics from four different models: the PSH model, the PSH model with time-covariate interactions, and our two proposed models (stopped PSH and weighted stopped PSH models). For demonstration, we selected two subgroups of patients with different sets of characteristics: the first subgroup included younger breast cancer patients (< 50 years old at the time of surgery) with 3 cm tumor size, received L+XRT, and treated with placebo; the second subgroup included older patients (≥ 50 years old at the time of surgery) with 1 cm tumor size, received mastectomy, and treated with tamoxifen. Figure 4 depicts the estimated predicted cumulative probabilities of LRR for the two subgroups of patients obtained from the aforementioned four different modeling approaches. The results show that the estimates from our proposed models (both unweighted and weighted stopped PSH models) agree with that obtained from the PSH model with time-covariate interactions but disagree with that obtained from the simple PSH model. From the simple PSH model, no substantial differences were found in the probabilities of developing LRR between the two subgroups.

We also compared the predictive capability of the four models by estimating the Brier scores at 1, 3, 5, 7, and 9 years (Table 3). In order to quantify how much the proposed models gained in terms of prediction accuracy, we chose the PSH model as the reference model and assessed the relative increment (or reduction) of prediction errors for other three regression models in comparison with the reference model (Table 3). The results show that our proposed stopped PSH model surpasses PSH model in prediction accuracy and it also surpasses the PSH model with time-covariate interactions. The weighted stopped PSH model also yielded smaller prediction errors, and the stopped PSH model had the smallest prediction errors.

One possible reason for the weighted stopped PSH model not performing as well as the stopped PSH model with this particular data is that there was only less than 10% of the samples censored before year 10, which may lead to biased estimates of the weights.

To account for the possible overfitting, we also computed a 3-fold cross-validated Brier scores and the standard errors (SE), where the SE estimates were obtained from estimating sampling distribution of the Brier scores by resampling the original dataset of size 2,767 for $B = 500$ times. Because that the number of main events (developing LRR) was too small, our proposed models and the PSH model with time-covariate interactions only showed minor improvements in prediction accuracy and became less efficient in comparison with the simple PSH model. An alternative strategy is to validate the performance of our models with an external dataset that is independent of the dataset used for fitting the model.

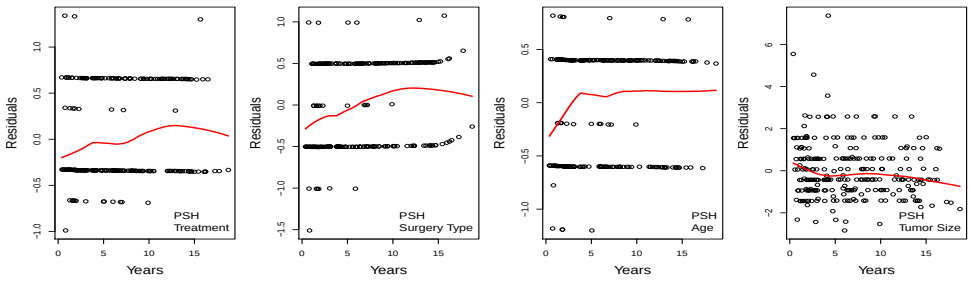
Table 2: Goodness-of-fit test for the PSH model.

Covariates	t	t^2	$t + t^2$
Treatment	0.005	0.011	0.020
Surgery type	$< .001$	$< .001$	$< .001$
Age	0.021	0.058	0.033
Tumor size	0.017	0.020	0.066

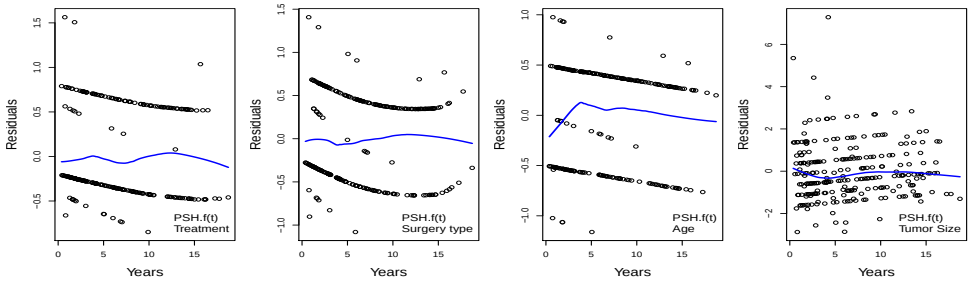
Table 3: Estimated Brier scores at various prediction horizon times (t_{hor}= 1, 3, 5, 7, and 9 years).

	t_{hor}				
	1	3	5	7	9
BS_{PSH}	0.611	3.376	4.691	5.880	6.751
$BS_{PSH.f(t)}$	0.607	3.368	4.675	5.879	6.756
$BS_{S\text{-}PSH}$	0.603	3.368	4.671	5.875	6.750
$BS_{WS\text{-}PSH}$	0.603	3.369	4.672	5.877	6.753
$RIPE_{PSH.f(t)/PSH}$	-0.614	-0.230	-0.342	-0.013	0.084
$RIPE_{S\text{-}PSH/PSH}$	-1.225	-0.224	-0.426	-0.083	-0.009
$RIPE_{WS\text{-}PSH/PSH}$	-1.223	-0.192	-0.401	-0.041	0.038

PSH: the PSH model; PSH.f(t): the PSH model with time-covariate interactions; S-PSH: the stopped PSH model; WS-PSH: the weighted stopped PSH model. RIPE is the relative increment of prediction errors where BS_{PSH} is a reference, i.e., $RIPE_{PSH.f(t)/PSH} = BS_{PSH.f(t)}/BS_{PSH}$ - 1. All entries are multiplyed by 100.



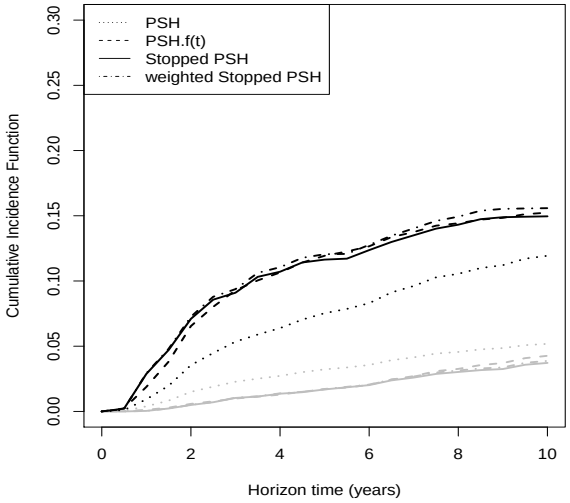
(a) The PSH model



(b) The PSH model with time-covariate interactions

The solid line is a locally weighted regression smooth with span= 0.5.

Figure 3: Plots of Shoenfeld-type residuals of the PSH model and the PSH model with time-covariate interactions.



Black curves are for $\boldsymbol{Z}_1$ = (placebo, L+XRT, age $<$ 50, 3 cm); and gray curves are for $\boldsymbol{Z}_2$ = (tamoxifen, mastectomy, age $\geq$ 50, 1 cm). PSH: the PSH model; PSH.f(t): the PSH model with time-covariate interactions.

Figure 4: Predicted cumulative incidences of locoreginal recurrence for two subgroups defined by $\boldsymbol{Z}_i$ = (treatment, surgery type, age, tumor size).

2.5 DISCUSSION

In this study, we propose a risk prediction modeling procedure to accurately estimate the cumulative incidence function for data with competing risks. The proposed methods are robust, simple to implement, not sensitive to the proportional subdistribution hazards assumption, and can incorporate multiple discrete and continuous prognostic factors without the need to test the model fit.

In the simulation studies, we compared cross-validated prediction errors and found that our proposed stopped PSH model has the same prediction accuracy as that obtained from the nonparametric estimates if all covariates are discrete; and it has the same prediction accuracy as the PSH model with time-covariate interactions when the functional form of time was correctly specified. Using real data from a breast cancer clinical trial, it illustrated that the proposed approach is straightforward and simple to implement for risk prediction. As compared with the PSH model with time-covariate interactions, our stopped PSH model has another appealing advantage that is it does not require modeling the functional forms of time-varying effects. Although the Schoenfeld-type residuals could be utilized to represent and test the functional forms, if the functional form is incorrectly specified, the residuals would have little use and the goodness-of-fit testing procedure would become less powerful [36]. There may also be concerned about the issue of overfitting if adding time-covariate interactions increases the number of parameters. As shown in our simulations, in the first non-PSH setting, adding time-covariate interactions may not capture all nonproportionality effects. In this situation, we found that the goodness-of-fit tests and the modeling procedure allowing time-covariate interactions into the PSH model are not applicable.

Our proposed models might be less efficient in some situations. As opposed to the PSH model which uses information over the entire follow-up period, the stopped PSH model uses information only up to the prediction horizon time. Therefore, the stopped PSH model might be less efficient than the PSH model at the early follow-up period. Also, if the censoring proportion is low prior to the horizon time, the nonparametric estimates of censoring probabilities could be less efficient. As a result, the weights derived from these censoring probabilities would yield unreliable results in the weighted stopped PSH model.

In estimation of the Brier score with the traditional method, one often uses the IPCW to deal with censoring. This approach could lead to biased estimates if the number of main events is low and the censoring percentage is high. An alternative strategy is to use the pseudo-observation approach to deal with high percent of censoring [5]. Moreover, if the number of main events is small, the cross-validation strategy is not effective. In this case, it is preferred to validate the performance of the models using an external dataset which is independent of the dataset used for fitting the model.

The proposed methods allow the users to predict the CIF at a given horizon time therefore can be used for risk stratification or therapy assignments. If a researcher's goal is to find the absolute risk of a specific event, our procedure is more attractive and less complex than the currently available models. However, our proposed models do not serve the purpose of estimating a covariate effect or capturing the overall trajectory of the effect through time.

3.0 LANDMARK PROPORTIONAL SUBDISTRIBUTION HAZARDS MODELS FOR DYNAMIC PREDICTION OF CUMULATIVE INCIDENCE PROBABILITIES

3.1 INTRODUCTION

A patient's baseline prognosis predicted at the moment of diagnosis or at the beginning of treatment often changes over time with the progression of disease. Patients who share similar baseline prognosis could have very different prognosis at later time points during follow-up because of population heterogeneity. In such cases, prognosis performed at the baseline are unable to reveal the underlying dynamic changes. Prognosis tools that can update patients' prognosis dynamically using information collected during follow-up about disease progression will be more helpful to patients and clinicians in their decision-making [4, 17].

A dynamic prediction refers to dynamically predicting a patient's prognosis at later time points during follow-up by incorporating all the time-dependent prognostic information collected up to the time points. The time-dependent information includes the potential time-varying covariate effects, i.e. the effect of some prognostic factors may diminish as time elapses from diagnosis, the intermediate clinical events, i.e. the occurrence of acute graft versus host disease (aGvHD) for acute lymphoid leukemia patients after bone marrow transplantation, and the longitudinally measured biomarkers, i.e. CD4 cell count for HIV patients and prostate-specific antigen (PSA) level for prostate cancer patients.

The quantities of primary interest in dynamic prediction are conditional survival probability for data containing no competing risks and conditional cause-specific cumulative incidence function for data with competing risks [31, 33]. The conditional survival probability is defined as the probability of surviving beyond a pre-specified time (the prediction horizon

time, t_{hor}) or surviving an additional w-years given that the patient is still alive at certain specified time points (the landmark time, t_{LM}) during follow-up. Similarly, the conditional cause-specific cumulative incidence function is the probability that the event of interest occurs before t_{hor} conditional on the patient who has not yet failed from any cause at t_{LM}. For those patients who have not experienced any event at t_{LM}, clinicians may predict the risk by estimating the next w-years cumulative incidence function based on all the prognostic information available at the landmark time point.

Several statistical methods have been used to incorporate the intermediate clinical events or the repeated measures of biomarkers in dynamic prediction. Van Houwelingen and Putter (2008) [32] utilized multi-state models to dynamically estimate the failure free survival during follow-up of acute lymphoid leukemia patients after transplantation given the history on aGvHD. Proust-Lima and Taylor (2009) [3] developed a joint modeling approach derived from a joint latent class model to dynamically predict the risk of recurrence given the longitudinal posttreatment PSA measures. Mauguen et al. (2013) [18] proposed a dynamic prognostic tool based on joint frailty models to predict patients' risk of death given the history on cancer recurrence.

