PARAMETRIC REGRESSION ANALYSIS AND DISCRIMINATION ANALYSIS OF COMPETING RISKS DATA

by

Haiwen Shi

B.E., Beijing Polytechnic University, 1993

M.S., University of Pittsburgh, 2004

M.A., University of Pittsburgh, 2009

Submitted to the Graduate Faculty of

the Department of Statistics in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

THE DIETRICH SCHOOL OF ARTS & SCIENCES

This dissertation was presented

by

Haiwen Shi

It was defended on

November 12th 2012

and approved by

Yu Cheng, Assistant Professor, Department of Statistics

Allan R. Sampson, Professor, Department of Statistics

Leon J. Gleser, Professor, Department of Statistics

Jong-Hyeon Jeong, Associate Professor, Department of Biostatistics

Dissertation Director: Yu Cheng, Assistant Professor, Department of Statistics

Yu Cheng, PhD

PARAMETRIC REGRESSION ANALYSIS AND DISCRIMINATION ANALYSIS OF COMPETING RISKS DATA

Haiwen Shi, PhD

University of Pittsburgh, 2012

This dissertation contains two parts focusing on regression analysis and diagnostic accuracy analysis of competing risks data. In the first part, we propose a parametric regression model for the cumulative incidence functions (CIFs) commonly used for competing risks data. The model adopts a modified logistic model as the baseline CIF and a generalized odds-rate model for covariate effects, and it explicitly takes into account the constraint that a subject with any given prognostic factors should eventually fail from one of the causes such that the asymptotes of the CIFs should add up to one. We hence model the CIF from the primary cause assuming the generalized odds-rate transformation and the modified logistic function as the baseline CIF. Under the additivity constraint, the covariate effects on the competing cause are modeled by a function of the asymptote of the baseline distribution and the covariate effects on the primary cause. The inference procedure is straightforward by using standard maximum likelihood theory. We demonstrate desirable finite-sample performance of our model by simulation studies in comparison with existing methods. Its practical utility is illustrated in an analysis of a breast cancer data set to assess the treatment effect of tamoxifen on breast cancer recurrence that is subject to dependent censoring by second primary cancers and deaths.

Diagnostic accuracy studies progressed in the past decade to involve complicated survival outcomes beyond the traditional dichotomous outcome. Another recent advance in diagnostic medicine is the appearance of novel measures for accuracy improvement due to the addition

of new markers . In the second part of this dissertation, we intend to integrate these two evolving areas and contribute a discussion on assessing accuracy improvement for censored survival outcomes. Furthermore, we consider competing-risk censoring in addition to the usual independent censoring and provide statistical procedures with inference details. In particular, we consider fitting regression models based on the CIF for the primary event. Parallel estimators are proposed using inverse probability weighting or based on the bivariate CIF. Both estimators perform very well in simulation studies and in an application to another breast cancer study.

Keywords: Area under the receiver operating characteristic curve, Cause-specific hazard function, Competing-risk censoring, Cumulative incidence function, Diagnostic and prognostic accuracy improvement, Integrated discrimination improvement, Long-term incidence, Modified three-parameter logistic model, Net reclassification improvement, Parametric modeling.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Competing risks has been an active research area in survival analysis. In practice, when there are composite outcomes, it is common to see that an event of interest is competing-risk censored by other events. For example, in the first breast cancer study considered in this dissertation, the primary event is local cancer recurrence that may be dependently censored by other competing events such as death or distant metastasis. In the second breast cancer study that motivated our work, the event of interest is metastasis which may be dependently censored by death. In a competing-risk setting, some standard quantities such as the survival function may not be well defined if the removal of competing events is not conceptually realistic. Instead, the cumulative incidence function (CIF) has been an established quantity to describe cumulative risks of an event of interest over time. Since the limit of a CIF is less than one, CIF is often called a subdistribution.

Naturally all subdistributions should add up to one when time goes to infinity, as a subject should eventually fail from one of the events. We refer to this as an additivity constraint among CIFs from all causes. In the first part of this dissertation, our primary goal is to investigate how covariates affect the CIFs in a regression setting. However, after carefully reviewing existing methods in the literature, we noticed that the additivity constraint had been ignored in some commonly used regression models for CIFs. For example, Fine and Gray (1999)'s and Scheike et al. (2008)'s semi-parametric models only work on one event each time. They did not jointly model all competing risks simultaneously so they did not clearly consider this constraint. Jeong and Fine (2007) also developed a new parametric model on CIFs without considering the additivity constraint. Hence, we propose a parametric regression model that is based on Jeong and Fine (2007)'s framework but specifically

takes into account this additivity constraint. Moreover, we update their model by introducing a more flexible baseline function – the modified three-parameter logistic model of Cheng (2009). In Chapter 2, we will give futher justification for our model specifications and will present extensive simulation studies. We will show that our parametric model has good practical performance, and more importantly, what the consequences of ignoring the additivity constraint by other methods would be in comparison with our model.

The focus of the second part of this dissertation is on evaluating the additional contributions of new markers in a regression model that already includes other conventional predictors. We consider an event outcome that is subject to random or competing-risk censoring. Traditionally, the area under receiver operating characteristic curve (AUC) is used to measure the added value of new markers. However, researchers have documented the limitation of AUC. For example, Pencina et al. (2008) shed light on a study in which AUC is not sensitive enough to capturing the additional contribution of HDL cholesterol in predicting the risk of cardiovascular disease over other covariates, even though it is a significant independent predictor. Meanwhile, they introduced a new criterion called integrated discrimination improvement (IDI). They demonstrated that it is more sensitive than AUC. Uno et al. (2009) developed Pencina's IDI by considering the distribution functions of the difference in predicted risks with and without new markers, conditional on whether or not patients have experienced the event of interest by time t. They showed the IDI is the area between the two conditional distribution functions and proposed some estimators for them using inverse weighting method. In Chapter 3, we will propose an alternative estimating method for the distribution functions based on Dabrowska (1988)'s bivariate survival function estimators. Furthermore, we will extend these estimators to a competing-risk setting and focus on predicting the risk of interest based on its cumulative incidence function (CIF). Parallel estimators are developed to estimate the conditional distribution functions based on the nonparametric estimator of the bivariate CIF (Cheng et al., 2007). The two estimating methods are compared through extensive simulation studies. We will also apply these estimators to the second breast cancer study to evaluate the added value of gene score, newly derived from microarray gene expression data, in predicting the risk of metastases.

2.0 CONSTRAINED PARAMETRIC MODEL FOR SIMULTANEOUS INFERENCE OF TWO CUMULATIVE INCIDENCE FUNCTIONS

2.1 INTRODUCTION

In practice, Cox regression models (Cox, 1972) and accelerated failure time models (Wei et al., 1990; Jin et al., 2003) are commonly used to analyze covariate effects on time-to-event outcomes. Recently, practitioners have become more aware of competing-risk censoring (Koller et al., 2012), where the event of interest may be dependently censored by some competing events. Competing risks arise commonly in the analysis of composite endpoints. For example, a breast cancer dataset used later in our data example (NSABP B-14) involves treating breast cancer patients with a hormonal therapy in an adjuvant setting. Therefore, patients first have a surgical procedure to remove tumors, are treated with tamoxifen, and followed for many years. For the disease-free survival endpoint, patients can experience local, regional, or distant recurrence of breast cancer, second primary cancer other than breast, or death prior to any disease, whichever occurs first. In this setting, if we are interested in breast cancer recurrence, other events such as second primary cancers and death prior to any disease should be considered as competing events because the events of primary interest are precluded from being observed once the competing events occur.

Fine and Gray (1999) extended the Cox model to competing-risk settings and proposed a proportional hazard model for the hazard of a subdistribution. A perhaps more familiar name for a subdistribution is cumulative incidence function (CIF), which describes the cumulative probability of a target event occurring up to a certain time point. The Fine and Gray (1999) model is appealing to practitioners as the CIF has a straightforward probability

interpretation, although the plateau of the CIF is less than one as a subject may fail from multiple causes. Regression analyses of univariate competing risks data based on CIFs have been well studied in the literature. Fine (2001) extended the standard log-linear regression model to competing-risk settings and proposed semiparametric regression models for the CIF. Klein and Andersen (2005) presented an alternative regression model based on pseudo-values of the CIF. Scheike et al. (2008) extended the Fine and Gray (1999) model to allow for time-varying coefficients which relax the proportional hazards assumption for hazards of CIFs. Scharfstein et al. (1998) investigated the efficiency of covariate coefficient estimators under the generalized odds-rate regression model, which was extended to the competing-risk setting by Jeong and Fine (2007), where they proposed a parametric regression model for the primary cause and competing cause CIFs given covariates by using two-parameter Gompertz models as baseline CIFs.

Without loss of generality, we consider in this dissertation only two causes $k = 1, 2$, one for the event of interest and the other for the competing event. As one subject will eventually fail from the event of interest or the competing event, the CIFs from both causes should add up to one as time goes to infinity. In our application to the breast cancer study, if we had followed all the patients long enough, we would have observed that each patient either had breast cancer recurrence, or died without breast cancer recurrence. This relationship is well observed in nonparametric and parametric estimations of CIFs without covariates. In the presence of covariates $\mathbf{z}$, a $P \times 1$ vector, we should also expect that

$$P(K = 1|\mathbf{z}) + P(K = 2|\mathbf{z}) = 1. \tag{2.1.1}$$

where K is a random variable representing the cause type. However, this additivity constraint has not been explicitly considered in previous regression analyses of CIFs. Fine and Gray (1999), Fine (2001), Klein and Andersen (2005) and Scheike et al. (2008) focused on modeling covariate effects on the cause 1 event only. If both causes 1 and 2 are of interest, one may run their models twice, one for each cause. However, it is not clear how to interpret the two sets of regression parameters when the two CIFs do not add up to one as time goes to

infinity. The same dilemma also applies to Jeong and Fine (2007), which jointly modeled both cause 1 and cause 2 events, assuming a generalized odds-rate model for each CIF.

Therefore, in this chapter we propose a parametric regression model that explicitly incorporates the constraint between the two CIFs. Gaynor et al. (1993) pointed out the difficulty in extrapolating the nonparametric estimates beyond the range of observed failure times. In contrast, a parametric model for the CIF provides an estimate of the upper-limit probability without extrapolation (Jeong and Fine, 2007; Cheng, 2009). Hence a parametric model is especially useful for handling the additivity constraint on two asymptotes. We adapt the parametric model proposed in Jeong and Fine (2007) in two ways. First, we replace the Gompertz model with a more flexible model for baseline CIFs. Though the Gompertz model has a nice proportional subdistribution hazards interpretation, it cannot model a sigmoidal CIF. Cheng (2009) developed a modified three-parameter logistic model for a one-sample CIF without covariates, and showed that it performs better than the Gompertz model, particularly when the CIF curves have a sigmoidal shape. More importantly, when we jointly model cause 1 and cause 2 CIFs, we incorporate the additivity constraint in (2.1.1) to the asymptotes for which existing regression methods fail to account.

By using a flexible baseline function and a generalized odds-rate model, our parametric model seems to enjoy flexibility similar to that of its semiparametric counterparts, as shown via simulations. Standard maximum likelihood theory is readily applied. The semiparametric models considered in Fine and Gray (1999) and Fine (2001) use inverse probabilities of censoring as the weights which are estimated based on data. Their estimates of regression coefficients may be biased if the censoring survival distribution is misspecified. In contrast, our parametric model does not require modeling for censoring times.