Although multi-state models and joint models are useful in understanding the underlying process of a disease progression over time, predictions using either of these two models are complex and the implementation could be cumbersome. In multi-state models, where all the intermediate and terminal events are defined as states, each transition from one state to another need to be modeled separately [22] which could result in overfitting. In addition, when the Markov assumption is not satisfied in practice, the multi-state models could not provide explicit expressions for the dynamic prediction probabilities and additional simulation procedures are needed [9]. The limitation of applying joint models as a dynamic prognostic tool is the computational complexity in jointly modeling the longitudinal covariate processes and time-to-event data [3, 33]. Furthermore, developing multi-state models or joint models for data with various types of time-dependent information will be even more complicated.

Van Houwelingen (2007) [31] developed a more simple and straightforward dynamic prognostic tool for data containing no competing risks. In contrast to the multi-state models

and joint modeling approach, the proposed landmark Cox dynamic prediction models can incorporate time-varying covariate effects, intermediate events, and repeated measurements of biomarkers simultaneously into one single model that also provides an explicit expression for the conditional survival probability [31–33]. The landmark method avoids complex procedures of modeling either the time-varying effects or the longitudinal covariate processes, thus it is less demanding in computation and easier to implement than other methods of dynamic prediction.

The landmark method was originally introduced by Anderson et al. (1983) [1] to correct the immortal time bias in the analysis of survival of "responders" and "nonresponders" in cancer patients to evaluate the effect of response to chemotherapy. The immortal time bias was introduced when the survival probabilities were compared between "responders" and "nonresponders" from the start of the study, since the patients' response status was determined at later time points during follow-up but not at the baseline. This bias leaded the survival probability to be overestimated for "responders" but underestimated for "nonresponders". The landmark method is applied to patients who are alive at a fixed pre-specified time point, the landmark, during follow-up with survival analyses conditional on these patients using their response status measured at the landmark time. All patients who died or were censored before landmark are excluded from the analysis [1, 7].

Van Houwelingen (2007) [31] applied the landmark method to the Cox proportional hazards model and proposed the landmark Cox model to dynamically predict the probability of failure at prediction horizon time t_{hor} conditional on the subjects who have not yet failed at landmark time t_{LM} given the event's history and the time-dependent covariates until t_{LM}. The proposed landmark Cox model can also accommodate the time-varying effects, because it is robust against violation of the proportional hazards assumption. In addition, van Houwelingen (2007) [31] further developed more comprehensive landmark supermodels that requires fitting only one model for making dynamic prediction at any landmark point from a pre-specified interval, instead of fitting a landmark Cox model for each landmark time point separately.

To extend the landmark dynamic prediction method to data containing competing risks, Nicolaie et al. (2013) [19] developed landmark supermodels based on the cause-specific

hazards function. However, in order to estimate the conditional CIFs, this approach requires fitting the landmark supermodels for all causes of failure separately then combining the results. Hence, there is a lack of one-to-one correspondence in the interpretation of landmark-dependent covariate effects on the dynamic predictive probabilities. To directly predict the conditional CIFs, Nicolaie et al. (2013) [20] proposed landmark supermodels based on the dynamic pseudo-observations that are updated at each landmark point. Cortese et al. (2013) [5] applied the landmark method to the Fine-Gray proportional sub-distribution hazards (PSH) model [8] at a small set of pre-defined fixed time points. However, they did not assess robustness of their proposed model against violations of the PSH assumption, nor did they construct a more comprehensive landmark supermodel to reduce complexity in computation when there are multiple landmark time points of interest. The major difficulty encountered by other researchers attempting to develop a landmark model/supermodel based on the Fine-Gray model was in constructing the landmark subset at each landmark time point to properly deal with the competing events [19].

In this study, we overcame the aforementioned challenges and extended the landmark approach to the Fine-Gray PSH model for data with competing risks. We proposed a simple landmark PSH model that can be used to directly predict the conditional CIFs in one step and will provide simple and accurate prediction of the conditional CIFs for covariates with possible time-varying effects bypassing complex modeling of the process of time-varying covariates effect. We further developed a landmark PSH supermodel which can be used to make dynamic predictions at arbitrary landmark points by fitting only one model. In addition, we adopted the concept of sliding prediction window from van Houwelingen and Putter (2012) [33] and allowed the horizon time to vary. In this case, the landmark PSH models provide a dynamic prediction tool to estimate conditional CIFs at $t_{LM} + w$, where w is a fixed width of the prediction window. Compared to the currently available landmark models for data containing competing risks, our proposed landmark PSH models are straightforward, easy to implement and can yield a simpler and explicit estimation form for the desired conditional CIFs.

In the next section, we introduce the dynamic prediction probabilities in competing risks, and present our proposed landmark PSH model and landmark PSH supermodels. In

Section 3.4, we assess the performance in prediction for the proposed models and compare with the existing methods through simulations. In Section 3.5, we apply the proposed models to predict the w-year fixed width cumulative incidences of locoregional recurrence given a set of prognostic factors from a breast cancer clinical trial. The discussions are provided in Section 3.6.

3.2 DYNAMIC PREDICTION WITH TIME-VARYING COVARIATE EFFECTS

3.2.1 Conditional Cumulative Incidence Function

In Section 3.2, we only consider time-fixed covariates, but the covariate effects might be changing over time. The target dynamic prediction probability is the conditional cause-specific cumulative incidence function that is the cumulative incidence probability of occurring the event of interest by a pre-specified prediction horizon time t_{hor} conditional on the individuals who have not yet failed from any cause at a landmark time point t_{LM}.

Following the same notations introduced in Section 2.2.1, the corresponding conditional CIF for cause 1 is defined as

$$F_{1,LM}(t_{hor}|\boldsymbol{Z}, t_{LM}) = \Pr(T \leq t_{hor}, \varepsilon = 1 | T > t_{LM}, \boldsymbol{Z}), \tag{3.1}$$

where T is the failure time, $\varepsilon \in \{1, \ldots, k\}$ is the cause of failure, and $\boldsymbol{Z}$ is a p-dimensional vector of time-fixed covariates. Using the definition of conditional probabilities, we can rewrite the conditional CIF as

$$F_{1,LM}(t_{hor}|\boldsymbol{Z}, t_{LM}) = \frac{F_1(t_{hor}; \boldsymbol{Z}) - F_1(t_{LM}; \boldsymbol{Z})}{S(t_{LM}; \boldsymbol{Z})} = \frac{F_1(t_{hor}; \boldsymbol{Z}) - F_1(t_{LM}; \boldsymbol{Z})}{1 - \sum_{j=1}^{k} F_j(t_{LM}; \boldsymbol{Z})},$$

where $F_j(t_{LM}; \boldsymbol{Z})$ is the CIF for cause j and $S(t_{LM}; \boldsymbol{Z})$ is the overall survival at time t_{LM}.

Several methods can be used to estimate $F_{1,LM}(t_{hor}|\boldsymbol{Z}, t_{LM})$, including the nonparametric estimates, the cause-specific hazards model, the Fine-Gray proportional subdistribution hazards (PSH) model [8], the multistate model [6], the pseudo-value approach [15], and the

direct binomial regression model [26]. However, all these methods require at least two steps in estimating the conditional CIFs; the first step is to estimate the probabilities of having no event up to the landmark time; and the second step is to estimate the cumulative incidences of developing the main event between the landmark time and the horizon time. Moreover, to predict the conditional CIFs at a set of landmark time points using the aforementioned methods, one has to fit the models separately for each landmark point; which could be computationally cumbersome. In addition, the cause-specific hazards model and the Fine-Gray model are sensitive to the proportionality assumption which assumes that the covariate effects are constant over time, yet often the assumption is violated in a long-term follow-up. Therefore, these methods are not ideal for dynamic prediction when dealing with competing risks data.

In order to directly estimate the conditional CIF in one step, we generalize the landmarking approach proposed by van Houwelingen (2007) [31] into the Fine-Gray PSH model and propose a landmark PSH model. The proposed landmark PSH model can also accommodate the potential time-varying covariate effects because it is robust against the misspecification of the proportional subdistribution hazards.

3.2.2 Landmark PSH Model

To directly estimate the $F_{1,LM}(t_{hor}|\boldsymbol{Z}, t_{LM})$, we define a conditional subdistribution hazard

$$\lambda_1(t|\boldsymbol{Z}, t_{LM}) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr\{t \leq T \leq t + \Delta t, \varepsilon = 1 | T \geq t \cup (t_{LM} \leq T \leq t \cap \varepsilon \neq 1), \boldsymbol{Z}\}$$

for $t \geq t_{LM}$, which is the subdistribution hazards $\lambda_1(t|\boldsymbol{Z})$ defined in Section 2.2.1 conditional on the subjects who have no event occurred from any cause prior to t_{LM}. It can be shown that

$$\begin{aligned} F_{1,LM}(t_{hor}|\boldsymbol{Z}, t_{LM}) &= \Pr(t_{LM} < T \leq t_{hor}, \varepsilon = 1|\boldsymbol{Z}) / \Pr(T > t_{LM}|\boldsymbol{Z}) \\ &= 1 - \Pr\{(T > t_{hor}) \cup (t_{LM} < T \leq t_{hor} \cap \varepsilon \neq 1)|\boldsymbol{Z}\} / \Pr(T > t_{LM}|\boldsymbol{Z}) \\ &= 1 - \exp[-\{\Lambda_1(t_{hor}|\boldsymbol{Z}, t_{LM}) - \Lambda_1(t_{LM} - |\boldsymbol{Z}, t_{LM})\}] \\ &= 1 - \exp\left\{-\int_{t_{LM}}^{t_{hor}} \lambda_1(t|\boldsymbol{Z}, t_{LM})\, dt\right\}. \end{aligned} \tag{3.2}$$

Following the form of the PSH model, we define a *landmark PSH model* which is given by

$$\lambda_1(t|\boldsymbol{Z}, t_{LM}) = \lambda_{10}(t|t_{LM}) \exp\{\boldsymbol{Z}^T \boldsymbol{\beta}_{LM}\} \qquad \text{for } t_{LM} \leq t \leq t_{hor}. \tag{3.3}$$

Therefore, the conditional CIF can be easily computed from $\hat{F}_{1,LM}(t_{hor}|\boldsymbol{Z}, t_{LM}) = 1 - \exp\{-\int_{t_{LM}}^{t_{hor}} \hat{\lambda}_1(t|\boldsymbol{Z}, t_{LM})\, dt\}$. The model can be fitted by applying the Fine-Gray PSH model to the data of individuals who have not failed from any cause at t_{LM} and ignoring the events occurring after the horizon time t_{hor} by adding administrative censoring at t_{hor}.

The landmark PSH model with truncation at t_{LM} and administrative censoring at t_{hor} is also robust against violations of the proportional subdistribution hazards assumption. Under the nonproportionality, based on the results showed Chapter 2, we can derive that the partial likelihood estimator $\hat{\boldsymbol{\beta}}_{LM}$ calculated from the landmark PSH model is a weighted average of the true time-varying effects in the interval $[t_{LM}, t_{hor}]$ and the corresponding $\hat{F}_{1,LM}(t_{hor}|\boldsymbol{Z}, t_{LM})$ is an approximately correct estimate for the underlying true conditional CIF at time t_{hor}.

So, the proposed landmark PSH model provides a convenient and straightforward way to predict the conditional CIF $F_{1,LM}(t_{hor}|\boldsymbol{Z}, t_{LM})$ in one model. In addition, it can incorporate the potential time-varying covariate effects into dynamic prediction via a simple model form without constructing complex procedures to model the time-varying effects.

3.2.3 Landmark PSH Supermodel

Different from the fixed horizon time, van Houwelingen and Putter (2012) [33] proposed the idea of sliding prediction window, which is taking a window with a fixed width of w and predicting the failure probabilities at $t_{hor} = t_{LM} + w$ given that the subjects have not experienced any event before t_{LM}. The idea of sliding window is the basis of dynamic prediction. Generalizing this idea to data with competing risks, our aim is to predict the cumulative probabilities at the next w years conditional on the individuals who are event-free at a set of landmark time points $t_{LM} = s$ within an interval $[s_0, s_L]$. The target conditional CIF is defined as $F_{1,LM}(s + w|\boldsymbol{Z}, s)$.

Applying landmark PSH models to dynamic prediction is less practical because it is necessary to fit a landmark PSH model for each landmark point separately. We used the smoothing strategy of van Houwelingen (2007) [31] for the Cox model to develop a landmark supermodel based on the PSH model. Similar to the landmark Cox model, the landmark PSH supermodel is also based on an augmented dataset which is constructed as follows: we first select a set of landmark points s from an interval $[s_0, s_L]$; for each s, we create a landmarking subset by selecting the individuals who have not failed from any cause at s and adding administrative censoring at $s + w$ for those who have events occur after $s + w$; and then stacking all the individual landmarking subsets into a super prediction dataset.