The rest of this chapter is organized as follows. We introduce our parametric model in the next section. Simulation studies are given in Section 2.3 to compare the performance of our parametric model with the Fine and Gray (1999) method, the semiparametric model by Scheike et al. (2008) and the parametric model in Jeong and Fine (2007). The methods are applied to breast cancer data in Section 2.4. We conclude the chapter with some remarks in Section 2.5.

2.2 METHOD

Let T and C be the event time and censoring time and $K = 1, 2$ be the random variable representing cause type. We observe $Y = \min(T, C)$ and $\eta = KI\{T < C\}$, where I is the indicator function. Fine and Gray (1999) and Fine (2001) assumed that

$$g_k\{F_k(t; \mathbf{z})\} = g_k\{F_{0k}(t)\} + \boldsymbol{\beta}_k^{'}\mathbf{z}, \tag{2.2.1}$$

where $F_k(t; \mathbf{z}), k = 1, 2$ are the cause k CIFs at time t given covariates $\mathbf{z}$; $F_{0k}(t)$ are the baseline CIFs for cause k, that is, $F_{0k}(t) = F_k(t; \mathbf{0})$; and g_k are some nondecreasing known functions, and may have different forms for the two causes. Klein (2006) proposed an additive model for CIFs by simply letting $g_k(x)$ be x, and estimated $\boldsymbol{\beta}_k$ under the constraint that the CIFs have to be between zero and one. The analysis will become more complicated when we jointly model two CIFs. The Fine and Gray method considered the complementary log-log transformation $g(u) = \log\{-\log(1 - u)\}$ which gives a proportional hazards interpretation for subdistribution hazards. In our application, we are interested in assessing the effect of tamoxifen as compared to placebo on cancer recurrence while controlling for patients' age and initial tumor sizes. In Section 2.4, we will show that the proportional subdistribution hazards assumption does not hold for age. Departures from the proportional hazards assumption have led to a more general model. Scheike et al. (2008) proposed a more general model

$$g_k\{F_k(t; \mathbf{z})\} = \mathbf{z}_1^{'}\boldsymbol{\alpha}_k(t) + h_k(\mathbf{z}_2, \boldsymbol{\gamma}_k, t),$$

for $k = 1, 2$, where g_k and h_k are known link functions, $\boldsymbol{\alpha}_k(t)$ are time-varying coefficients for a sub-vector of covariates $\mathbf{z}_1$ of dimension Q ($Q < P$), and $\mathbf{z}_2$ are the remaining covariates in $\mathbf{z}$ with time-independent coefficients $\boldsymbol{\gamma}_k$. This model allows for time-varying effects for some covariates and is more flexible than the Fine and Gray model. However, there is no guarantee that the two asymptotes would add up to one, if both causes 1 and 2 CIF satisfy the above models. We hence propose a generalized odds-rate model with the modified logistic function for baseline CIFs.

Cheng (2009) modified a three-parameter logistic function and provided a flexible parametric model to characterize the CIF. For cause $k = 1, 2$, the CIF is expressed as:

$$F_k(s; \psi_k) = \frac{p_k \exp\{b_k(s - c_k)\} - p_k \exp(-b_k c_k)}{1 + \exp\{b_k(s - c_k)\}}, \tag{2.2.2}$$

where $\psi_k = (b_k, c_k, p_k)$ is the vector of three parameters. The parameter p_k corresponds to the long-term probability of the cause k event, b_k describes how fast the CIF approaches its asymptote p_k, and c_k describes the "center" of the rising.

To avoid intensive computing which is common for most semiparametric or nonparametric regression models, we adopt flexible parametric forms for our regression model on the CIFs. Recently the generalized odds-rate model (Dabrowska and Doksum, 1998) was extended to competing-risk settings by Jeong and Fine (2007), where

$$g_k(u; \alpha_k) = \log[\{(1 - u)^{-\alpha_k} - 1\}/\alpha_k], 0 < \alpha_k < \infty. \tag{2.2.3}$$

Here, α_k is an extra parameter for each event type. As demonstrated in Cheng (2009), the modified three-parameter logistic model is a flexible function that characterizes the CIF. In contrast to the Gompertz model used in Jeong and Fine (2006, 2007), the model in (2.2.2) is especially useful to capture CIFs that have a sigmoidal shape. In this chapter, we hence assume that F_{0k} satisfies the parametric form given in (2.2.2). By adopting parametric forms for g_k and F_{0k} in (2.2.1), we sacrifice some flexibility. However, this parametric model is flexible enough to handle a variety of situations and will substantially reduce computing intensity.

Let $f_k(t; \mathbf{z}) = \dot{F}_k(t; \mathbf{z})$, $k = 1, 2$, where the superscript dot denotes a derivative. Similar to Jeong and Fine (2007), we consider the following likelihood function

$$\prod_{i=1}^{n} \left[\left\{ \prod_{k=1}^{2} f_k(y_i; \mathbf{z})^{I\{\delta_i = k\}} \right\} \left\{ 1 - \sum_{k=1}^{2} F_k(y_i; \mathbf{z}) \right\}^{I\{\delta_i = 0\}} \right]. \tag{2.2.4}$$

The cause 1 and cause 2 CIFs are both assumed to satisfy the regression models in (2.2.3) and to be fitted simultaneously based on the above full likelihood function. Ideally, we would like to maximize the full-likelihood under the constraint in (2.1.1). This turns out to be a difficult task. As an example, we consider the simple case $g(u) = \log\{-\log(1 - u)\}$.

Assuming the cause $k = 1, 2$ baseline CIFs satisfying (2.2.2) and the complementary log-log link function, we have

$$F_k(t; \mathbf{z}) = 1 - \left[1 - \frac{p_k \exp\{b_k(t - c_k)\} - p_k \exp(-b_k c_k)}{1 + \exp\{b_k(t - c_k)\}}\right]^{\exp(\boldsymbol{\beta}_k' \mathbf{z})}. \tag{2.2.5}$$

At baseline, $F_{01}(\infty) + F_{02}(\infty) = 1$ results in $p_2 = 1 - p_1$. For a subject with covariates $\mathbf{z}$,

$$F_1(\infty; \mathbf{z}) + F_2(\infty; \mathbf{z}) = 1 - (1 - p_1)^{\exp(\boldsymbol{\beta}_1' \mathbf{z})} + 1 - p_1^{\exp(\boldsymbol{\beta}_2' \mathbf{z})} = 1.$$

This gives $\exp(\boldsymbol{\beta}_2' \mathbf{z}) = \frac{1}{\log p_1} \log\{1 - (1 - p_1)^{\exp(\boldsymbol{\beta}_1' \mathbf{z})}\}$. That is, the covariate effects on the cause 2 event can be replaced by a function of the cause 1 covariate effects and the cause 1 baseline asymptote. Under the constraint of two asymptotes adding up to one, only the parameters for baseline logistic functions and the cause 1 covariate effects can vary freely. When there are more than one covariate in the model, we lose more than one degree of freedom in the maximization procedure with this single constraint. As commented in Grambauer et al. (2010), specifying proportional subdistribution hazards model for all CIFs "implies rather tedious algebraic complications." To handle this difficulty, we explicitly model the cause 1 CIF as in (2.2.3); for the competing cause, we simply adopt the modified logistic function as follows:

$$F_2(t; \mathbf{z}) = \frac{p_2(\mathbf{z})[\exp\{b_2(t - c_2)\} - \exp(-b_2 c_2)]}{1 + \exp\{b_2(t - c_2)\}}, \tag{2.2.6}$$

where $p_2(\mathbf{z}) = 1 - F_1(\infty; \mathbf{z}) = (1 - p_1)^{\exp(\boldsymbol{\beta}_1' \mathbf{z})}$. That is, we do not explicitly model the covariate effects on the competing cause. However, we allow the cause 2 asymptote to depend on covariates. Here we are taking advantage of using the modified logistic function as baseline, since the long-term probability of type k events is directly controlled by the asymptote parameter p_k, $k = 1, 2$. The likelihood function in (2.2.4) is similarly used to obtain the maximum likelihood estimators for unknown parameters. The variance of the maximum likelihood estimators will be computed from the observed Fisher information matrix.

2.3 SIMULATION STUDIES

2.3.1 Cause 1 event

We performed extensive simulation studies to evaluate the finite-sample performance of our parametric regression model (Log) as compared to the Gompertz parametric model by Jeong and Fine (Gom), the Fine and Gray method (FG) and the time-varying coefficients model proposed by Scheike et al. (Sch). Four scenarios were considered. We let the baseline CIFs satisfy either a modified logistic function (LOG) or a Gompertz function (GOM). The covariate effects follow either a proportional subdistribution hazards model (PSH) or a generalized odds-rate model (GOR).

Below we give details on how we simulated the data with the modified logistic function as baseline and the proportional subdistribution hazards for covariate effects (LOG+PSH). Essentially we simulated the event times and cause indicators by directly inverting the cause 1 CIF or the cumulative distribution of the cause 2 event conditional on the cause 2 event occurring first. The similar simulation strategy was used in Fine and Gray (1999) and discussed in Beyersmann et al. (2012, Sec 5.3). We assume that the baseline CIF for the primary cause follows (2.2.2) with $k = 1$. The regression model on the cause 1 CIF satisfies $g_1\{F_1(t;\mathbf{z})\} = g_1\{F_{01}(t)\} + \beta_{11}z_1 + \beta_{12}z_2$, where z_1 and z_2 were drawn from the standard normal distribution and $g_1(u) = \log\{-\log(1-u)\}$.Then the cause 1 CIF conditional on the covariates $\mathbf{z}^{'} = (z_1, z_2)$ has the form of (2.2.5) with $k = 1$ and $\boldsymbol{\beta}_1^{'} = (\beta_{11}, \beta_{12})$. Therefore, we simulated the event time by $F_1^{-1}(U; z)$, where $U \sim$ uniform$(0, 1)$ and F_1^{-1} is the inverse function of $F_1(t;\mathbf{z})$. Note that F_1 is improper and may not be invertible. When $U < 1 - (1-p_1)^{\exp(\boldsymbol{\beta}_1^{'}\mathbf{z})}$, we simulated the event time by

$$T = F_1^{-1}(U;\mathbf{z}) = c_1 + \frac{1}{b_1}\log\left\{\frac{1-(1-U)^{\exp(-\boldsymbol{\beta}_1^{'}\mathbf{z})} + p_1\exp(-b_1c_1)}{p_1 - 1 + (1-U)^{\exp(-\boldsymbol{\beta}_1^{'}\mathbf{z})}}\right\}$$

with $k = 1$. When $U \geq 1 - (1-p_1)^{\exp(\boldsymbol{\beta}_1^{'}\mathbf{z})}$, $k = 2$ and the event time T comes from the cause 2 event. As mentioned in Section 2.2, one needs to keep the additivity constraint

(2.1.1). Thus, for the competing cause ($k = 2$), we simply use (2.2.6) to model the CIF of the competing cause ($k = 2$). In this case, the conditional distribution of T given $k = 2$ is

$$F(t|k=2) = P(T \leq t|k=2) = \frac{P(T \leq t, k=2)}{P(k=2)} = \frac{\exp\{b_2(t-c_2)\} - \exp(-b_2c_2)}{1+\exp\{b_2(t-c_2)\}}.$$

Then we simulated the event time T by $F^{-1}(V|k=2)$, where V ~uniform(0,1). We also simulated independent censoring time C following uniform(a, b), where a and b are constants greater than zero. The observable time is $Y = \min(T, C)$ and the corresponding cause indicator is $\eta = kI\{T < C\}$. Different values of a and b are used and the percentage of censoring is around 10-20% for all simulations reported in Tables 2.1 and 2.2.

In each run of our simulations, we generated 100 or 200 pairs of event times and associated cause indicators (approximately 30% cause 1 events, 50% cause 2 events, and 20% censoring). Next, we fitted the simulated data by using our proposed parametric model with the modified logistic baseline and the complementary log-log transformation (Log) (the special case of the generalized odds-rate transformation when $\alpha \to 0$), the Gompertz parametric regression model by Jeong and Fine (2007) with the complementary log-log transformation (Gom), the Fine and Gray semiparametric method (FG) and the more general semiparametric model by Scheike et al. (Sch). For the parametric models, the regression coefficient estimates were obtained by using the R function "nlminb" to minimize the minus likelihood function (2.2.4). However, when we applied the Gompertz baseline model to the simulated data, the optimization often failed to converge. This showed the limitation of the Gompertz model in fitting sigmoidal CIF curves. For the semiparametric models, we used the R function "crr" in the package **cmprsk** (Fine and Gray, 1999) and the R function "comp.risk" in the package **timereg** (Scheike et al., 2008; Scheike and Zhang, 2011). The same procedure was repeated 2,000 times and the results were summarized in the first panel of Table 2.1 (LOG + PSH). In the table, we listed the averages of the estimates (AVE), the model-based standard errors (MSE) which were computed based on the observed Fisher information matrices, the empirical standard errors (ESE), and the coverage (Cov) rates of the 95% Wald confidence intervals for the regression coefficients β_{11} and β_{12}. Since the Scheike et al. (2008) method adopts time-varying coefficients for the two covariates, we reported the results

on coefficients from the other three models. To include the Scheike et al. method in our comparisons, we also reported the results on predicted CIFs at times 3 and 5 from the four models for individuals with covariates $Z_1 = -1$ and $Z_2 = 2$. Since the R package **cmprsk** does not provide the variance estimators for the predicted CIFs and Scheike et al. (2008) pointed out that their model includes Fine and Gray's model as a special case by specifying constant effects for all covariates, we used the R package **timereg** with constant covariates for predicting CIFs from the Fine and Gray model. The proposed parametric model with the modified logistic baseline has comparable or slightly better performance in estimating regression coefficients and predicting CIFs as compared to Fine and Gray's semiparametric model. For the model by Scheike et al. (2008), the standard errors provided by their R package seem to overestimate the true variability, where the MSEs are more than double the ESEs in the scenario of LOG+PSH. We also observed inflated MSEs in the other three scenarios below.

To evaluate the effect of mis-specifying baseline CIFs on estimating regression coefficients and predicting CIFs, we generated the data based on a Gompertz baseline and the proportional subdistribution hazards model (GOM+PSH). The Gompertz model was used by Jeong and Fine (2006, 2007) to model the baseline CIF. Specifically, under the Gompertz model the baseline CIF for cause 1 is $F_{01}(t) = 1 - \exp[\tau_1\{1 - \exp(\rho_1 t)\}/\rho_1]$, where $\rho_1 < 0$. By using the complementary log-log link function, the CIF for the primary cause is $F_1(t; \mathbf{z}) = P(T \leq t, k = 1; \mathbf{z}) = 1 - \exp[\tau_1\{1 - \exp(\rho_1 t)\}\exp(\boldsymbol{\beta}_1'\mathbf{z})/\rho_1]$. Similarly with the LOG+PSH scenario, we simulated the event time T by $F_1^{-1}(U; z)$, where $U \sim$ uniform(0,1), if F_1 is invertible. Otherwise we simulated the event time T based on the conditional distribution $F(t|k = 2) = P(T \leq t|k = 2)$. Once again, in this set of simulations, we simply used the Gompertz function to model the CIF for the competing cause $P(T \leq t, k = 2) = 1 - \exp[\tau_2\{1 - \exp(\rho_2 t)\}/\rho_2]$, $\rho_2 < 0$. The effect of covariates on the cause 2 CIF is through the constraint $F_2(\infty; \mathbf{z}) = 1 - F_1(\infty; \mathbf{z}) = \exp[\tau_1 \exp(\boldsymbol{\beta}_1'\mathbf{z})/\rho_1]$. Since $F_2(\infty; \mathbf{z}) = 1 - \exp(\tau_2/\rho_2)$, we have $\tau_2 = \rho_2 \log\{1 - \exp[\tau_1 \exp(\boldsymbol{\beta}_1'\mathbf{z})/\rho_1]\}$ and the conditional distribution $F(t|k = 2; \mathbf{z}) = \frac{1-\{1-\exp[\tau_1 \exp(\boldsymbol{\beta}_1'\mathbf{z})/\rho_1]\}^{\{1-\exp(\rho_2 t)\}}}{\exp[\tau_1 \exp(\boldsymbol{\beta}_1'\mathbf{z})/\rho_1]}$. The other simulation settings remain the same as in the first scenario, resulting in about 40% cause 1 events, 50% cause 2

Table 2.1: Simulation results where the data were simulated from our proposed modified logistic (panel LOG + PSH) or Gompertz model (panel GOM + PSH) with complementary log-log transformation or with generalized odds-rate transformation (panels LOG + GOR and GOM + GOR), where AVE is the average of the estimates, MSE is the average of the model-based standard errors, ESE is the empirical standard error, and Cov is the coverage rates of the 95% Wald CIs

LOG + PSH		$\hat{\beta}_{11}$			$\hat{\beta}_{12}$			$\hat{F}_1(3)$				$\hat{F}_1(5)$			
DIM	VAR	Log	Gom	FG	Log	Gom	FG	Log	Gom	FG	Sch	Log	Gom	FG	Sch
100	True	0.50	0.50	0.50	0.50	0.50	0.50	0.32	0.32	0.32	0.32	0.42	0.42	0.42	0.42
	AVE	0.51	-	0.51	0.52	-	0.52	0.33	-	0.33	0.33	0.44	-	0.43	0.43
	MSE	0.19	-	0.19	0.19	-	0.19	0.11	-	0.12	0.36	0.13	-	0.14	0.33
	ESE	0.20	-	0.20	0.20	-	0.20	0.12	-	0.13	0.14	0.14	-	0.15	0.16
	Cov	0.95	-	0.94	0.95	-	0.94	0.92	-	0.90	0.98	0.92	-	0.89	0.94
200	True	0.50	0.50	0.50	0.50	0.50	0.50	0.32	0.32	0.32	0.32	0.42	0.42	0.42	0.42
	AVE	0.51	-	0.51	0.51	-	0.51	0.32	-	0.33	0.33	0.43	-	0.43	0.43
	MSE	0.13	-	0.13	0.13	-	0.13	0.08	-	0.09	0.25	0.09	-	0.10	0.22
	ESE	0.14	-	0.14	0.14	-	0.14	0.08	-	0.09	0.10	0.10	-	0.11	0.11
	Cov	0.94	-	0.94	0.94	-	0.94	0.93	-	0.92	0.99	0.93	-	0.91	0.97

GOM + PSH		$\hat{\beta}_{11}$			$\hat{\beta}_{12}$			$\hat{F}_1(1)$				$\hat{F}_1(5)$			
DIM	VAR	Log	Gom	FG	Log	Gom	FG	Log	Gom	FG	Sch	Log	Gom	FG	Sch
100	True	0.50	0.50	0.50	0.50	0.50	0.50	0.51	0.51	0.51	0.51	0.56	0.56	0.56	0.56
	AVE	0.52	0.51	0.51	0.53	0.52	0.52	0.52	0.52	0.52	0.52	0.57	0.57	0.57	0.57
	MSE	0.17	0.17	0.17	0.17	0.17	0.17	0.13	0.12	0.14	0.22	0.13	0.13	0.14	0.20
	ESE	0.18	0.17	0.18	0.18	0.17	0.18	0.13	0.13	0.15	0.16	0.14	0.13	0.15	0.17
	Cov	0.94	0.95	0.93	0.94	0.94	0.94	0.91	0.91	0.90	0.91	0.90	0.90	0.90	0.88
200	True	0.50	0.50	0.50	0.50	0.50	0.50	0.51	0.51	0.51	0.51	0.56	0.56	0.56	0.56
	AVE	0.52	0.51	0.51	0.52	0.51	0.51	0.52	0.52	0.52	0.52	0.57	0.57	0.57	0.57
	MSE	0.12	0.11	0.12	0.12	0.11	0.12	0.09	0.09	0.10	0.15	0.10	0.09	0.10	0.14
	ESE	0.12	0.13	0.12	0.12	0.13	0.12	0.09	0.10	0.10	0.11	0.10	0.10	0.11	0.11
	Cov	0.94	0.91	0.94	0.94	0.92	0.94	0.94	0.92	0.92	0.94	0.93	0.91	0.92	0.91

LOG + GOR		$\hat{\beta}_{11}$			$\hat{\beta}_{12}$			$\hat{F}_1(3)$				$\hat{F}_1(5)$			
DIM	VAR	Log	Gom	FG	Log	Gom	FG	Log	Gom	FG	Sch	Log	Gom	FG	Sch
100	True	0.50	0.50	0.50	0.50	0.50	0.50	0.26	0.26	0.26	0.26	0.34	0.34	0.34	0.34
	AVE	0.56	-	0.26	0.57	-	0.27	0.27	-	0.27	0.27	0.35	-	0.35	0.34
	MSE	0.51	-	0.20	0.52	-	0.20	0.10	-	0.11	0.41	0.11	-	0.13	0.47
	ESE	0.48	-	0.21	0.47	-	0.21	0.11	-	0.12	0.12	0.12	-	0.14	0.14
	Cov	0.96	-	0.73	0.96	-	0.75	0.90	-	0.89	0.99	0.91	-	0.89	0.98
200	True	0.50	0.50	0.50	0.50	0.50	0.50	0.26	0.26	0.26	0.26	0.34	0.34	0.34	0.34
	AVE	0.54	-	0.25	0.54	-	0.25	0.27	-	0.27	0.27	0.35	-	0.35	0.34
	MSE	0.36	-	0.14	0.37	-	0.14	0.07	-	0.07	0.28	0.08	-	0.09	0.31
	ESE	0.32	-	0.14	0.33	-	0.14	0.07	-	0.08	0.09	0.08	-	0.10	0.10
	Cov	0.95	-	0.56	0.96	-	0.56	0.92	-	0.91	0.99	0.93	-	0.92	0.99

GOM + GOR		$\hat{\beta}_{11}$			$\hat{\beta}_{12}$			$\hat{F}_1(1)$				$\hat{F}_1(5)$			
DIM	VAR	Log	Gom	FG	Log	Gom	FG	Log	Gom	FG	Sch	Log	Gom	FG	Sch
100	True	0.50	0.50	0.50	0.50	0.50	0.50	0.41	0.41	0.41	0.41	0.45	0.45	0.45	0.45
	AVE	0.50	0.49	0.21	0.52	0.50	0.22	0.42	0.41	0.41	0.41	0.46	0.46	0.45	0.45
	MSE	0.44	0.45	0.17	0.44	0.44	0.17	0.11	0.11	0.13	0.27	0.11	0.11	0.13	0.25
	ESE	0.33	0.32	0.18	0.32	0.32	0.18	0.10	0.10	0.13	0.14	0.11	0.10	0.14	0.14
	Cov	0.97	0.97	0.57	0.97	0.97	0.60	0.95	0.96	0.91	0.96	0.95	0.95	0.91	0.95
200	True	0.50	0.50	0.50	0.50	0.50	0.50	0.41	0.41	0.41	0.41	0.45	0.45	0.45	0.45
	AVE	0.52	0.50	0.21	0.52	0.50	0.21	0.41	0.41	0.41	0.41	0.45	0.45	0.45	0.45
	MSE	0.32	0.32	0.12	0.32	0.32	0.12	0.08	0.07	0.09	0.19	0.08	0.08	0.10	0.17
	ESE	0.27	0.26	0.12	0.27	0.26	0.12	0.07	0.07	0.09	0.10	0.08	0.07	0.10	0.10
	Cov	0.95	0.95	0.33	0.96	0.95	0.32	0.95	0.96	0.92	0.98	0.96	0.96	0.92	0.97

events and 10% censoring cases.