3.2.3.1 Stratified Landmark PSH Supermodel As discussed in Chapter 2, we have shown that the $\hat{\boldsymbol{\beta}}_{LM}$ calculated from model (3.3) is consistent for a weighted average of possibly time-varying effects over the interval $[s, s + w]$. As discussed in van Houwelingen (2007) [31], we can also expect the effect of s on $\boldsymbol{\beta}_{LM}(s)$ in a smooth way and model $\boldsymbol{\beta}_{LM}(s)$ as continuous functions of s. Then, we have the *stratified landmark PSH supermodel*

$$\lambda_1(t|\boldsymbol{Z}, s, w) = \lambda_{10}(t|s, w) \exp\{\boldsymbol{Z}^T \boldsymbol{\beta}_{LM}(s)\} \qquad \text{for } s \leq t \leq s + w, \tag{3.4}$$

where $\lambda_{10}(t|s, w)$ is the unspecified baseline subdistribution hazard for each s. For simplicity, we can fit a linear model for $\boldsymbol{\beta}_{LM}(s)$, which takes the form

$$\boldsymbol{\beta}_{LM}(s) = \boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta}) = \sum_{j=1}^{m_\beta} \boldsymbol{\theta}_j f_j(s)$$

where $f(s)$ is a set of parametric functions of s and $\boldsymbol{\theta}_j$ is a vector of parameters.

Model (3.4) can be fitted by applying a PSH model with landmark-covariate interactions $\boldsymbol{Z} * f_j(s)$ to the stacked dataset with stratification on s. To obtain the consistent estimates, we impose that any landmark time point s satisfies $s + w < \tau$, where τ is chosen such that $G(\tau) > 0$ and $S(\tau) > 0$. The consistent estimates of $\boldsymbol{\theta}$ can be obtained by maximizing a pseudo partial log-likelihood

$$ipl(\boldsymbol{\theta}) = \sum_{i=1}^{n} I(\varepsilon_i = 1) \left(\sum_{s: s \leq T_i \leq s+w} \ln \left[\frac{\exp\{\boldsymbol{Z}_i^T \boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta})\}}{\sum_{j \in R^s(T_i)} \exp\{\boldsymbol{Z}_j^T \boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta})\}} \right] \right)$$

where $R^s(T_i) = \{j : (T_i \leq T_j \leq s+w) \cup (s < T_j \leq T_i \cap \varepsilon_j \neq 1)\}$. The $ipl(\boldsymbol{\theta})$ is similar as that shown in van Houwelingen (2007) [31]; the difference is the modified risk sets $R^s(T_i)$. In the presence of random right censoring, the $ipl(\boldsymbol{\theta})$ need to be modified by including the IPCW weights. In terms of the counting process, the weighted $ipl(\boldsymbol{\theta})$ takes the form

$$ipl_w(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^{\infty} \sum_{s:s\leq t\leq s+w} \ln \left[\frac{\exp\{\boldsymbol{Z}_i^T \boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta})\}}{\sum_j \omega_j^s(t) Y_j^s(t) \exp\{\boldsymbol{Z}_j^T \boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta})\}} \right] \omega_i^s(t)\, dN_i^s(t)$$

where $N_i^s(t) = I(s < T_i \leq t, \varepsilon_i = 1)$, $Y_i^s(t) = I(t \leq T_i \leq s+w) + I(s < T_i < t, \varepsilon_i \neq 1)$ and $\omega_i^s(t) = \frac{\hat{G}^s(t)}{\hat{G}^s(X_i)} I(s < X_i < t, \varepsilon_i \neq 1) + I(t \leq X_i \leq s+w)$, $\hat{G}^s(t)$ is the Kaplan-Meier estimate of the conditional censoring survival distribution $G^s(t) = \Pr(C \geq t | C \geq s)$. To obtain the standard errors for the estimated parameters, a robust sandwich estimator is required to adjust for the correlation between the risk sets because the same subject is repeatedly used when we estimate the parameters based on the stacked dataset.

For each stratum s, the Breslow-type estimator of the baseline conditional subdistribution hazard is

$$\hat{\lambda}_{10}(T_i|s,w) = \frac{1}{\sum_{j\in R^s(T_i)} \exp\{\boldsymbol{Z}_j^T \hat{\boldsymbol{\beta}}_{LM}(s)\}} \qquad \text{for } s \leq T_i \leq s+w. \tag{3.5}$$

Similarly, in the presence of random right censoring, the Breslow-type estimator need to be modified by the IPCW weights.

In the implementation, the `R` function `coxph()` can be used to fit model (3.4) and also provide the robust sandwich estimates for the standard errors; but it requires a transformation for each landmarking subset in the process of constructing the super prediction dataset. Before stacking the subsets into a big dataset, following the strategy in Geskus (2011) [12], we need to transform each subset into the counting process style and include time-varying IPCW weights for the subjects experienced competing events when random right censoring is also present. We will then construct a new super dataset by stacking all the transformed subsets together and run the PSH model in `coxph()` stratified by s.

3.2.3.2 Proportional Baselines Landmark PSH Supermodel

In practice, the limitation of the stratified landmark PSH model is that it estimates the baseline subdsitribution hazard separately for each landmark point. As the Breslow estimator (3.5) showed, the dependence of $\hat{\lambda}_{10}(T_i|s,w)$ on s is through $\hat{\boldsymbol{\beta}}_{LM}(s)$; so that we can also expect that the $\lambda_{10}(t|s,w)$ varies continuously with s. Following the strategy of van Houwelingen (2007) [31], we can directly model the $\lambda_{10}(t|s,w)$ as

$$\lambda_{10}(t|s,w) = \lambda_{10}(t)\exp\{\boldsymbol{\gamma}(s)\}.$$

This lead to the *proportional baselines landmark PSH supermodel* (PBLM-PSH supermodel)

$$\lambda_1(t|\boldsymbol{Z},s,w) = \lambda_{10}(t)\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_{LM}(s) + \boldsymbol{\gamma}(s)\} \tag{3.6}$$

for $s_0 \leq t \leq s_L + w$. Similarly, $\boldsymbol{\gamma}(s)$ can be fitted as a linear model

$$\boldsymbol{\gamma}(s) = \boldsymbol{\gamma}(s|\boldsymbol{\eta}) = \sum_{j=1}^{m_{\lambda_{10}}} \eta_j g_j(s),$$

where $g(s)$ is a set of parametric functions of s and $\boldsymbol{\eta}$ is a vector of parameters. Note that $g(s)$ do not need to contain the constant term and have the restriction of $g_j(s_0) = 0$ for all $j(j = 1,\ldots,m_{\lambda_{10}})$ due to identifiability of the baseline subdistribution hazard, some discussion as shown in van Houwelingen (2007) [31].

Model (3.6) can still be fitted by applying a PSH model with landmark-covariate interactions $\boldsymbol{Z} * f_j(s)$ to the stacked dataset. Instead of stratifying on the landmark point, it directly fits s as $g(s)$ into model. To consistently estimate the parameters $(\boldsymbol{\theta},\boldsymbol{\eta})$, a Breslow partial log-likelihood is required for tied events. This is because in the stacked dataset, one subject with event time T_i has $n_i = \#\{s : s \leq T_i \leq s+w, s \in [s_0, s_L]\}$ repeated observations. Thus, fitting the model is equivalent to maximizing a different pseudo partial log-likelihood which is given by

$$ipl^*(\boldsymbol{\theta},\boldsymbol{\eta}) = \sum_{i=1}^{n} I(\varepsilon_i = 1)\ln\left[\frac{\sum_{s:s\leq T_i\leq s+w}\exp\{\boldsymbol{Z}_i^T\boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta}) + \boldsymbol{\gamma}(s|\boldsymbol{\eta})\}}{\sum_{s:s\leq T_i\leq s+w}\sum_{j\in R^s(T_i)}\exp\{\boldsymbol{Z}_j^T\boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta}) + \boldsymbol{\gamma}(s|\boldsymbol{\eta})\}}\right].$$

Similarly, in the presence of random right censoring, the $ipl^*(\boldsymbol{\theta}, \boldsymbol{\eta})$ need to be weighted by the IPCW weights. In terms of the counting process, the weighted $ipl^*(\boldsymbol{\theta}, \boldsymbol{\eta})$ takes the form

$$ipl_w^*(\boldsymbol{\theta}, \boldsymbol{\eta}) = \sum_{i=1}^{n} \int_0^{\infty} \ln \left[\frac{\sum_{s:s\leq t\leq s+w} \omega_i^s(t) \exp\{\boldsymbol{Z}_i^T \boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta}) + \boldsymbol{\gamma}(s|\boldsymbol{\eta})\}}{\sum_{s:s\leq t\leq s+w} \sum_j \omega_j^s(t) Y_j^s(t) \exp\{\boldsymbol{Z}_j^T \boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta}) + \boldsymbol{\gamma}(s|\boldsymbol{\eta})\}} \right] d\bar{N}_i(t),$$

where $\bar{N}_i(t) = I(s_0 < T_i \leq t \leq s_L + w, \varepsilon_i = 1)$. Again, robust estimates of the standard errors for the estimated parameters can be obtained by using the sandwich procedure.

For complete and censoring complete data, the estimate of $\lambda_{10}(T_i)$ is given by

$$\hat{\lambda}_{10}(T_i) = \frac{\#\{s : s \leq T_i \leq s + w, \varepsilon_i = 1\}}{\sum_{s:s\leq T_i\leq s+w} \sum_{j\in R^s(T_i)} \exp\{\boldsymbol{Z}_j^T \hat{\boldsymbol{\beta}}_{LM}(s) + \hat{\boldsymbol{\gamma}}(s)\}} \quad \text{for } s_0 \leq T_i \leq s_L + w \tag{3.7}$$

and the cumulative subdistribution hazard is estimated by $\hat{\Lambda}_{10}(t) = \sum_{(T_i\leq t, \varepsilon_i=1)} \hat{\lambda}_{10}(T_i)$. When random right censoring exists, $\hat{\Lambda}_{10}(t)$ takes the form

$$\hat{\Lambda}_{10}(t) = \sum_{i=1}^{n} \int_0^{t} \frac{\sum_{s:s\leq u\leq s+w} \omega_i^s(u)}{\sum_{s:s\leq u\leq s+w} \sum_j \omega_j^s(u) Y_j^s(u) \exp\{\boldsymbol{Z}_j^T \boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta}) + \boldsymbol{\gamma}(s|\boldsymbol{\eta})\}} d\bar{N}_i(u).$$

Thus, the target dynamic prediction probabilities $F_{1,LM}(s + w|\boldsymbol{Z}, s)$ have a simple and explicit estimation form, which is given by

$$\hat{F}_{1,LM}(s + w|\boldsymbol{Z}, s) = 1 - \exp\left[-e^{\boldsymbol{Z}^T \hat{\boldsymbol{\beta}}_{LM}(s) + \hat{\boldsymbol{\gamma}}(s)} \{\hat{\Lambda}_{10}(s + w) - \hat{\Lambda}_{10}(s-)\}\right] \tag{3.8}$$

for all $s \in [s_0, s_L]$. Compared to model (3.4) which can only get the prediction for the pre-specified landmark points because the baseline subdistribution hazard is specific for each s, the proportional baselines landmark PSH supermodel can provide the prediction of $F_{1,LM}(s + w|\boldsymbol{Z}, s)$ in any period of length w starting anywhere in $[s_0, s_L]$. Note that model (3.6) assume that the effect of s on the baseline subdistribution hazard in an additive way. As discussed in van Houwelingen (2007) [31], this assumption hold if the follow-up is not too long or the effect of covariates is not too big. So, if we choose an optimal width w for the prediction window and a rational range $[s_0, s_L]$ for the landmark points, according to the robustness of the PSH model, the proportional baselines landmark PSH supermodel directly provides a correct approximation for the conditional cumulative incidence function at time $s + w$ for any $s \in [s_0, s_L]$.

In practice, fitting the PBLM-PSH supermodel in the stacked dataset requires software that allows for delayed entry or left truncation at s. Following the same data transformation strategies we mentioned before, the `coxph()` can be still used; and the provided robust sandwich covariance matrix for $(\boldsymbol{\theta}, \boldsymbol{\eta})$ can be used for the significance testing of the estimated regression coefficients. As we discussed before, the landmark effect on the baseline subdistribution hazard is through $\boldsymbol{\beta}_{LM}(s)$, so that there is a correlation between $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. Van Houwelingen (2007) [31] suggested to center the covariates, so that the correlation would be reduced.