We applied our proposed modified logistic regression model to the simulated GOM+PSH data and noticed that the parameter c can be any reasonable negative number with satisfactory fitting to the baseline CIFs. The modified logistic function uses three parameters to capture CIF curves that can be sigmoidal. When the data are generated from the Gompertz model, the CIF curves only have one bend point. Hence the third parameter c is not necessary. We simply fixed $c = -5$ in the model fitting and obtained the other unknown parameters by maximizing the likelihood function (2.2.4). We also considered the Gompertz regression model by Jeong and Fine (2007) with the complementary log-log link function. To have a fair comparison, we modeled the cause 1 CIF under the proportional subdistribution hazards model and used a simple Gompertz function with the constraint for the cause 2 CIF in just the same way as the data were simulated. We also applied the semiparametric Fine and Gray and Scheike et al. models to the simulated GOM+PSH data using the aforementioned R packages.

The results from 2,000 simulations are summarized in the second panel of Table 2.1 (GOM + PSH). When the baseline CIFs are from the Gompertz model, the performance of our proposed logistic model is almost identical to that of the modified Gompertz model and of the Fine and Gray method. This is in line with the results reported in Cheng (2009) that the model fit from the modified logistic function is close to those from a Gompertz model and a nonparametric estimator of CIF when the true CIF is generated from a Gompertz models. The more general method by Scheike et al. (2008) also performs well, though the standard errors provided by the package are noticeably larger than the ones from the other three models. The performance improves slightly when the sample size increases from 100 to 200. We also simulated the data with the sample size of 500 and 40% censoring and observed similar results; see Table A1 in the appendix.

Next, we evaluated and compared the performance of the four models when the proportional subdistribution hazards assumption does not hold. We adopted the generalized odds-rate (GOR) transformation (Dabrowska and Doksum, 1998; Jeong and Fine, 2007) for the cause 1 CIF. That is, $g_1\{F_1(t;\mathbf{z});\alpha_1\} = g_1\{F_{01}(t);\alpha_1\} + \boldsymbol{\beta}_1'\mathbf{z}$, where $g_1\{\nu;\alpha_1\} =$

$\log[\{(1-\nu)^{-\alpha_1}-1\}/\alpha_1]$. We set $\alpha_1 = 5$ in our simulations. We considered both the modified logistic function and Gompertz function for the baseline CIFs and simply used either the modified logistic function or the Gompertz function with the constraint for the cause 2 CIF similar to the previous scenarios. Analogously, we fitted our modified logistic regression and the modified Gompertz model with the generalized odds-rate transformation, the Fine and Gray method and the semiparametric model by Scheike et al. (2008) to the simulated LOG+GOR and GOM+GOR data. In both scenarios, there are around 30-35% cause 1 events, 50% cause 2 events and 20-15% censored events.

The results are summarized in the third and fourth panels of Table 2.1. Analogous to the LOG+PSH, when the baseline CIFs satisfy the modified logistic function, the Gomperz model often fails to converge. As the proportional hazards assumption is clearly violated, the Fine and Gray method results in serious downward bias in estimating covariate effects. Their standard error estimates are also systematically smaller than those from the modified logistic model. However, the significance levels from the two models are similar. For example, the average p values for both coefficients are around 0.19 from the modified logistic model and around 0.17 from Fine and Gray's model when the sample size of the LOG+GOR data is 200. The predictions of CIFs from both models are also comparable. As before, the Scheike et al. model tends to overestimate the standard errors in their CIF estimators. Latouche et al. (2007) and Grambauer et al. (2010) showed that the Fine and Gray model provides a reasonable summary of covariate effects (least false parameters) even when the proportional subdistribution hazards assumption does not hold. Moreover, since the Fine and Gray model is a semiparametric approach, the misspecified covariate effects may be compensated by the nonparametric estimation of the baseline CIF. Hence, the predicted CIF is close to the true value despite the misspecified covariate effects. For the LOG+GOR data, our modified logistic model clearly has the best performance among the four. In the last panel for the GOM+GOR data, once again we fixed $c = -5$ when we applied our modified logistic model to the simulated data. The other three methods were also implemented. The Fine and Gray method again seriously underestimates regression coefficients though their predictions of CIFs are satisfactory. When the true data are from the Gompertz model, the modified

Gompertz model performs well. However, our modified logistic model performs just as well as the Gompertz model does. Both work well for the sample of 100 and clearly outperform the general semiparametric model by Scheike et al. (2008). The latter overestimates the variability in predicting CIFs so its coverage rate is close to 1. On the other hand, the coverage rates of regression parameters from the Fine and Gray method are much lower than the nominal level and actually decrease when the sample size increases.

Next, we focused on our parametric model and the Fine and Gray method, and investigated how the censoring distribution affects the prediction of CIFs. We simulated censoring times from the log logistic model. That is, $\log(C) = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \sigma\epsilon$, where γ_0 is a constant and chosen so that the censoring rate is around 20%, $\gamma_1 = \gamma_2 = 0.5$, $\sigma = 0.5$, and $\epsilon \sim$ standard logistic distribution. Therefore, the censoring time depends on covariates in a nonproportional hazards fashion. The simulation settings for the event times remain the same. We applied our parametric model and the Fine and Gray method to each simulated data set. For the Fine and Gray method, we used the R package **timereg** by specifying a Cox model for estimating the censoring distribution. The results are generally consistent with those reported in Table 2.1, and the coverage rates of the predicted CIFs from our parametric model are slightly better than those from the Fine and Gray method (89-95% versus 82-96%); see Table A2 in the appendix for more detail.

2.3.2 Cause 2 event

So far we have been focusing on the cause 1 event. In practice, both events may be of interest. Since our parametric modeling specifies the expressions for both CIFs, we obtained the predicted cause 2 CIFs along with the cause 1 quantities for each data set that we simulated previously. For the Fine and Gray method, one may run the model twice, once for each cause. Therefore, we also applied this method to our simulated data focusing on the cause 2 event. As in the previous subsection, we considered the four scenarios, each simulating 2,000 data sets of size 200. The estimated average, model-based standard errors, empirical standard errors and the coverage rates of 95% confidence intervals of the cause 2

Table 2.2: Simulation results on the cause 2 CIFs; the data were simulated from the modified logistic or Gompertz base with a proportional subdistribution hazard transformation (LOG+PSH and GOM+PSH) or with a generalized odds-rate transformation (LOG + GOR and GOM + GOR); referring to Table 2.1 for the definition of AVE, MSE, ESE, and Cov

		$\hat{F}_2(3)$		$\hat{F}_2(5)$				$\hat{F}_2(1)$		$\hat{F}_2(5)$	
DATA	VAR	Log	FG	Log	FG	DATA	VAR	Log	FG	Log	FG
LOG+PSH	True	0.07	0.07	0.49	0.49	GOM+PSH	True	0.31	0.31	0.44	0.44
	AVE	0.06	0.06	0.48	0.49		AVE	0.33	0.32	0.43	0.44
	MSE	0.02	0.03	0.09	0.11		MSE	0.07	0.09	0.10	0.12
	ESE	0.02	0.04	0.09	0.10		ESE	0.08	0.11	0.10	0.16
	Cov	0.92	0.82	0.94	0.96		Cov	0.93	0.94	0.92	0.91
LOG+GOR	True	0.08	0.08	0.57	0.57	GOM+GOR	True	0.40	0.40	0.55	0.55
	AVE	0.08	0.08	0.56	0.57		AVE	0.41	0.41	0.55	0.55
	MSE	0.02	0.03	0.07	0.11		MSE	0.06	0.10	0.08	0.12
	ESE	0.02	0.06	0.07	0.11		ESE	0.06	0.12	0.08	0.15
	Cov	0.93	0.90	0.93	0.95		Cov	0.95	0.94	0.95	0.91

CIFs from both models are summarized in Table 2.2.

We can see that on average the Fine and Gray method can still predict the cause 2 CIF pretty accurately, though its empirical standard errors are often double the empirical standard errors from our parametric model. When the data are simulated from a modified logistic baseline with the complementary log-log transformation, the coverage rate of the Fine and Gray model is noticeably worse than that from the parametric model. Since we naively applied the Fine and Gray method to both causes, we actually assumed that both CIFs satisfy (2.2.1) with the complementary log-log link function. Let p_1 and p_2 be the asymptotes of baseline cause 1 and cause 2 CIFs, where $p_2 = 1 - p_1$. Under the log-linear model in (2.2.1) with $g_k(u) = \log\{-\log(1-u)\}$ and by the additivity constraint in (2.1.1), we have $1-(1-p_1)^{\exp(\boldsymbol{\beta}_1'\mathbf{z})} = p_1^{\exp(\boldsymbol{\beta}_2'\mathbf{z})}$, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are regression coefficients for causes 1 and 2, respectively. When we have two covariates, the above relationship is not likely to hold. This has also been noticed by Latouche et al. (2007) and Grambauer et al. (2010). Hence the performance of the Fine and Gray method is not satisfactory under this setting. Under the other three scenarios, the effects of ignoring the additivity constraint between two

CIFs on predicting CIFs appear to be less severe.

2.4 BREAST CANCER STUDY

We applied our proposed modified logistic model along with the other three to the National Surgical Adjuvant Breast and Bowel Project Breast Cancer Study (NSABP B-14). This study investigated the effects of tamoxifen for node negative and hormonal receptor positive patients. The data contain information on time (in years), event (0=censored, 1=recurrence, 2=other events), treatment group (trt=1,placebo; trt=2, tamoxifen), age, and tumor size (=tsize), for 2,582 eligible patients who had follow-up and known tumor sizes. Recurrence is the event of interest and other events are treated as competing ones. Before fitting the data, we coded trt=0 for placebo and trt=1 for tamoxifen, and centered the age at mean 55 and tumor size at mean 22. We first ran the Kolmogorov-Smirnov test and the Cramer von Mises test for time invariant effects of treatment group, age and tumor size (Scheike et al., 2008). The assumption of proportional hazards for subdistribution did not hold for age. Figure 2.1 shows the estimated time-varying coefficient for age $\alpha(t)$ in the Breast cancer Study using Scheike et al.'s model. On the left, we show the time-varying coefficient estimate for age (solid line) along with its 95% pointwise confidence intervals (broken lines) and 95% confidence bands (dotted lines). The age effect is time varying, though the confidence bands generally cover 0. The package **timereg** also outputs the test process for the null hypothesis $\alpha(t)$ being a constant. The solid line on the right panel of the figure represents the test process based on the estimated time-varying coefficient $T(t, \hat{\alpha}) = \hat{\alpha}(t) - \frac{1}{\tau}\int_0^\tau \hat{\alpha}(t)dt$, where τ is some reasonably large time point. The gray area corresponds to the simulated test processes under the null hypothesis. There is a clear departure from the null in the observed test process.

To handle the departure of the proportional subdistribution hazards assumption for age, we ran two additional models. One is a stratified Fine and Gray model (SFG) proposed

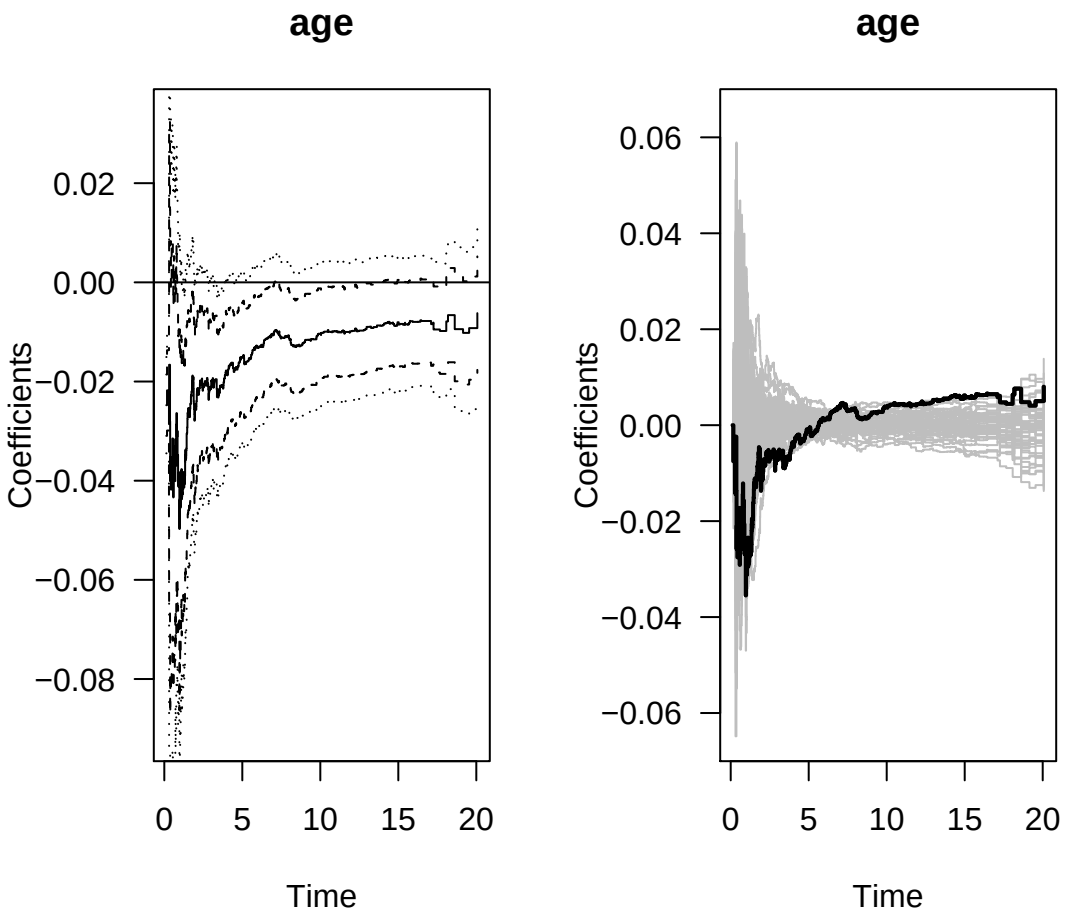

Figure 2.1: The estimates of time-varying coefficient for age in the Breast cancer Study

by Zhou et al. (2011) (R package **crrSC**) for treatment and tumor size stratified on age (dichotomized at mean). The other is the standard Fine and Gray model including terms of treatment, tumor size, age as well as the time by age interaction (FGt). The estimated coefficients are summarized in Table 2.3. The coefficient estimates for treatment and tumor size are almost identical under these three models. The coefficient estimates for age are very similar in the Fine-Gray model with or without the age by time interaction. The coefficient for the interaction term is not significantly different from zero.

Hence we applied our modified logistic model and the modified Gompertz model with the generalized odds-rate transformation, as well as the semiparametric model by Scheike et al.

Table 2.3: The estimates of the regression coefficients for the Breast cancer Study based on the Fine-Gray model (FG), the Fine-Gray model with age by time interaction (FGt), and the stratified Fine-Gray model (SFG), where we list the coefficient estimates (Est), standard deviations (STD), the values of the Wald statistics (z value) and corresponding p values

	$\hat{\beta}_{1\_trt}$			$\hat{\beta}_{1\_age}$			$\hat{\beta}_{1\_tsize}$			$\hat{\beta}_{1\_age*t}$
VAR	FG	FGt	SFG	FG	FGt	SFG	FG	FGt	SFG	FGt
Est	-0.48	-0.48	-0.48	-0.01	-0.02	-	0.02	0.02	0.02	0.001
STD	0.09	0.09	0.09	0.004	0.007	-	0.003	0.003	0.003	0.001
z value	-5.60	-5.59	-5.62	-2.58	-2.67	-	7.40	7.41	7.52	1.36
p value	<0.001	<0.001	<0.001	0.01	0.01	-	<0.001	<0.001	<0.001	0.16

(2008) to the breast cancer data. The original Fine and Gray (1999) model without the age by time interaction was also included for comparison because our simulations suggest that this model can have good prediction on CIFs even though the estimates of the regression coefficients may be biased.

Since the proportional hazards assumption does not hold for age, we specified a time-varying coefficient for age in Scheike's model and did not include the coefficient estimate for age from this model in Table 2.4. For the remaining three models, the coefficient estimates and standard errors from the modified logistic model and the modified Gompertz model are generally close to each other, and the values from the Fine and Gray model and the Scheike et al. model are consistently smaller than those from the parametric models. This is in line with the simulation studies which suggest that the parametric models and semiparametric models may have different covariate coefficients when there is a departure from proportional subdistribution hazards assumption. However, the p values from the three models are comparable and fairly small for all three prognostic factors. The coefficients for treatment and age are all negative while those for tumor size are all positive across the four models. This suggests that the patients who were older, received Tamoxifen treatment and had smaller tumor sizes were less likely to have cancer recurrence. This also implies that these patients would have higher long-term incidences from competing events such as deaths or second primary cancers.

Next, we selected two patient cases to compare the predicted CIFs from the four models.

Table 2.4: The estimates of the regression coefficients for the Breast cancer Study based on our proposed modified logistic (Log) and the modified Gompertz model (Gom) with generalized odds-rate transformation, the Scheike et al. model (Sch) and the Fine-Gray model (FG); referring to Table 2.3 for the definitions of Est, STD, z value, and p value

	$\hat{\beta}_{1\_trt}$				$\hat{\beta}_{1\_age}$				$\hat{\beta}_{1\_tsize}$			
VAR	Log	Gom	FG	Sch	Log	Gom	FG	Sch	Log	Gom	FG	Sch
Est	-0.82	-0.68	-0.48	-0.48	-0.02	-0.03	-0.01	-	0.04	0.04	0.02	0.02
STD	0.16	0.16	0.09	0.09	0.01	0.01	0.004	-	0.01	0.01	0.003	0.003
z value	-5.15	-4.19	-5.60	-5.47	-3.22	-4.59	-2.58	-	6.05	5.37	7.40	5.95
p value	<0.001	<0.001	<0.001	<0.001	0.001	<0.001	0.01	-	<0.001	<0.001	<0.001	<0.001

Patient I has good prognostic factors: Tamoxifen (trt = 1), 65 years old (age = 10) and tumor size 12 (tsize = -10), while patient II has poor prognostic factors: placebo (trt = 0), 45 years old (age = -10) and tumor size 32 (tsize = 10). The four predicted CIFs for each patient are plotted in Figure 2.2. The solid lines are the estimated CIFs from 0 to 20 years from the modified logistic model, the broken lines are for the modified Gompertz model and the dotted lines are from Fine and Gray's model. The Scheike semiparametric model is represented by the gray lines. We first observe that the predicted CIF curves have only one bend point. In this case, the two parametric models have similar coefficient estimates. The predicted CIFs from the two parametric models, Fine and Gray's model and the Scheike et al. model are really close, which is consistent with our simulation results. The predicted 20-year recurrence rate is about 0.1 for patient I and about 0.4 for patient II based on the four models. In the figure, we also include the confidence bands (gray broken lines) from the Scheike et. al. model, which is available from the R package **timereg**.

2.5 DISCUSSION

In this chapter, we have proposed a simultaneous inference for the CIF of a primary event and the CIF of a competing event using a flexible baseline function and a generalized odds-rate