3.2.4 Measure of Predictive Accuracy

To evaluate the dynamic predictive accuracy of the proposed procedures, we adapted the time-dependent Brier score which is an estimate of the mean-squared prediction errors of the predicted event probabilities at $t_{hor} = s + w$ over the observed event status for subjects who are still alive at landmark s [5, 28]. For competing risks data, the expected time-dependent Brier score at landmark s for the prediction at horizon t_{hor} is defined as

$$\mathbf{BS}_{LM}(t_{hor}, s) = E\left[\{I(T \leq t_{hor}, \varepsilon = 1) - F_{1,LM}(t_{hor}|\boldsymbol{Z}, s)\}^2 | T > s\right],$$

where $t_{hor} > s$. For complete data, $\mathbf{BS}_{LM}(t_{hor}, s)$ can be consistently estimated by

$$\widehat{\mathbf{BS}}_{LM}(t_{hor}, s) = \frac{1}{n_s} \sum_{i \in R_s} \{I(T_i \leq t_{hor}, \varepsilon_i = 1) - \hat{F}_{1,LM}(t_{hor}|\boldsymbol{Z}_i, s)\}^2,$$

where $R_s = \{i : T_i > s\}$ and n_s is the number of subjects in R_s. When random right censoring exists, we utilize a pseudovalue-based consistent estimator of $\mathbf{BS}_{LM}(t_{hor}, s)$, which was proposed by Cortese et al. (2013) [5]

$$\widehat{\mathbf{BS}}_{LM}(t_{hor}, s) = \frac{1}{\tilde{n}_s} \sum_{i \in \tilde{R}_s} \left[\hat{Q}^{(i)}_{1,LM}(t_{hor}|s)\{1 - 2\hat{F}_{1,LM}(t_{hor}|\boldsymbol{Z}_i, s)\} + \hat{F}_{1,LM}(t_{hor}|\boldsymbol{Z}_i, s)^2\right],$$

where $\tilde{R}_s = \{i : X_i > s\}$ and $\tilde{n}_s$ is the number of subjects in $\tilde{R}_s$. $\hat{Q}^{(i)}_{1,LM}(t_{hor}|s)$ is a jackknife pseudovalue for ith subject who is still alive at s, which is defined by

$$\hat{Q}^{(i)}_{1,LM}(t_{hor}|s) = \tilde{n}_s \hat{F}_{1,LM}(t_{hor}|s) - (\tilde{n}_s - 1)\hat{F}^{(i)}_{1,LM}(t_{hor}|s),$$

where $\hat{F}_{1,LM}(t_{hor}|s)$ is the nonparametric estimate of the conditional cumulative incidence function $\Pr(T \leq t_{hor}, \epsilon = 1|T > s)$ and $\hat{F}^{(i)}_{1,LM}(t_{hor}|s)$ is the same estimate but based on the data where the ith subject has been removed.

3.3 DYNAMIC PREDICTION WITH TIME-DEPENDENT COVARIATES

3.3.1 Landmark PSH Supermodel

When there are time-dependent covariates that may be the occurrence of intermediate events and/or repeated measurements of biomarkers, the clinical interest is in the dynamic prediction of cumulative incidences given the covariates history available up to the landmark time point s. The target conditional CIF is defined as

$$F_{1,LM}(s+w|\boldsymbol{Z}(s), s) = \Pr(T \leq s+w, \varepsilon = 1|T > s, \boldsymbol{Z}(s)),$$

where $\boldsymbol{Z}(s)$ include the time-fixed covariates measured at baseline and the time-dependent covariates whose values are fixed at landmark time s.

To incorporate time-dependent covariates and time-varying covariate effects simultaneously, the landmark PSH supermodels provide a simpler way in dynamic prediction for data with competing risks. The stratified landmark PSH supermodel becomes

$$\lambda_1(t|\boldsymbol{Z}(s), s, w) = \lambda_{10}(t|s, w) \exp\{\boldsymbol{Z}(s)^T \boldsymbol{\beta}_{LM}(s)\} \quad \text{for } s \leq t \leq s+w, \tag{3.9}$$

where $\boldsymbol{\beta}_{LM}(s)$ is also a continuous function of s, can be modeled as $\boldsymbol{\beta}_{LM}(s) = f(s)\boldsymbol{\theta}$. The parameter $\boldsymbol{\beta}$ can be consistently estimated by maximizing

$$ipl(\boldsymbol{\theta}) = \sum_{i=1}^{n} I(\varepsilon_i = 1) \left(\sum_{s: s \leq T_i \leq s+w} \ln \left[\frac{\exp\{\boldsymbol{Z}_i(s)^T \boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta})\}}{\sum_{j \in R^s(T_i)} \exp\{\boldsymbol{Z}_j(s)^T \boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta})\}} \right] \right).$$

The estimate of the landmark-specific baseline conditional subdistribution hazards $\lambda_{10}(t|s, w)$ is given by

$$\hat{\lambda}_{10}(T_i|s, w) = \frac{1}{\sum_{j \in R^s(T_i)} \exp\{\boldsymbol{Z}_j(s)^T \hat{\boldsymbol{\beta}}_{LM}(s)\}} \quad \text{for } s \leq T_i \leq s+w.$$

As before, we can assume multiplicative effects of landmark s on $\lambda_{10}(t|s,w)$ and have the proportional baseline landmark PSH supermodel

$$\lambda_1(t|\boldsymbol{Z}(s),s,w) = \lambda_{10}(t)\exp\{\boldsymbol{Z}(s)^T\boldsymbol{\beta}_{LM}(s) + \boldsymbol{\gamma}(s)\} \tag{3.10}$$

for $s_0 \leq t \leq s_L + w$, where $\boldsymbol{\beta}_{LM}(s) = f(s)\boldsymbol{\theta}$ and $\boldsymbol{\gamma}(s) = g(s)\boldsymbol{\eta}$. The consistent estimates of the parameters $(\boldsymbol{\theta}, \boldsymbol{\eta})$ can be obtained by maximizing

$$ipl^*(\boldsymbol{\theta},\boldsymbol{\eta}) = \sum_{i=1}^{n} I(\varepsilon_i = 1)\ln\left[\frac{\sum_{s:s\leq T_i\leq s+w}\exp\{\boldsymbol{Z}_i(s)^T\boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta}) + \boldsymbol{\gamma}(s|\boldsymbol{\eta})\}}{\sum_{s:s\leq T_i\leq s+w}\sum_{j\in R^s(T_i)}\exp\{\boldsymbol{Z}_j(s)^T\boldsymbol{\beta}_{LM}(s|\boldsymbol{\theta}) + \boldsymbol{\gamma}(s|\boldsymbol{\eta})\}}\right].$$

The estimate of $\lambda_{10}(T_i)$ is given by

$$\hat{\lambda}_{10}(T_i) = \frac{\#\{s : s \leq T_i \leq s+w, \varepsilon_i = 1\}}{\sum_{s:s\leq T_i\leq s+w}\sum_{j\in R^s(T_i)}\exp\{\boldsymbol{Z}_j(s)^T\hat{\boldsymbol{\beta}}_{LM}(s) + \hat{\boldsymbol{\gamma}}(s)\}} \quad \text{for } s_0 \leq T_i \leq s_L + w.$$

Thus, the conditional CIF $F_{1,LM}(s+w|\boldsymbol{Z}(s),s)$ has a simple estimation form

$$\hat{F}_{1,LM}(s+w|\boldsymbol{Z}(s),s) = 1 - \exp\left[-e^{\boldsymbol{Z}(s)^T\hat{\boldsymbol{\beta}}_{LM}(s)+\hat{\boldsymbol{\gamma}}(s)}\{\hat{\Lambda}_{10}(s+w) - \hat{\Lambda}_{10}(s-)\}\right]$$

for all $s \in [s_0, s_L]$.

3.3.2 Adjusted Landmark PSH Supermodel

In the landmark PSH supmermodel 3.9 and 3.10, we assume that the Markov property is held where the distribution of the future depends only on the current value, and we use $\boldsymbol{Z}(s)$ as a proxy for the true value of $\boldsymbol{Z}(t)$. But, if the $\boldsymbol{Z}(t)$ is changing rapidly over time, the big variation between $\boldsymbol{Z}(s)$ and $\boldsymbol{Z}(t)$ might cause an attenuated covariate effects $\boldsymbol{\beta}_{LM}(s)$ and subsequently provide bias estimation of the conditional CIF.

To adjust for this issue, we can explore a set of suitable landmark points to postulate the process of $\boldsymbol{Z}(t)$ and model the attenuation process. Van Houwelingen and Putter (2012) [33] proposed a simple method to adjust the attenuated covariate effects by adding a monotonically decreasing function $\varphi(t-s)$ with $\varphi(0)=1$ to $\varphi(\infty)=0$. Utilizing this strategy, the stratified landmark PSH supermodel can be modified as

$$\lambda_1(t|\boldsymbol{Z}(s),s,w)=\lambda_{10}(t|s,w)\exp\{\boldsymbol{Z}(s)^T(\boldsymbol{\beta}_{LM}(s)*\varphi(t-s))\} \quad \text{for } s\le t\le s+w, \quad (3.11)$$

where $*$ denotes coordinate-wise multiplication. Similarly, the PBLM-PSH supermodel can also be adjusted. As a note, to fit the adjusted landmark PSH supermodels, when construct the super stacked dataset, at each landmark-specific subset, we need to expand each individual's record at each combinition of landmark point s and event time t.

3.4 SIMULATION STUDIES

We evaluated the performance of the proposed methods for dynamic prediction using simulated data under two different settings of nonproportional subdistribution hazards (non-PSH). The time-dependent Brier scores were compared among the proposed landmark PSH model, the landmark PSH supermodel, the nonparametric method, and the standard PSH model. For simplicity, only two failure types were considered: type 1 failure is the main event of interest; type 2 failure indicates competing events.

In the first non-PSH setting, we generated the type 1 failure times from a two-parameter Weibull mixture distribution with the subdistribution function

$$F_{1,1^{st}nonPSH}(t;Z_i)=p(1-\exp[-\{\lambda_1\exp(Z_i\beta_1)t\}^{\alpha_1}])$$

with $(\alpha_1, \lambda_1, \beta_1) = (3.2, 0.18, -0.81)$, where the rate parameter depends on the covariate Z_i. In the second non-PSH setting, we let the coefficient of Z_i be a function of time; and the subdistribution of the main event became

$$F_{1,2^{nd}nonPSH}(t; Z_i) = 1 - (1 - p[1 - \exp\{-(\lambda_2 t)^{\alpha_2}\}])^{\exp\{Z_i\beta_{21} + Z_i\beta_{22}\ln(t+1)\}}$$

with $(\alpha_2, \lambda_2, \beta_{21}, \beta_{22}) = (3.2, 0.12, 0.8, 0.3)$. In both settings, we considered Z_i as a discrete covariate from Bernoulli(0.5), and let $p = 0.3$, which produced about 30% main events at $Z_i = 0$ when there was no censoring. We generated the type 2 failure times from an exponential distribution $\Pr(T_i \leq t|\varepsilon_i = 2, Z_i) = 1 - \exp\{-\exp(Z_i\beta_c t)\}$ by taking $\Pr(\varepsilon_i = 2|Z_i) = 1 - \Pr(\varepsilon_i = 1|Z_i)$, where $\beta_c = 0.5$. Sample size of $n = 1000$ was chosen and the data were simulated repeatedly for $N = 1000$ times. The censoring times were generated independently from a uniform distribution which resulted in about 20% censoring.

The performance of the proposed landmark PSH model and the landmark PSH supermodel in dynamic prediction of the conditional CIFs using a fixed width of w at a set of landmark points were compared with the performance of nonparametric method and the standard PSH model. We chose a prediction window of width $w = 3$ for the first non-PSH setting and $w = 2$ for the second non-PSH setting. For fitting the landmark PSH supermodel, we set up a fine grid of landmark points with equidistant step of 0.1 from 0 to 5 for the first non-PSH setting and from 0 to 4 for the second non-PSH setting. In both settings, we took ordinary polynomials for the basis functions as $f(s) = \{1, s, s^2\}$ and $g(s) = \{s, s^2\}$. Since the performance of the stratified landmark PSH supermodel is as good as that of the proportional baselines landmark supermodel (PBLM-PSH supermodel), we only present the simulation results of the PBLM-PSH supermodel in this section. Note that the landmark PSH supermodel in absence of censoring cannot be fitted using the `coxph()` function in R, because that the competing risks data cannot be transformed into the counting process format if there is no censoring; and that the model cannot be fitted by the `crr()` function in R either since `crr()` does not allow delayed entries. Therefore, we only fitted our landmark PSH model (LM-PSH model) in complete cases. Figure 3.4 depicts the true and estimated conditional cumulative incidence probabilities obtained from different approaches. We found

that in two different non-PSH scenarios, the performance of the landmark PSH model and the PBLM-PSH supermodel are as good as that of the nonparametric methods.