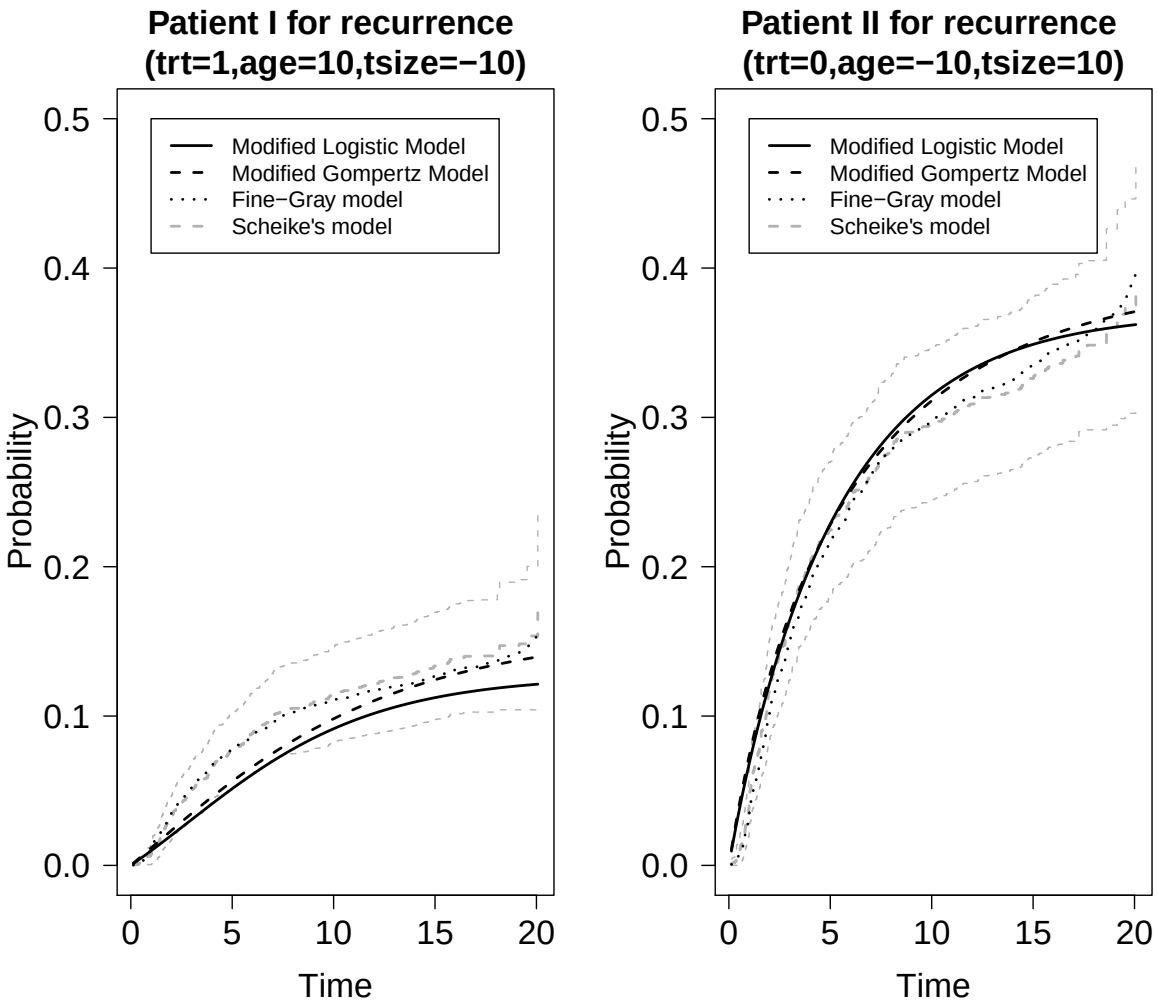

Figure 2.2: Breast Cancer Study

model, for which standard maximum likelihood theory is readily applied. This parametric model provides a robust alternative to its semiparametric counterparts: it not only accounts for nonproportionality but also provides a reliable variance estimator. It does not require modeling for censoring times. In addition, the parametric method can easily be extended to the case where the event time is interval-censored or left-truncated. More importantly, this parametric model explicitly incorporates the additivity constraint between the asymptotes of the CIFs from both causes for any given covariates.

To the best of our knowledge, this is the first attempt to address the interrelationship

between two CIFs in regression analysis, a circumstance for which other methods fail to account. Ignoring the additivity constraint clearly results in interpretation issues, since in reality subjects with certain characteristics would fail from one of the competing events eventually. We have also shown through simulations that ignoring the constraint may lead to larger variability and lower coverage rates in estimating CIFs.

As illustrated in our example in (2.2.5), it is difficult to explicitly incorporate the additivity constraint in (2.1.1) in the joint modeling of CIFs. Instead of explicitly modeling covariate effects on the competing cause, we let the covariates influence the cause 2 event through the relationship between the causes 1 and 2 plateaus. This may be a potential limitation of the method. In contrast, for models based on cause-specific hazard (CSH) functions (Korn and Dorey, 1992; Cheng et al., 1998; Hyun et al., 2009), we do not have this problem, as the hazard functions from two causes can vary from zero to infinity. In addition, modeling one cause hazard function will not affect modeling for the other. However, covariates' effects on CSH functions do not directly translate into the effects on CIFs (Gray, 1988; Pepe and Mori, 1993; Fine and Gray, 1999). As CIFs have appealing probability interpretations, there is substantial interest in developing regression analysis based on CIFs. Hence, we have proposed a parametric regression method which explicitly models the cause 1 CIF through a generalized odds-rate model and a modified logistic function as baseline, and models the cause 2 CIF using a modified logistic baseline with the asymptote determined by the additivity constraint. This parametric model allows joint modeling two CIFs and clearly specifies how the two CIFs interplay with one another among subjects with certain characteristics.

3.0 ASSESSING ACCURACY IMPROVEMENT FOR COMPETING-RISK CENSORED OUTCOMES

3.1 INTRODUCTION

Diagnostic tests are applied to detect the disease status in clinical research. Statistical problems arising from such practices include how to assess the accuracy of the test and how to design tests with adequate sensitivity and specificity. In the past decade, diagnostic accuracy studies have progressed to involve complicated survival outcomes (Heagerty et al., 2000) beyond the traditional dichotomous outcome. When time-to-event data are concerned, the test becomes prognostic, since it is used in a population that is yet event free to decide who will or will not develop the event of interest at some later time. In particular, time-dependent receiver operating characteristic (ROC) curves and area under the ROC curve (AUC) are proposed for right-censored (Heagerty et al., 2000; Heagerty and Zheng, 2005) and interval-censored failure times (Li and Ma, 2011).

Very often more than one tests are used to build a stronger marker with higher accuracy for the outcome. We intend to investigate the accuracy improvement over the course of variable addition. AUC or other ROC-based measures are well-known to be insensitive to detecting the added values of new markers (Greenland and O'Malley, 2005; Pepe et al., 2004; Ware, 2006). Based on such observations, new indicators have been developed to complement the AUC measure. Among others, Pencina et al. (2008) developed two very

useful criteria: the net reclassification improvement (NRI) and integrated discrimination improvement (IDI). NRI is constructed by re-distributing membership of disease categories. It was originally proposed for categorical outcomes only and has been recently extended to survival outcomes (Pencina et al., 2011). IDI is the sum of the improvement in the integral of sensitivity and that in the integral of specificity over all possible cutoff values from the model without the new biomarkers to the model with the new biomarkers. For more discussion on IDI, see Chi and Zhou (2008), Cook (2008), Greenland (2008), Kraemer (2008), Pepe et al. (2008) and Ware and Cai (2008).

Our goal in this chapter is to integrate the time-dependent diagnostic accuracy study and the IDI statistic. From our review of the literature, the closest development to this goal is Uno et al. (2009) where the authors focused on distribution functions of the difference in predicted risks, conditional on either that subjects have developed the disease or that they have not yet by a specific time point. The two conditional distribution functions were estimated by an inversely weighing method and the inference was carried out based on a perturbation-resampling method (Uno et al., 2009). In this chapter, we adopt the same definition of IDI but will propose a set of alternative estimators to the conditional distributions based on a two-dimensional Kaplan-Meier estimator (Dabrowska, 1988).

Furthermore, we consider competing-risk outcomes which are more general than a single event with random censoring and also more challenging due to dependence among competing events. There has been extensive research in handling competing-risk censoring including Tsiatis (1975), Prentice et al. (1978), Gray (1988), Gaynor et al. (1993), Pepe and Mori (1993) and Fine and Gray (1999), among many others. However, the work on evaluating discrimination ability of biomarkers on an outcome that is subject to competing-risk censoring is limited. Foucher et al. (2011), Saha and Heagerty (2010) and Zheng et al. (2011) extended some of the existing methods to handle an outcome that is subject to competing-risk censoring. However, these methods used AUC to evaluate the overall accuracy of the marker of interest and did not focus on evaluating the added values of the new markers.

With the presence of competing-risk censoring, at a fixed time point, the whole population will be divided into two groups: the "disease" group including subjects who have

developed the event of interest and the "healthy" group including not only those who have not had any event but also those who have experienced competing events. It is worth pointing out that the "disease" and"healthy" groups are defined based on following-up an event prospectively which differ from a retrospective case-control study. Consequently, a "healthy" subject at one time may become "diseased" at a later time point. We adapt the conditional distributions considered in Uno et al. (2009) to this new "disease" and "healthy" grouping, and also propose two sets of estimators based on the inverse probability weighting and the bivariate cumulative incidence function (CIF) estimator which was developed by Cheng et al. (2007). Some summary statistics are extended to the competing-risk setting for evaluating the overall improvement by new biomarkers in risk prediction. In addition, we re-sample the difference in risks based on the two conditional distribution functions and display the two conditional density curves to better illustrate the added predictive value of the new biomarkers.

The rest of the chapter is organized as follows. In Section 3.2, we provide details on the two sets of estimators for the conditional distributions, the summary statistics and the graphic displays. It is followed by simulation studies in Section 3.3 to evaluate and compare the two sets of estimators for the conditional distribution functions. In Section 3.4, we apply these estimators to a breast cancer dataset to assess whether or not adding a new marker, gene score, into a model can improve the prediction over conventional markers. We conclude with some remarks in Section 3.5.

3.2 METHODS

In this chapter we discuss both typical survival outcomes and competing-risk outcomes. In a competing-risk setting, there are usually two or more types of events. Without loss of generality, we consider only two events, the primary event with the event type indicator $k = 1$, and the competing event with $k = 2$, as multiple competing events can be grouped

together. We define T as the time to first event, whichever occurs first. When we consider the composite endpoint T, it may also be subject to independent censoring by C. As in a typical survival study, we observe $X = \min(T, C)$ and the censoring indicator $\Delta = I(T \leq C)$, where $I(\cdot)$ is an indicator function. Let $\mathbf{z}_1$, a $p \times 1$ vector, denote conventional predictors and let $\mathbf{z}_2$, a $q \times 1$ vector, denote the new biomarkers. At any pre-specified time t_0, we define

$$P_{10}(\mathbf{z}_1; t_0) = \Pr(T \leq t_0 | \mathbf{z}_1), \text{ and } P_{20}(\mathbf{z}_1, \mathbf{z}_2; t_0) = \Pr(T \leq t_0 | \mathbf{z}_1, \mathbf{z}_2),$$

where P_{10} and P_{20} are the cumulative risks of developing the composite event by time t_0, given the conventional markers only and given both the conventional and new biomarkers. Though survival functions are more commonly used in survival analysis, we adopt the cumulative distribution functions of the composite event, since they are simply the survival functions subtracted from 1, and they are more closely related to the following CIFs. When we focus on a specific event, say $k = 1$, we are interested in

$$P_1(\mathbf{z}_1; t_0) = \Pr(T \leq t_0, k = 1 | \mathbf{z}_1), \text{ and } P_2(\mathbf{z}_1, \mathbf{z}_2; t_0) = \Pr(T \leq t_0, k = 1 | \mathbf{z}_1, \mathbf{z}_2),$$

which correspond to cumulative risks of the primary or cause 1 event by time t_0. With the presence of censoring due to a competing event, the survival function of the latent failure time of cause 1 event is not well defined. Instead, the CIF and the cause-specific hazard function have been widely used in the competing-risk literature (Kalbfleisch and Prentice, 2002). We adopt the CIF in our current study as it has an appealing probability interpretation.

The above probabilities are functions of the markers (or covariates) and therefore are random variables. Treating probabilities as random variables might be like a surprise for some practitioners. However, such an idea is not new and has emerged in quite many areas. For example, the commonly used P-values are treated as random in microarray studies (Efron, 2010). The risks are unobservable and must be evaluated indirectly through data. In practice, we usually construct a regression model that helps calibrate the values of these probabilities and treat them as observations of the random variables. Conventionally, we still call the calibrated values as risks. For binary outcomes, the logistic regression model and other appropriate binary regression models may be used (Pfeiffer and Gail, 2011). For

typical survival and competing-risk outcomes, one can usually adopt the Cox proportional hazards model or the Fine and Gray model (Fine and Gray, 1999).