For each approach we evaluated the prediction errors in the dynamic conditional CIFs by estimating time-dependent Brier scores, and we used a 3-fold cross-validation to correct for possible overfitting. Table 4 shows averaged estimates of the cross-validated time-dependent Brier score and its empirical standard deviation. To quantify the improvement of predictive accuracy for the proposed landmark PSH models to the standard PSH model under nonproportional hazards, we utilized a relative increment (or reduction) of prediction errors by treating the nonparametric estimates as a reference. The relative increment of prediction errors are presented in Figure 6. As expected, in both non-PSH settings the predictive accuracy of the LM-PSH model was almost the same as that obtained from the nonparametric method. As compared with the LM-PSH model or the nonparametric method, the PBLM-PSH supermodel has slightly lower accuracy, yet the differences in prediction errors are negligible.

Table 4: Cross-validated (3-fold) estimates for the time-dependent Brier score and its empirical standard deviation (SD).

(a) 1st non-PSH setting: prediction window with fixed width of $w = 3$

Landmark		NP		PSH		LM-PSH		PBLM-PSH	
t_{LM}	Censoring	Ave.	SD	Ave.	SD	Ave.	SD	Ave.	SD
0	0	2.059	0.430	2.087	0.441	2.074	0.434		
	20%	2.058	0.470	2.066	0.474	2.058	0.470	2.076	0.476
1	0	4.065	0.498	4.323	0.554	4.116	0.504		
	20%	6.986	0.889	7.126	0.923	6.986	0.889	7.093	0.909
2	0	5.182	0.411	6.124	0.500	5.271	0.423		
	20%	12.529	0.854	13.207	0.897	12.529	0.854	12.818	0.934
3	0	4.678	0.406	6.405	0.385	4.790	0.427		
	20%	14.211	1.104	15.512	0.882	14.212	1.104	14.666	1.238
4	0	3.525	0.448	5.556	0.356	3.624	0.467		
	20%	12.700	1.820	14.054	1.513	12.707	1.819	13.233	1.939
5	0	2.775	0.409	4.424	0.331	2.851	0.423		
	20%	11.338	2.269	12.082	2.055	11.356	2.270	12.043	2.441

Table 4 continued

(b) 2nd non-PSH setting: prediction window with fixed width of $w = 2$

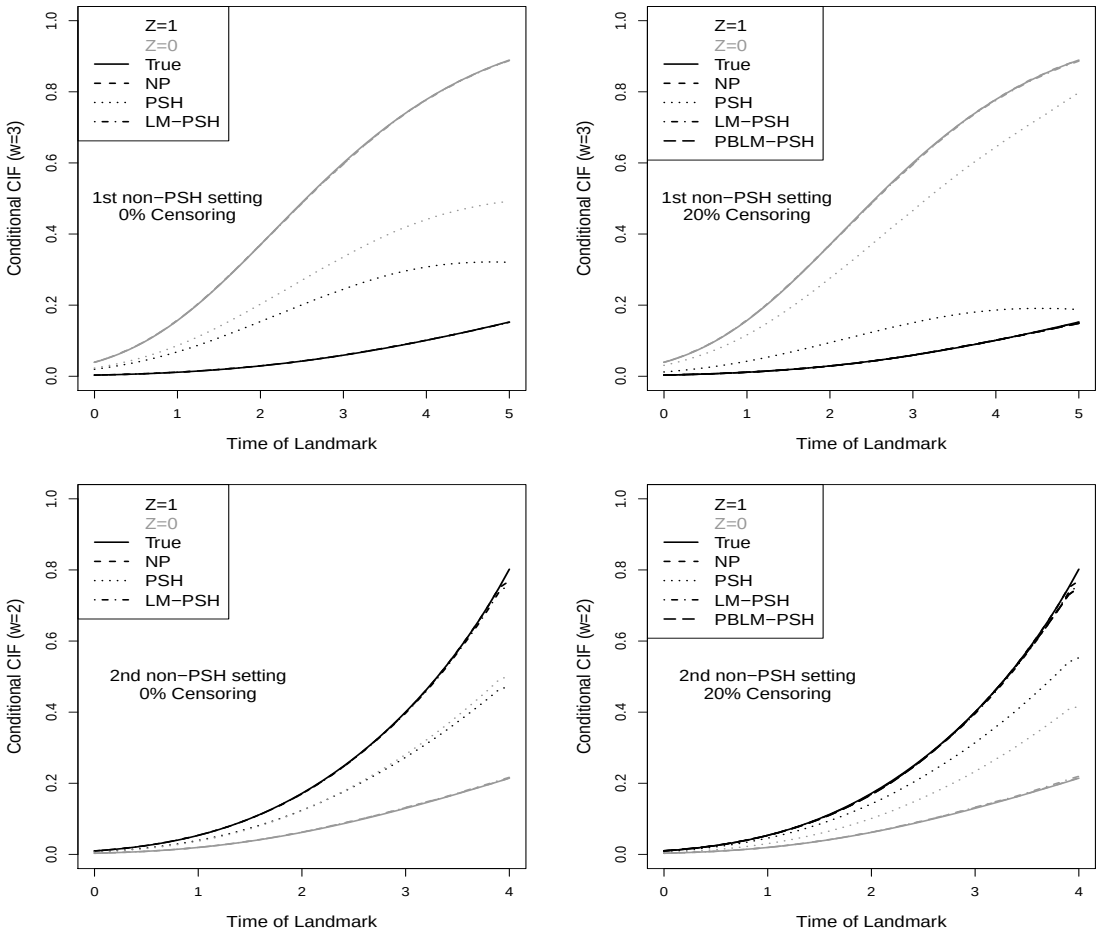
Landmark		NP		PSH		LM-PSH		PBLM-PSH	
t_{LM}	Censoring	Ave.	SD	Ave.	SD	Ave.	SD	Ave.	SD
0	0	0.639	0.253	0.640	0.253	0.643	0.255		
	20%	0.640	0.270	0.641	0.271	0.640	0.270	0.646	0.273
1	0	2.257	0.456	2.275	0.462	2.282	0.461		
	20%	3.536	0.792	3.550	0.796	3.536	0.792	3.579	0.802
2	0	4.662	0.522	4.797	0.543	4.731	0.535		
	20%	10.078	1.247	10.211	1.257	10.078	1.247	10.253	1.277
3	0	6.326	0.452	7.003	0.453	6.451	0.477		
	20%	17.716	1.271	18.642	1.181	17.718	1.271	18.180	1.407
4	0	4.739	0.457	7.100	0.344	4.859	0.477		
	20%	17.713	1.751	21.588	1.323	17.127	1.750	17.936	1.983

NP: the nonparametric method; PSH: the standard PSH model; LM-PSH: the landmark PSH model; PBLM-PSH: the proportional baselines landmark PSH supermodel. All entries are multiplyed by 100.

3.5 APPLICATION

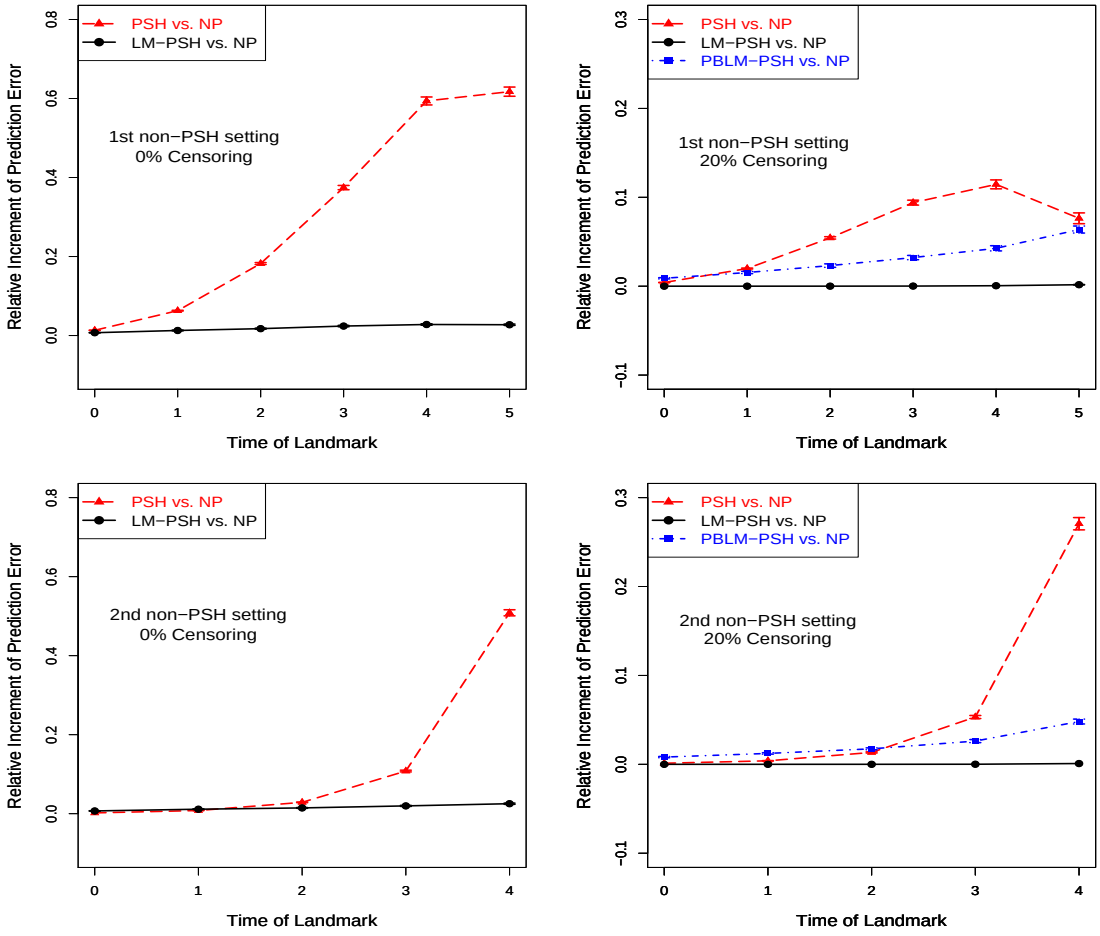
3.5.1 Application 1

To illustrate the use of our proposed landmark PSH supermodel in prediction of the conditional cumulative incidence probabilities for a moving (or dynamic) time interval with fixed width, we used the data from the NSABP B-20 trial which was a multicenter phase III clinical trial for women with estrogen receptor positive and historically nodes-negative primary breast cancer. In this trial, 2,363 patients were randomly selected to receive one of the following three regimens: tamoxifen 10mg daily for 5 years, tamoxifen 10mg daily for 5 years plus Metrotrexate (M) and Fluorouracil (F), and tamoxifen 10mg daily for 5 years plus M and F, and Cyclophosphamide (C). For simplicity, we will call these treatments as TAM, TAM+MF, and TAM+CMF, respectively. Among the 2,272 clinically eligible patients followed up, 119 developed locoregional recurrence (LRR); 482 had other events before LRR, including distant recurrence, second primary cancers, and death; and the remaining 1,671 were censored. The median follow-up time was 11.2 years. Our main interest in this application is to dynamically predict the conditional cumulative incidence of LRR for the subsequent w years given no event occurred before the landmark time point s. Prognostic



Black curves: the $Z = 1$ group; gray curves: the $Z = 0$ group. True: underlying true conditional CIFs. NP: nonparametric method; PSH: the standard PSH model; LM-PSH; the landmark PSH model; PBLM-PSH: the proportional baselines landmark PSH supermodel.

Figure 5: Predicted conditional cumulative incidence functions over w years (averaged over 1000 simulations) at a set of landmark time points.



The prediction errors were cross-validated (3-fold) estimates for the time-dependent Brier scores, where $w = 3$ for the 1st non-PSH setting and $w = 2$ for the 2nd non-PSH setting. NP: nonparametric method; PSH: the standard PSH model; LM-PSH; the landmark PSH model; PBLM-PSH: the proportional baselines landmark PSH supermodel.

Figure 6: Relative increment of prediction errors (and their standard deviation) at a set of landmark time points.

covariates of interest include the treatment type (TAM, TAM+MF, or TAM+CMF), surgery type (lumpectomy plus radiation therapy [L+XRT] vs. mastectomy), age at the study entry (< 50 vs. ≥ 50 years old), clinical tumor size (≤ 2 vs. > 2 cm), and tumor grade (well, moderate, and poor). All covariates were measured at the baseline.

Figure 7(a) and 7(b) shows that the estimated cumulative incidence of LRR is stable after 12 years (i.e., only a few LRR events occurred); and that few random censoring events occurred during the first 10-year of follow-up. Therefore, we set up a grid of landmark time points from 0 to 8 years and chose the prediction interval of width $w = 2$. To fit a proportional baselines landmark PSH (PBLM-PSH) supermodel to this dataset, we took 41 equally space landmark points s ($0 \leq s \leq 8$) and set the basis functions for $\boldsymbol{\beta}_{LM}(s)$ and $\boldsymbol{\gamma}(s)$ as $\boldsymbol{\beta}_{LM}(s) = \theta_1 + \theta_2 s + \theta_3 s^2$ and $\boldsymbol{\gamma}(s) = \eta_1 s + \eta_2 s^2$, respectively. The frequency of LRR in each of the landmark sub-dataset for $s = 0, 1, \ldots, 8$ years are shown in Figure 7(c) for illustration.

We began our analysis by selecting covariates which effects are dependent on the landmark points using the backward selection procedure. We tested the landmark-covariate interactions for each covariate using the Wald test based on the robust covariance matrix of the estimated coefficients from the PBLM-PSH supermodel. We found that only the effect of tumor grade (poor vs. well) are significantly dependent on the landmark points ($\chi^2_{[2]} = 6.76, p = 0.034$). Effects of other covariates do not deviate significantly from time-invariant effects. The effect of treatment TAM+MF is not significantly different from that of the treatment TAM. The effects of well and moderate tumor grade on the incident LRR are almost undifferentiable. The estimated log subdistribution hazard ratio and its corresponding robust standard error for a given prognostic factor are reported in Table 5. Multivariate Wald test for the baseline parameters (η_1, η_2) is significant ($\chi^2_{[2]} = 8.59, p = 0.014$), indicating that the baseline subdistribution hazard is also dependent on the choice of landmark points.

Figure 8(a) shows the comparison between the estimated log subdistribution hazards ratio and the associated pointwise 95% confidence intervals for patients with poorly differentiated tumor and those with well differentiated tumor adjusting for other covariates. As shown in the figure, the effect of tumor grade (poor vs. well) decreases over the landmark time from $s = 0$ on, and becomes stable after $s = 5$ years. The subdistribution hazard ratio

of poor to well tumor grade changes from 3.58(= exp(1.28)) at the beginning of the follow-up to 0.74(= exp(−0.30)) at year 5. The risk effect of poor tumor grade is diminishing over time of landmarking. This result is in line with many breast cancer studies which reported an attenuated prognostic effect of tumor grade as follow-up progressed. We also graphically examined the estimated landmark-dependent effect of tumor grade and the 95% confidence intervals obtained from the landmark PSH model (figure not shown). We found that the curves from these two different landmark PSH models are almost identical, which indicates that the landmark PSH supermodel can provide a good smoothing on the landmark effect. Figure 8(b) and 8(c) demonstrate the estimated cumulative baseline subdistribution hazard $\hat{\Lambda}_{10}(t)$ and the estimated landmark effect on the baseline subdistribution hazard. Along with the increase of s, $\exp(\hat{\boldsymbol{\gamma}}(s))$ rises rapidly at the beginning then slows down, and decreases in the right tail. The right-tail decreasing could be explained by that patients with better prognosis are more likely to be included in the landmark subset for the landmark time points close to 8 years.

For each level of tumor grades (poor, moderate, and well) and for each treatment group (TAM, TAM+MF, or TAM+CMF), Figure 9 depicts the predicted dynamic 2-year fixed width cumulative incidences of LRR for patients younger than 50 years old with tumor larger than 2cm and were receiving L+XRT. In all three treatment groups, we found similar patterns of dynamic predictive cumulative incidence over the landmark time changing from 0 to 8 years between the moderately differentiated tumor group and the well differentiated tumor group. In contrast, patients with poorly differentiated tumor grade had apparently higher risks of LRR for the subsequent 2 years of the given landmark time (less than 3.5 years from baseline), especially for those who received TAM or TAM+MF. There is a higher risk of LRR shown at the beginning of follow-up for patients with poorly differentiated tumor grade; yet if they survive for about one and half years, their 2-year cumulative incidence of LRR decreased dramatically. After 3.5 years from the baseline, patients with poorly differentiated tumor grade had similar 2-year risk of LRR as those with moderate and well differentiated tumor grades. This indicates that the effect of tumor grade is insignificant for those patients who are still alive or who have not failed from any cause up to the landmark points that are after 3.5 years.

Table 5: Estimated regression parameters of the proportional baselines landmark PSH supermodel for locoregional recurrence.

Covariate	Time function	Parameter estimate	Robust standard error
TAM + MF vs. TAM	Constant	-0.258	0.222
TAM + CMF vs. TAM	Constant	-1.305	0.307
Age (≥ 50 vs. < 50)	Constant	-0.780	0.210
Mastectomy vs. L + XRT	Constant	-0.521	0.210
Clinical tumor size (> 2 vs. ≤ 2 cm)	Constant	0.620	0.213
Tumor grade (moderate vs. well)	Constant	-0.216	0.251
Tumor grade (poor vs. well)	Constant	1.275	0.469
	s	-0.515	0.300
	s^2	0.040	0.038
Baseline parameters			
η_1	s	0.229	0.121
η_2	s^2	-0.017	0.013

3.5.2 Application 2

In this application example, we used the same data set as the one used in previous Section 3.5.1. But, at this time, the main event of interest is distant metastasis, and death is competing events. For early stage breast cancer patients who received surgery, development of locoregional recurrence (LRR) is an important prognostic clinical event affecting the risk of distant metastasis. Therefore, we treat LRR as intermediate clinical event to dynamically predict the risk of distant metastasis within the subsequent 3-years for a breast cancer patient, based on her LRR status measured during follow-up and other prognostic covariates measured at baseline. We also compared the dynamic 3-year fixed width probabilities of distant metastasis and death based on a patients LRR history.

Among the 2,272 clinically eligible patients followed up, 241 developed distant metastasis, 127 died due to other causes before distant metastasis could occur, and the remaining 1,904 were censored. In this data, 17.8% of the patients developed LRR before progressing to distant metastasis; but only 7.1% patients experienced LRR before death. Figure 10(a) shows that both the estimated cumulative incident distant metastasis and the estimated

mortality rates are stable after 13 years (i.e., few events occurred). Figure 10(b) depicts the estimated distribution of LRR. Only a few random censoring events occurred during the first 10 years of follow-up (figure not shown). We chose the range of landmark points from 0 to 10 years and prediction window with a fixed width of 3-years. To fit a PBLM-PSH supermodel to this dataset, we took 51 equally spaced landmark points s ($0 \leq s \leq 10$) and set the basis functions for $\boldsymbol{\beta}_{LM}(s)$ and $\boldsymbol{\gamma}(s)$ as $\boldsymbol{\beta}_{LM}(s) = \theta_1 + \theta_2 s + \theta_3 s^2$ and $\boldsymbol{\gamma}(s) = \eta_1 s + \eta_2 s^2$, respectively. The frequencies of distant metastasis and death in each of the landmark sub-dataset for $s = 0, 1, \ldots, 10$ years are shown in Figure 10(c).

We began our analysis using the backward selection procedure to select those covariates the effects of which were dependent on the landmark points. We tested the landmark-covariate interactions for each covariate via the Wald test based on the robust covariance matrix of the estimated coefficients from the PBLM-PSH supermodel. We found that for distant metastasis, the effects of TAM+MR and age are significantly dependent on the landmark points, whereas for death, only the effect of LRR status is dependent on the landmark points. The estimated coefficient and the corresponding robust standard error for a given prognostic factor are summarized in Table 6. Multivariate Wald tests for the baseline parameters (η_1, η_2) are significant for both distant metastasis and death, indicating that the baseline subdistribution hazard also depends on the choice of landmark points.

For different locoregional recurrence status (no LRR developed over the course of study, with LRR occurred at 3, 5, and 7 years) and for each treatment group (TAM, TAM+MF, or TAM+CMF), Figure 11 depicts the predicted dynamic 3-year fixed width cumulative incidences and the associated bootstrap 95% confidence intervals of distant metastasis and death for a patient 50 years old or younger with poorly differentiated tumor, size larger than 2cm and were receiving L+XRT. If the patient did not have a LRR, her risk of having distant metastasis within the subsequent 3 years is very close to the risk of death for any treatment group. However, if the patient had a LRR, she had a much higher risk of distant metastasis as compared to the risk of death, especially for TAM+MF treatment group. Similar results were obtained for patients who experienced LRR at 3, 5, and 7 years, respectively.

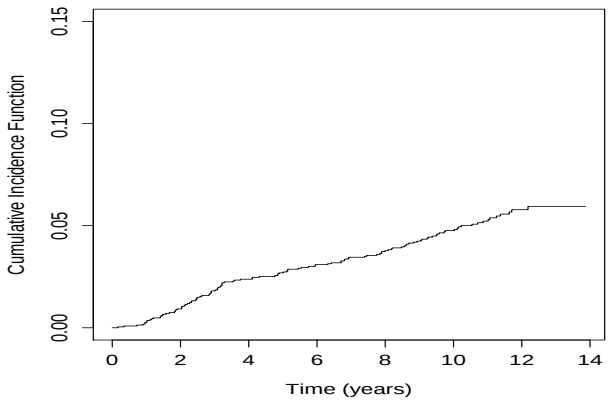
3.6 DISCUSSION

In this study, we developed dynamic predictive models for data containing competing risks by extending the landmark approach to the Fine-Gray PSH model. The resulting landmark PSH models can be used to directly predict the dynamic cumulative incidences of failure for a specific cause within a given prediction window of a fixed width by incorporating all available information updated up to this landmark time under the condition that the patient has not failed at the landmark time.

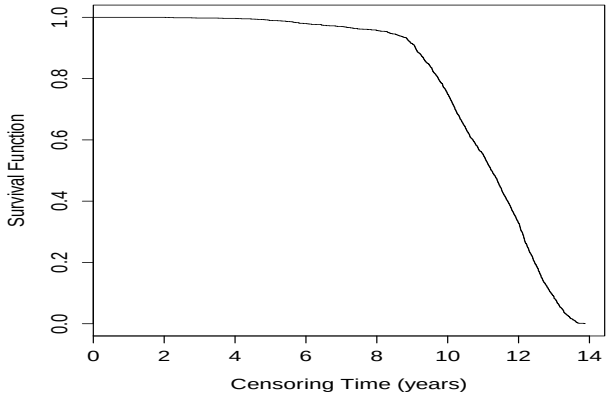
Our proposed models have several advantages over currently available methods in predicting conditional CIFs. The first is that our landmark PSH models can provide accurate estimates even when a covariate effect is time-varying (a violation of the PSH assumption), whereas the model developed by Cortese et al. (2013) [5], which also incorporates the landmark approach into the Fine-Gray model, will not give an unbiased estimate if a covariate in the model violates the PSH assumption. The second advantage is that unlike the standard Fine-Gray PSH model which does not allow the use of internal time-dependent covariate for predicting the CIF [2, 14], our proposed landmark PSH models can incorporate both internal and external time-dependent information through a simple model form without simultaneously modeling the covariate changing process and the time-to-event outcome process. The landmark approach can provide simpler explicit form of estimates and it is much easier to incorporate time-dependent covariates as compared to the multistate models and joint models which are more complicated and prone to overfitting. Furthermore, our landmark PSH supermodel could be more straightforward and simpler in implementation using the existing software packages. Being different from the landmark supermodel based on the cause-specific hazards [19], our landmark PSH supermodel predicts the conditional CIFs in one step and provides a direct interpretation of predictive probabilities. In comparison with the landmark supermodel based on the pseudo-observations [20], our landmark PSH supermodel is simpler in computation, whereas the GEE-based method used in pseudo-observations would have convergence issues for large sample size especially when dealing with many landmark points of interest.

Through simulations, we evaluated the prediction performance of our proposed models and compared them with other existing methods. We determined how closely the estimated conditional CIFs would approximate the true probabilities for our models, which is the first of its kind done in this area of study. We further utilized time-dependent Brier score to assess a model's discrimination and calibration capabilities simultaneously. The simulation results showed that our models performed well in prediction even when the PSH assumption was violated. Although some other studies applied time-dependent ROCs to evaluate their prediction models [23] we did not choose the same approach because the ROCs and AUCs can only be used to assess the capability of discrimination not calibration. In the future, we will use ROC-based method as a tool to evaluate the discriminative accuracy of the marker for a specific marker of interest.