3.2.1 Estimating conditional distributions of the composite event time

Uno et al. (2009) estimated P_{10} and P_{20} based on a random sample $\{(X_i, \Delta_i, \mathbf{z}_{1i}, \mathbf{z}_{2i}), i = 1, \ldots, n\}$ using the Cox model. More specifically, two Cox models were fitted to the data, without and with the new biomarkers $\mathbf{z}_2$. Then for each subject the predicted risks of developing an event at time t_0, $\hat{P}_{10}$ and $\hat{P}_{20}$, were calculated based on the two models as

$$\begin{aligned} \hat{P}_{10}(\mathbf{z}_1; t_0) &= 1 - \hat{S}_{10}(t_0)^{\exp(\mathbf{z}_1^T \hat{\boldsymbol{\beta}}_1)}, \\ \hat{P}_{20}(\mathbf{z}_1, \mathbf{z}_2; t_0) &= 1 - \hat{S}_{20}(t_0)^{\exp(\mathbf{z}_1^T \hat{\boldsymbol{\beta}}_2 + \mathbf{z}_2^T \hat{\boldsymbol{\beta}}_2^*)}, \end{aligned}$$

where $\hat{S}_{j0}(t)$ $(j = 1, 2)$ are the estimated baseline survival functions for the two Cox models, $\hat{\boldsymbol{\beta}}_j$ are the estimated regression coefficients for $\mathbf{z}_1$ under two models, and $\hat{\boldsymbol{\beta}}_2^*$ is the estimated regression coefficients for $\mathbf{z}_2$ in the second model. We note that $\hat{P}_{10}$ and $\hat{P}_{20}$ may themselves be of interest to practitioners and their distributions could provide some insight on risk management. In fact, Pfeiffer and Gail (2011) studied distributions of such quantities for binary outcomes and defined some criteria for risk prediction based on a Lorenz transformation and an inverse Lorenz transformation. The same criteria can be defined for survival outcomes easily by introducing $I(T \leq t_0)$ as a binary outcome.

We are interested in evaluating whether the model with new markers better discriminates subjects who will develop the outcome of interest by time t_0 from those who will not across various levels of $\mathbf{z}_1$ and $\mathbf{z}_2$. To this end, we adopt the definitions proposed in Uno et al. (2009), and focus on the difference in predicted risks between the models with or without the markers, $\hat{D}(t_0) = \hat{P}_{20}(\mathbf{z}_1, \mathbf{z}_2; t_0) - \hat{P}_{10}(\mathbf{z}_1; t_0)$, and the conditional distributions of $\hat{D}$:

$$F(s; t_0) = \Pr(\hat{D}(t_0) \leq s | T \leq t_0) \text{ and } G(s; t_0) = \Pr(\hat{D}(t_0) \leq s | T > t_0),$$

where $s \in [-1, 1]$. We provide a few insights here. For a good biomarker, we expect an increase in the predictive risk of the event for subjects who have the event and a decrease

in the predictive risk for the non-event population. Hence, its corresponding $F(s;t_0)$ should be close to 0 for $s<0$ and rise fast for $s>0$ while its corresponding $G(s;t_0)$ should rise fast for $s<0$ and stay flat for $s>0$. These two distribution functions act similarly to sensitivity and specificity widely used in diagnostic medicine and therefore can be used to describe the prediction accuracy for survival outcomes. Adopting the terminology therein, we call F as the true positive probability (TPP) and G as the false positive probability (FPP) for the estimated risk difference. The focus of this chapter is to estimate these two conditional probabilities as well as their extensions to competing-risk settings, and to develop accuracy improvement diagnostics based on these two distribution functions.

Uno et al. (2009) proposed the following inversely weighted estimators:

$$\hat{F}(s;t_0)=\frac{\sum_{i=1}^{n}\Delta_i\{\hat{H}(X_i)\}^{-1}I\{\hat{D}_i\le s,X_i\le t_0\}}{\sum_{i=1}^{n}\Delta_i\{\hat{H}(X_i)\}^{-1}I(X_i\le t_0)}$$

and

$$\hat{G}(s;t_0)=\frac{\sum_{i=1}^{n}I\{\hat{D}_i\le s,X_i>t_0\}}{\sum_{i=1}^{n}I(X_i>t_0)},$$

where $\hat{H}(\cdot)$ is the estimator for the survival function of censoring, i.e. $H(t)=\Pr(C>t)$, and $\hat{D}_i=\hat{P}_{20}(\mathbf{z}_{1i},\mathbf{z}_{2i})-\hat{P}_{10}(\mathbf{z}_{1i})$ is the difference in the estimated risk probabilities of the two models given covariates for the i-th subject. We call these estimators as the inverse probability weighting (IPW) estimators. The inference of these estimators is based on a perturbation-resampling method (Uno et al., 2009).

We now propose a new set of estimators for $F(s;t_0)$ and $G(s;t_0)$. Our new estimators are based on the fact that the two conditional distribution functions can be written as

$$F(s;t_0)=\frac{\Pr(\hat{D}\le s,T\le t_0)}{\Pr(T\le t_0)}\text{ and }G(s;t_0)=\frac{\Pr(\hat{D}\le s,T>t_0)}{\Pr(T>t_0)}.$$

That is, the conditional distribution functions can be expressed in terms of the joint survival function of $\hat{D}$ and T, $S_{\hat{D},T}(s,t_0)=\Pr(\hat{D}>s,T>t_0)$ and the marginal quantities, $S_{\hat{D}}(s)=P(\hat{D}>s)$ and $S_T(t_0)=P(T>t_0)$. We adopt the Dabrowska (1988) estimator for the bivariate survival function, $\hat{S}_{\hat{D},T}(s,t_0)$, since it has good practical performance and is easy to implement. The marginal quantities can be estimated by the Kaplan-Meier estimator $\hat{S}_T(t_0)$ or one minus the empirical cumulative distribution function $\hat{S}_{\hat{D}}(s)$. Plugging in these

nonparametric estimators for the bivariate and univariate quantities, we provide alternative estimators for $F(s;t_0)$ and $G(s;t_0)$ as follows:

$$\hat{F}'(s;t_0) = \frac{1-\hat{S}_T(t_0)-\hat{S}_{\hat{D}}(s)+\hat{S}_{\hat{D},T}(s,t_0)}{1-\hat{S}_T(t_0)}$$

and

$$\hat{G}'(s;t_0) = \frac{\hat{S}_T(t_0)-\hat{S}_{\hat{D},T}(s,t_0)}{\hat{S}_T(t_0)}.$$

We call our estimators as Dabrowska's estimators (Dab) and use the bootstrap method for inference. More specifically, after we fit the two models with and without the new biomarkers, we obtain the data $\{(\hat{D}_i, X_i, \Delta_i), i = 1,\ldots,n\}$. Then we treat the data as the "raw" data and resample with replacement to obtain B bootstrap samples $\{(\hat{D}_i^b, X_i^b, \Delta_i^b), i = 1,\ldots,n\}, b = 1,\ldots,B$. For each bootstrap sample, we obtain the estimators $\hat{F}'^b(s;t_0)$ and $\hat{G}'^b(s;t_0)$, yielding B replications for each estimator based on which the bootstrap standard deviation is calculated.

Since the Dabrowska estimator was developed for bivariate events both subject to random censoring, our estimator has the potential to handle the case where the risk difference is censored. For example, the biomarkers of some subjects may be beyond the detection limit, hence the predicted risk is right censored by the risk at the detection limit. When the risk difference is observed for each subject, our estimators and the inverse weighting estimators have comparable performance based on our simulation studies; see Section 3.3 for more details. The exact computation formula for our estimator may look complicated and need rather lengthy description of notation. We refer to Dabrowska (1988) for more details. To facilitate applications, we have made our code downloadable at the following website:

$$http://www.stat.pitt.edu/yucheng/software.html$$

and readers with a basic understanding of `R` should have no difficulty in implementing our code.

3.2.2 Estimating conditional distributions in a competing-risk setting

We now consider a more general setting where the primary event of interest may be dependently censored by competing events. With the presence of competing events, we are interested in predicting cumulative risks of the primary event over time given conventional predictors $P_1(\mathbf{z}_1; t_0)$ and additionally new markers $P_2(\mathbf{z}_1, \mathbf{z}_2; t_0)$. To estimate P_1 and P_2, we may use the multistate model proposed by Andersen et al. (2002) and the proportional subdistribution hazards model by Fine and Gray (1999), which are the two popular methods among CIF regression models. In the following we provide some details on the multistate model and will give some discussion on the Fine and Gray model in the data analysis for the risk of metastasis. Since the likelihood function for univariate competing risks data factors into a separate component for each cause-specific hazard function (Kalbfleisch and Prentice, 2002), we can apply the Cox model to a specific cause while treating the other event as if it were independent censoring. The multistate model essentially is to run the Cox model twice, one for each event type. For the model with conventional predictors only, the cumulative cause-specific hazard functions can be estimated through $\hat{\Lambda}_k(u; \mathbf{z}_1) = \hat{\Lambda}_{k0}(u) \exp(\mathbf{z}_1^T \hat{\boldsymbol{\beta}}_1^k)$, where $\hat{\Lambda}_{k0}(u)$ are the estimated baseline cumulative cause-specific hazard functions and $\hat{\boldsymbol{\beta}}_1^k$ are the estimated coefficients of traditional predictors from the Cox model for the cause k event, $k = 1, 2$. Then, the overall survival function is estimated as $\hat{S}(u; \mathbf{z}_1) = \exp\{-\hat{\Lambda}_1(u; \mathbf{z}_1) - \hat{\Lambda}_2(u; \mathbf{z}_1)\}$. The CIF at t_0 is then estimated through its relationship with the cause-specific hazard function and the overall survival function. That is, $\hat{P}_1(\mathbf{z}_1; t_0) = \int_0^{t_0} \hat{S}(u-; \mathbf{z}_1) d\hat{\Lambda}_1(u; \mathbf{z}_1)$. Following the same method, for the model with both conventional predictors and new biomarkers we obtain an estimator $\hat{P}_2(\mathbf{z}_1, \mathbf{z}_2; t_0)$. The difference in predicted risks for the cause 1 event is $\hat{D}^*(t_0) = \hat{P}_2(\mathbf{z}_1, \mathbf{z}_2; t_0) - \hat{P}_1(\mathbf{z}_1; t_0)$.

Focusing on the cause 1 event, we divide the population into those who have developed the cause 1 event by t_0 and those who have not. We define the conditional distributions as

$$F_1(s; t_0) = \Pr\{\hat{D}^*(t_0) \le s | T \le t_0, k = 1\} \text{ and } G_1(s; t_0) = \Pr\{\hat{D}^*(t_0) \le s | (T \le t_0, k = 1)^c\}. \quad (3.2.1)$$

As we have mentioned in the introduction, the "healthy" group in G_1 includes all patients who have not experienced any disease and those who have experienced the competing event.