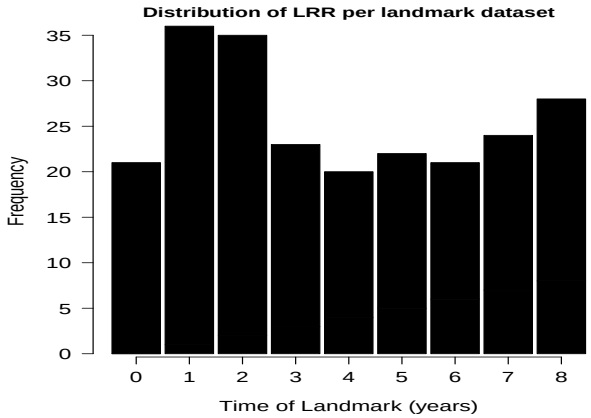
In application studies, we found that the baseline conditional subdistribution hazard was significantly dependent on the landmark s through $\exp(\boldsymbol{\gamma}(s))$. When applying PBLM-PSH supermodel to other data, it is possible that the baseline subdistribution hazard is independent of s. In such case, we suggest that one keeps $\exp(\boldsymbol{\gamma}(s))$ in the model to maintain coherence; more discussions on this topic can be found in van Houwelingen (2007) [31]. If there are too many covariates significantly dependent on the landmark time point, it is better to reduce the dimension by combining covariates into a prognostic index before fitting the landmark PSH supermodel.



(a) Nonparametric estimate of the cumulative incidence of locoregional recurrence (LRR).

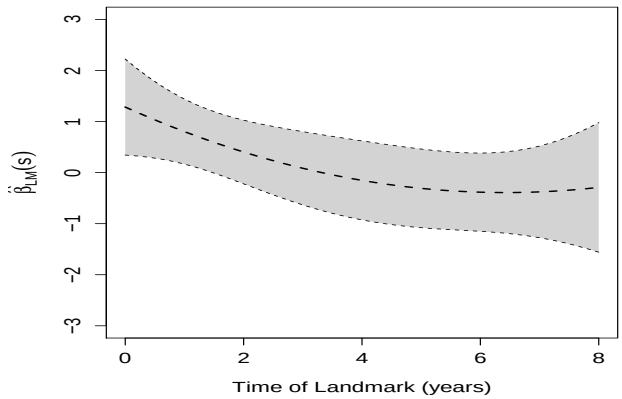


(b) Kaplan-Meier estimate of the censoring distribution.

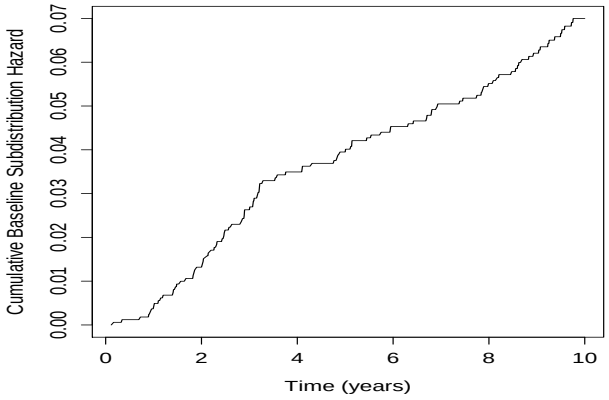


(c) Frequencies of LRR for each of the specified landmark data sets.

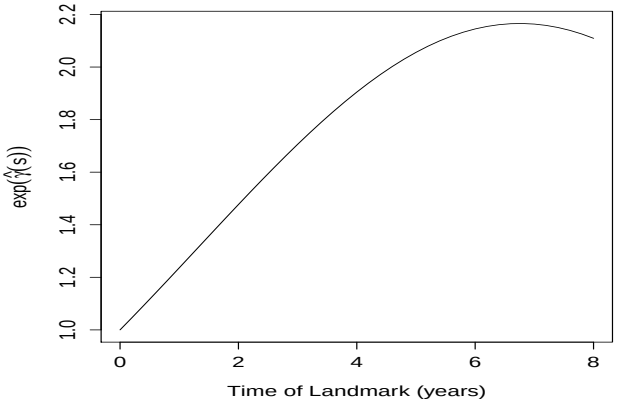
Figure 7: Descriptive analysis of the NSABP B-20 data.



(a) Estimated regression coefficients $\boldsymbol{\beta}_{LM}(s)$ of tumor grade (poor vs. well) and associated pointwise 95% confidence intervals.

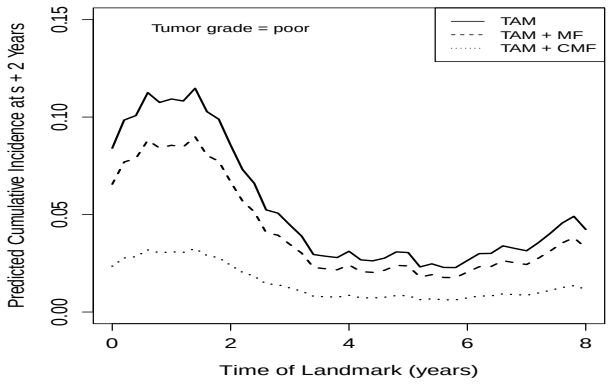


(b) Estimated cumulative baseline subdistribution hazards $\hat{\Lambda}_{10}(t)$.

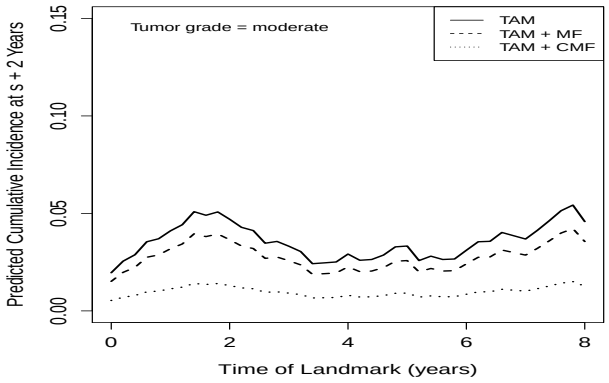


(c) Estimated $\boldsymbol{\gamma}(s)$ on exponential scale.

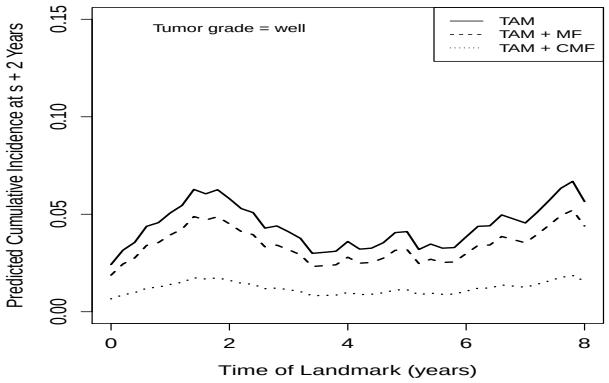
Figure 8: Regression results from the PBLM-PSH supermodel.



(a) Poorly differentiated tumor grade



(b) Moderate differentiated tumor grade



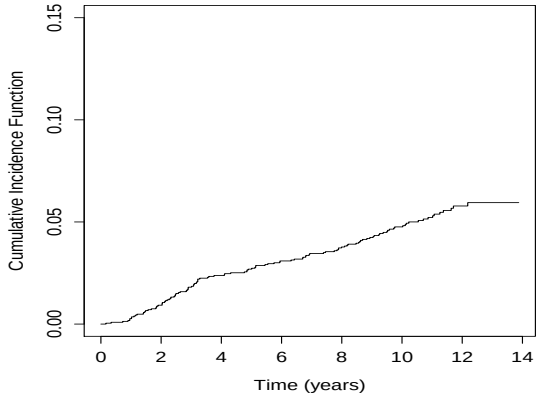
(c) Well differentiated tumor grade

Figure 9: The esimated 2-year fixed width predictive cumulative incidences of locoregional recurrence for patients younger than 50-years old with tumor larger than 2cm and treated with L + XRT, in each of the treatment groups (TAM, TAM + MF, TAM + CMF) and different levels of tumor grade.

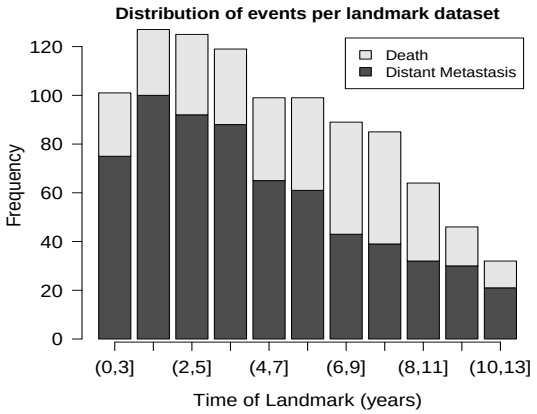
Table 6: Estimated regression parameters of the proportional baseline landmark PSH super-model for distant metastasis and death.

		Distant metastasis		Death	
Covariate		$\hat{\boldsymbol{\beta}}$	SE($\hat{\boldsymbol{\beta}}$)	$\hat{\boldsymbol{\beta}}$	SE($\hat{\boldsymbol{\beta}}$)
TAM + MF vs. TAM	Constant	-1.091	0.331	-0.312	0.238
	s	0.482	0.160		
	s^2	-0.050	0.017		
TAM + CMF vs. TAM	Constant	-0.470	0.170	0.160	0.219
Age (≥ 50 vs. < 50)	Constant	-0.540	0.257	1.057	0.218
	s	0.407	0.135		
	s^2	-0.035	0.015		
Mastectomy vs. L + XRT	Constant	0.424	0.144	-0.115	0.189
Clinical tumor size (> 2 vs. ≤ 2 cm)	Constant	0.284	0.141	0.334	0.196
Tumor grade (moderate vs. well)	Constant	0.221	0.179	0.104	0.221
Tumor grade (poor vs. well)	Constant	0.752	0.189	0.124	0.258
Local-regional recurrence status	Constant	2.315	0.237	3.575	1.230
	s			-0.492	0.582
	s^2			0.008	0.056
Baseline parameters					
η_1	s	-0.400	0.088	-0.001	0.014
η_2	s^2	0.035	0.009	0.002	0.001

(a) Nonparametric estimate of the cumulative incidence of distant metastasis and death.



(b) Nonparametric estimate of the cumulative incidence of locoregional recurrence



(c) Frequencies of distant metastasis and death for each of the specified landmark data sets.

Figure 10: Descriptive analysis of the B-20 data in application 2

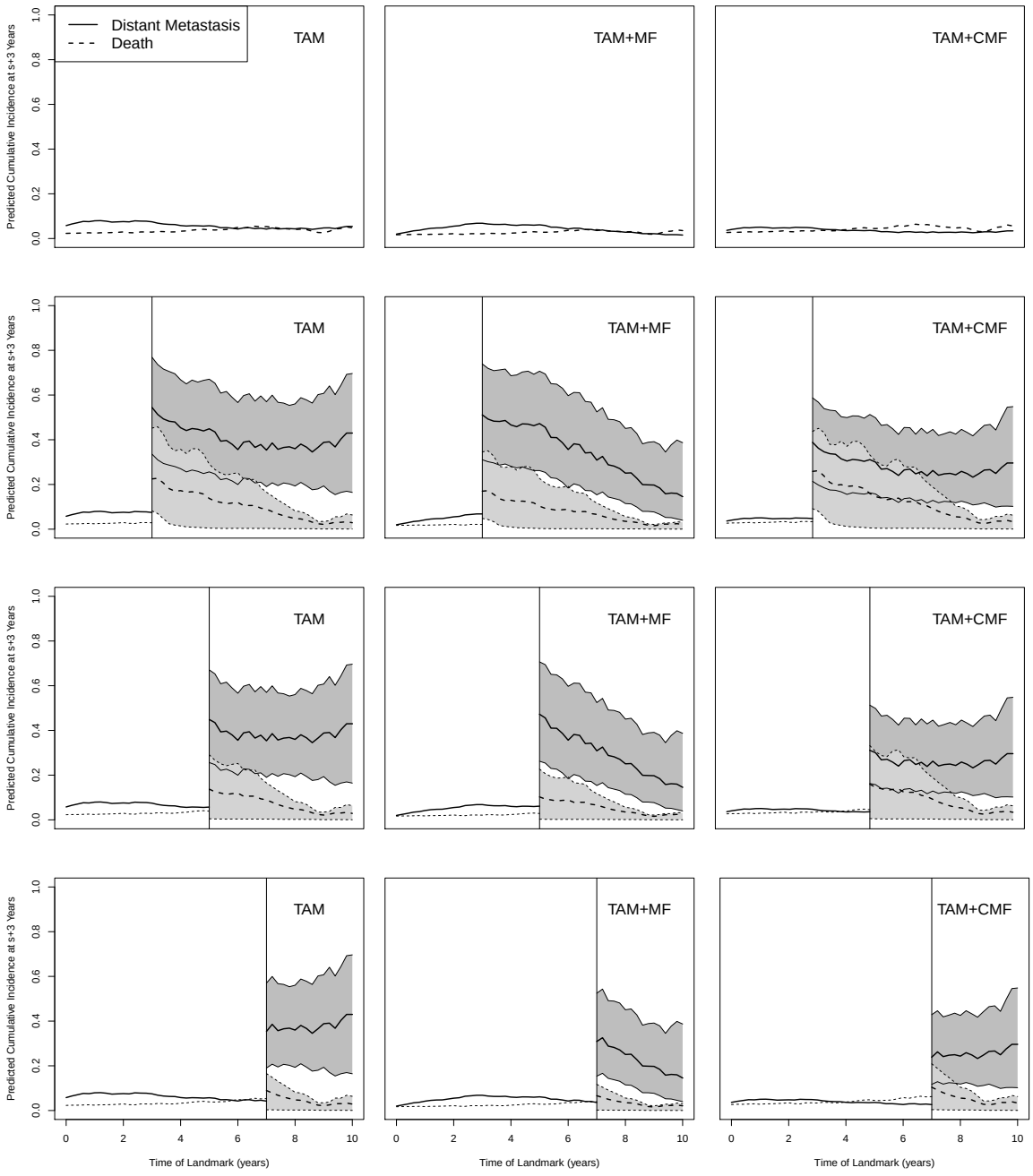


Figure 11: The predicted 3-year fixed width cumulative incidences of distant metastasis (solid lines), death (dashed lines) and associated 95% confidence intervals (shaded areas) for different landmark time points, for a patient younger than 50-tears old with poor tumor grade, tumor larger than 2cm and treated with lumpectomy, for each of the treatment groups and with locoregional recurrence occurred at none, 3 years, 5 years and 7 years.

APPENDIX A

ANDERSEN-GILL-TYPE CONDITIONS

Following Xu and O'Quigley (2000) [35], we define $w(t) = S(t)/E\{I(X \geq t)\}$, which is a non-negative bounded function uniformly in t. Here, we assume that there is enough information on $F_1(t)$ in the tails in the presence of censoring.

Andersen-Gill-type conditions:

A. (Finite interval). $\int_0^1 \lambda_{10}(t)dt < \infty$.

B. (Asymptotic stability). There exists a neighbourhood $\mathcal{B}$ of β such that 0 and $\beta(t)$, in which $t \in [0,1]$, are in the interior of $\mathcal{B}$, and for $r = 0, 1, 2$

$$\sup_{t\in[0,1]} |\boldsymbol{S}^{(r)}(\boldsymbol{\beta}(t), t) - \boldsymbol{s}^{(r)}(\boldsymbol{\beta}(t), t)| \xrightarrow{P} 0,$$

$$\sup_{t\in[0,1]} |\boldsymbol{S}^{(r)}(\boldsymbol{\beta}, t) - \boldsymbol{s}^{(r)}(\boldsymbol{\beta}, t)| \xrightarrow{P} 0, \quad \sup_{t\in[0,1]} |nW(t) - w(t)| \xrightarrow{P} 0.$$

C. (Lindeberg condition). There exists $\delta > 0$ such that

$$n^{-1/2} \sup_{i,t} |\boldsymbol{Z}_i| Y_i(t) I\{\boldsymbol{Z}_i^T \boldsymbol{\beta}(t) > -\delta |\boldsymbol{Z}_i|\} \xrightarrow{P} 0,$$

$$n^{1/2} \sup_{i,t} W(t) |\boldsymbol{Z}_i| Y_i(t) I\{\boldsymbol{Z}_i^T \boldsymbol{\beta}(t) > -\delta n W(t) |\boldsymbol{Z}_i|\} \xrightarrow{P} 0.$$

D. (Asymptotic regularity conditions). All (deterministic) functions in $\mathcal{B}$ are uniformly continuous in $t \in [0,1]$; $\boldsymbol{s}^{(r)}$, $r = 0, 1, 2$, are continuous functions of $\boldsymbol{\beta} \in \mathcal{B}$, and are bounded on $\mathcal{B} \times [0,1]$; $\boldsymbol{s}^{(0)}(\boldsymbol{\beta}(t), t)$ and $\boldsymbol{s}^{(0)}(\boldsymbol{\beta}, t)$ are bounded away from zero. For all $\boldsymbol{\beta} \in \mathcal{B}$, $t \in [0,1]$, $\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t) = \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{s}^{(0)}(\boldsymbol{\beta}, t)$, $\boldsymbol{s}^{(2)}(\boldsymbol{\beta}, t) = \frac{\partial^2}{\partial \boldsymbol{\beta}^2} \boldsymbol{s}^{(0)}(\boldsymbol{\beta}, t)$.

APPENDIX B

DERIVATION OF THE APPROXIMATION (2.12) IN CHAPTER 2

Let $\hat{\lambda}_{10}(\boldsymbol{\beta}(t), t)$ be the Breslow estimator of the baseline subdistribution hazard under a coorect model specification, and $\hat{\lambda}_{10}(\boldsymbol{\beta}(t), t)$ is consistent for the true value $\lambda_{10}(t)$. The $\hat{\lambda}_{10}(t)$ is the Breslow estimator calculated from the model using arbitrary $\boldsymbol{\beta}$, and $\hat{\lambda}_{10}(t)$ converges in probability to a limiting value $\lambda^*_{10}(t)$.

Following the same derivations in van Houwelingen (2007) [31], we have

$$\frac{\hat{\lambda}_{10}(\boldsymbol{\beta}(t), t)}{\hat{\lambda}_{10}(t)} = \frac{\boldsymbol{S}^{(0)}(t, \boldsymbol{\beta})}{\boldsymbol{S}^{(0)}(t, \boldsymbol{\beta}(t))} \xrightarrow{p} \frac{\boldsymbol{s}^{(0)}(t, \boldsymbol{\beta})}{\boldsymbol{s}^{(0)}(t, \boldsymbol{\beta}(t))} = \frac{\boldsymbol{s}^{(0)}_*(t, \boldsymbol{\beta})}{\boldsymbol{s}^{(0)}_*(t, \boldsymbol{\beta}(t))}.$$

As in the deviation of equation (2.5), under random right censoring, it can be shown that

$$\begin{aligned} \frac{\boldsymbol{s}^{(0)}_*(t, \boldsymbol{\beta})}{\boldsymbol{s}^{(0)}_*(t, \boldsymbol{\beta}(t))} &= E\left[\exp\{\boldsymbol{Z}^T(\boldsymbol{\beta}^*_{hor} - \boldsymbol{\beta}(t))\} | T^* = t\right] \\ &\approx \exp\left[E(\boldsymbol{Z}|T^* = t)^T\{\boldsymbol{\beta}^*_{hor} - \boldsymbol{\beta}(t)\}\right]. \end{aligned} \tag{B.1}$$

Under the assumption that the $\boldsymbol{\beta}(t)$ does not vary too much over time, the approximation (2.12) can be derived from (B.1), given by

$$\lambda^*_{10}(t) \approx \lambda_{10}(t) \exp[E(\boldsymbol{Z}|T^* = t)^T\{\boldsymbol{\beta}(t) - \boldsymbol{\beta}^*_{hor}\}].$$

BIBLIOGRAPHY

[1] J. R. Anderson, K. C. Cain, and R. D. Gelber. Analysis of survival by tumor response. *Journal of Clinical Oncology*, 1:710–719, 1983.

[2] J. Beyersmann and M. Schumacher. Time dependent covariates in the proportional subdistribution. *Biostatistics*, 9:765–776, 2008.

[3] C. Proust-Lima and J. M. G. Taylor Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics*, 10:535–549, 2009.

[4] E. Christensen, P. Schlichting, P. K. Andersen, L. Fauerholdt, G. Schou, B. V. Pedersen, E. Juhl, H. Poulsen, and N. Tygstrup. Updating prognosis and therapeutic effect evaluation in cirrhosis with cox multiple-regression model for time-dependent variables. *Scandinavian Journal of Gastroenterology*, 21(2):163–174, 1983.

[5] G. Cortese, T. A. Gerds, and P. K. Andersen. Comparing predictions among competing risks models with time-dependent covariates. *Statistics in Medicine*, 32:3089–3101, 2013.

[6] P. K. Cortese, G.and Andersen. Competing risks and time-dependent covariates. *Biometrical Journal*, 52:138–158, 2010.

[7] U. Dafni. Landmark analysis at the 25-year landmark point. *Circulation: Cardiovascular Quality and Outcomes*, 4:363–371, 2011.

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

[9] M. Fiocco, H. Putter, and van Houwelingen J. C. Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models. *Statistics in Medicine*, 27(21):4340–4358, 2008.

[10] B. Fisher, J. Costantino, C. Redmond, R. Poisson, D. Bowman, J. Couture, N. V. Dimitrov, N. Wolmark, D. L. Wickerham, E. R. Fisher, R. Margolese, A. Robidoux, H. Shibata, J. Terz, A.H.G. Paterson, M. Feldman, Farrar W., J. Evans, L. Lickley, and M. Ketner. A randomized clinical trial evaluating tamoxifen in the treatment of

patients with node-negative breast cancer who have estrogen receptor- positive tumors. *The New England Journal of Medicine*, 320:479–484, 1989.

[11] T. A. Gerds and M. Schumacher. Consistent estimation of the expected brier score in general models with right-censored event times. *Biometrical Journal*, 48:1029–1040, 2006.

[12] R. B. Geskus. Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring. *Biometrics*, 67:39–49, 2011.

[13] J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. New York: Wiley, 1980.

[14] J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. New York: Wiley, 2002.

[15] J. P. Klein and P. K. Andersen. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrika*, 61:223–229, 2005.

[16] D. Y. Lin. Goodness-of-fit analysis for the cox regression model based on a class of parameter estimators. *Journal of the American Statistical Association*, 86:725–728, 1991.

[17] E. B. Madsen, P. Hougaard, and E. Gilpin. Dynamic evaluation of prognosis from time-dependent variables in acute myocardial-infraction. *American Journal of Cardiology*, 51(10):1579–1583, 2013.

[18] A. Mauguen, B. Rachet, Mathoulin-Pélissier, G. MacGrogan, A. Laurent, and V. Rondeau. Dynamic prediction of risk of death using history of cancer recurences in joint frailty models. *Statistics in Medicine*, 32(30):5366–5380, 2013.

[19] M. A. Nicolaie, J. C. van Houwelingen, T. M. de Witte, and H. Putter. Dynamic prediction in competing risks by landmarking. *Statistics in Medicine*, 32:2031–2047, 2013.

[20] M. A. Nicolaie, J. C. van Houwelingen, T. M. de Witte, and H. Putter. Dynamic pseudo-observations: A robust approach to dynamic prediction in competing risks. *Biometrics*, 2013.

[21] M. Pepe and M. Mori. Kaplan-meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine*, 12:737–751, 1993.

[22] M. Fiocco , H. Putter, and R. B. Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, 26:2389–2430, 2007.

[23] D. Rizopoulos. Dynamic prediction and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67:819–829, 2011.

[24] J. Robins, A. Rotnitzky, and L. P. Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89:846–866, 1994.

[25] J. M. Robins and A. Rotnitzky. Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology, Methodological issues*, pages 297–331, 1992.

[26] T. H. Scheike, M. J. Zhang, and T. A. Gerds. Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, 10:1–16, 2008.

[27] R. Schoop, J. Beyersmann, M. Schumacher, and H. Binder. Quantifying the predictive accuracy of time-to-event models in the presence of competing risks. *Biometrical Journal*, 53:88–112, 2011.

[28] R. Schoop, E. Graf, and M. Schumacher. Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. *Bometrics*, 64:603–610, 2008.

[29] C. A. Struthers and J. D. Kalbfleisch. Misspecified proportional hazard models. *Biometrika*, 73:363–369, 1986.

[30] L. Q. Sun, J. X. Liu, J. G. Sun, and M. J. Zhang. Modelling the subdistribution of a competing risk. *Statistica Sinica*, 16:1367–1385, 2006.

[31] H. C. van Houwelingen. Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*, 34:70–85, 2007.

[32] H. C. van Houwelingen and H. Putter. Dynamic prediction by landmarking as an alternative for multi-state modeling: An application to acute lymphoid leukemia data. *Lifetime Data Analysis*, 14:447–463, 2008.

[33] H. C. van Houwelingen and H. Putter. *Dynamic prediction in clinical survival analysis*. CRC Press, 2012.

[34] R. H. Xu. Inference for the proportional hazards model. Ph.D. Thesis. University of California, San Diego., 1996.

[35] R. H. Xu and J. O'Quigley. Estimating average regression effect under non-proportional hazards. *Biostatistics*, 1:423–439, 2000.

[36] B. Q. Zhou, J. P. Fine, and G. Laird. Goodness-of-fit test for proportional subdistribution hazards model. *Statistics in Medicine*, 32:3804–3811, 2013.

[37] B. Q. Zhou, A. Latouche, V. Rocha, and J. P. Fine. Competing risks regression for stratified data. *Biometrics*, 67:661–670, 2011.