This definition of a disease-free status is consistent with the augmented at-risk set used in Fine and Gray (1999) and Fine (2001), where subjects who have failed from cause 2 event before or at t are included in the at-risk set for cause 1 event. Foucher et al. (2011) considered the set $\{T > t_0, k = 1\}$ which is less intuitive as $\Pr(T > t_0, k = 1)$ is nonparametrically nonidentifiable. However, if the new biomarkers are expected to significantly improve the prediction of cause 2 event, it may not be reasonable to include those subjects who have failed from cause 2 event by t_0 in the disease-free group. In this case, we may consider computing two F functions for each event, $F_1(s; t_0) = \Pr(\hat{D}^* \leq s | T \leq t_0, k = 1)$ and $F_2(s; t_0) = \Pr(\hat{D}^* \leq s | T \leq t_0, k = 2)$, and comparing with the "control" group that only includes those who have not failed from any event yet, $G^*(s; t_0) = \Pr(\hat{D}^* \leq s | T > t_0)$, as suggested in Pepe et al. (2008) and considered in Saha and Heagerty (2010) and Zheng et al. (2011). We will focus on the definitions in (3.2.1) in this chapter and note that the proposed estimating strategy in the following can easily be adapted to the alternative definitions.

Parallel to the inverse probability weighting estimator and Dabrowska estimator for the outcome with independent censoring, we develop the following two sets of estimators of $F_1(s; t_0)$ and $G_1(s; t_0)$ for the outcome with competing-risk censoring. We first extend the inversely weighted estimators to the competing risks setting as follows:

$$\hat{F}_1(s; t_0) = \frac{\sum_{i=1}^n \Delta_i \{\hat{H}(X_i)\}^{-1} I\{\hat{D}_i^* \leq s, X_i \leq t_0, k_i = 1\}}{\sum_{i=1}^n \Delta_i \{\hat{H}(X_i)\}^{-1} I(X_i \leq t_0, k_i = 1)}$$

and

$$\hat{G}_1(s; t_0) = \frac{\sum_{i=1}^n [I\{\hat{D}_i^* \leq s, X_i > t_0\}\{\hat{H}(t_0)\}^{-1} + \Delta_i \{\hat{H}(X_i)\}^{-1} I\{\hat{D}_i^* \leq s, X_i \leq t_0, k_i = 2\}]}{\sum_{i=1}^n [I(X_i > t_0)\{\hat{H}(t_0)\}^{-1} + \Delta_i \{\hat{H}(X_i)\}^{-1} I(X_i \leq t_0, k_i = 2)]},$$

where, k_i is the observed event type for ith observation, and $\hat{D}_i^*(t_0) = \hat{P}_2(\mathbf{z}_{1i}, \mathbf{z}_{2i}; t_0) - \hat{P}_1(\mathbf{z}_{1i}; t_0)$ is the difference in the estimated cause 1 CIFs for the i-th subject. For the $\hat{F}_1(s; t_0)$, we still use the same weight as the IPW estimators except that we concentrate on the cause 1 event. $\hat{G}_1(s; t_0)$ is more complicated here because it consists of two sets of people. The first set contains those who have not had any event by time t_0. They are weighted by $\hat{H}(t_0)$ to compensate for those who are censored at or after t_0. For those who experienced cause 2 event at $X_i \leq t_0$, we use $\hat{H}(X_i)$ as their weights to account for those

who were censored at X_i. By taking expectation, we can see that the numerator of $\hat{G}_1(s;t_0)$ is the joint function of $\hat{D}^*$, T, and k, and the denominator is the joint function of T and k. Hence we can develop consistency and weak convergence of these estimators following similar arguments for the IPW estimator. The inference is based on bootstrap. In the following we refer to these estimators as the modified IPW (mIPW) estimators.

Moreover, we propose another set of estimators for the conditional distribution functions $F_1(s;t_0)$ and $G_1(s;t_0)$. Like our alternative estimators in the one-event setting, we can rewrite

$$F_1(s;t_0) = \frac{\Pr(\hat{D}^* \leq s, T \leq t_0, k=1)}{\Pr(T \leq t_0, k=1)}$$

and also

$$G_1(s;t_0) = \frac{\Pr\{\hat{D}^* \leq s, (T \leq t_0, k=1)^c\}}{\Pr\{(T \leq t_0, k=1)^c\}} = \frac{\Pr(\hat{D}^* \leq s) - \Pr(\hat{D}^* \leq s, T \leq t_0, k=1)}{1 - \Pr(T \leq t_0, k=1)}.$$

Then, applying the bivariate CIF estimator $\hat{F}_{11}$ for $F_{11}(s,t_0) = \Pr(\hat{D}^* \leq s, T \leq t_0, k=1)$ (Cheng et al., 2007) and the univariate CIF estimator $\hat{F}_T$ for $F_T(t_0) = \Pr(T \leq t_0, k=1)$, we propose the following alternative estimators:

$$\hat{F}_1'(s;t_0) = \frac{\hat{F}_{11}(s,t_0)}{\hat{F}_T(t_0)}$$

and

$$\hat{G}_1'(s;t_0) = \frac{n^{-1}\sum_{i=1}^n I\{\hat{D}_i^* \leq s\} - \hat{F}_{11}(s,t_0)}{1 - \hat{F}_T(t_0)}.$$

In the following we will refer to our alternative estimators as the CFK estimators. Due to page limit we omit the exact form of $\hat{F}_{11}$ which requires a lengthy introduction of notation. The `R` code will be available at the website mentioned earlier. The statistical inference procedure relies on the bootstrap method.

3.2.3 Evaluating the added predictive ability of new markers

3.2.3.1 Extending existing criteria After obtaining the above estimators for the conditional distributions, we compute three criteria considered in Uno et al. (2009) to evaluate the added value of new markers. For a single-event outcome evaluated at t_0, the IDI proposed in Pencina et al. (2008) corresponds to the area between the two distribution functions, $F(s;t_0)$ and $G(s;t_0)$. That is,

$$\text{IDI}\ (t_0) = \int_{-1}^{1} G(s;t_0)ds - \int_{-1}^{1} F(s;t_0)ds.$$

It can be readily estimated by plugging in the estimated functions for G and F. The integration can be carried out via trapzoid approximation or the Monte Carlo method. We note that in the absence of censoring our estimation formula for IDI reduces to a simple difference in means.

The vertical distance between two distribution functions at $s = 0$ is called improvement in the AUC (IAUC) in Uno et al. (2009). That is, IAUC=$\Pr(\hat{D} \geq 0|T \leq t_0) - \Pr(\hat{D} \geq 0|T > t_0)$. We can show that the IAUC is exactly two times the extended NRI (Pencina et al., 2011) for survival data, and can be estimated by using the empirical distribution for $\hat{D}$.

The third criterion is the difference in the medians of the two conditional distribution functions, i.e. $\text{argmin}_s\{F(s;t_0) \geq 0.5\} - \text{argmin}_s\{G(s;t_0) \geq 0.5\}$. Again, it can be estimated by plugging in the distribution estimators.

These three criteria can be analogously extended to competing-risk settings where $\hat{D}$ is replaced by $\hat{D}^*$ which is the difference in predicted risks of developing cause 1 event, and the "disease" and "healthy" groups contain those who have developed the cause 1 event by t_0, and the remaining subjects. The three extended criteria can be estimated similarly as those in the composite-event setting.

The inference procedure is based on bootstrapping. As described in Section 3.2.1, we generate B bootstrap samples and obtain B estimates of the IDI, IAUC and difference in the medians. Kerr et al. (2011) and Demler et al. (2012) pointed out that asymptotic normality may not hold for the estimated IDI or AUC if the added new biomarkers are not statistically significant. Therefore, instead of obtaining the bootstrap standard deviation based on the B

replications and constructing inference procedures based on asymptotic normality, we report the lower 2.5 percentile and upper 2.5 percentile of the B estimates as the 95% bootstrap confidence intervals in our data analysis.

3.2.3.2 Comparing density curves The above three criteria are based on the cumulative distribution functions for event and non-event populations. However, it is often not easy to visually inspect the difference in cumulative distribution functions. Hence we sample data from the estimated cumulative distribution functions $F(s;t_0)$ and $G(s;t_0)$ for a composite-event outcome, or from $F_1(s;t_0)$ and $G_1(s;t_0)$ for a competing-risk outcome, following the idea of Wei (2008). Then we display the smoothed density curves of the two distributions side by side in a single graph; see Figure 3.3 for an example. The plot often provides a better illustration of the added values of the new biomarkers in separating the event and non-event populations.

3.3 SIMULATION STUDIES

The goal of the simulation studies is to evaluate and compare the Uno estimators and our proposed Dabrowska estimators of the conditional distribution functions $F(s;t_0)$ and $G(s;t_0)$ in the composite-event setting, and the modified Uno estimator and the CFK estimator of $F_1(s;t_0)$ and $G_1(s;t_0)$ in the competing-risk setting. We first considered the composite outcome with independent censoring. In this setting, we simulated the event time from a Weibull model with three hypothetical covariates, Z_1, Z_2, and Z_3 as follows:

$$\log(T) = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \sigma W,$$

where Z_1 and Z_3 were generated from a standard normal distribution and Z_2 was generated from a Bernoulli (0.7) distribution, and W has the standard extreme value distribution. This error distribution gives the proportional hazard interpretations for all covariates. We

selected $\beta_0 = 2.25, \beta_1 = 0.05, \beta_2 = -0.05, \beta_3 = 0.25$, and $\sigma = 0.1$ so that the simulated event times are not very skewed. Note that β_3, the coefficient for the new marker, is five times of β_1 and β_2 which are the coefficients for two conventional predictors. Therefore, we expect that the new model including Z_1, Z_2, Z_3 would have improved predictive ability than the old model with Z_1 and Z_2 only. The censoring time was simulated from a uniform $[a, b]$, where a and b were selected to control the censoring rates at 10% or 50%.

For each simulated dataset containing 100 or 200 observations of $(X, \Delta, Z_1, Z_2, Z_3)$, we ran the Cox regression model without or with the new marker Z_3, upon which we computed $\hat{P}_{10}$ and $\hat{P}_{20}$, and $\hat{D} = \hat{P}_{20} - \hat{P}_{10}$ for each subject. The R function "coxph" in the package **survival** was used for the Cox model. We then obtained the Uno estimates $\hat{F}(s; t_0)$ and $\hat{G}(s; t_0)$ and the Dabrowska estimates $\hat{F}'(s; t_0)$ and $\hat{G}'(s; t_0)$ at $s = -0.1, 0, 0.1$ and $t_0 = 8, 10$, following the estimation procedures described in Section 3.2.1. We considered two sample sizes ($n = 100$ or $n = 200$) and two censoring rates (10% and 50%) and did 1,000 simulations for each of four combinations. The simulation results are summarized in Table 3.1.

The true values were computed based on a large sample ($n = 1,300,000$) without censoring. AVE and BSE are the averages of the 1,000 point estimates and bootstrap standard errors. Though the perturbation method may be used for the Uno estimator, we apply the bootstrap approach to both estimators for a fair comparison, since the Dabrowska estimator usually relies on bootstrap for inference. ESE is the empirical standard error calculated from the 1,000 point estimates. Cov is the coverage rate of 95% confidence intervals constructed based on the bootstrap standard errors and asymptotic normality. When the censoring proportion is low and $n = 100$, the two estimators perform equally well on predicting the risk probabilities, except for $G(0.1; 10)$ where the coverage rates for both estimators are as low as 70%. This is expected as for larger t_0, the number of "healthy" subjects is small. The coverages are again close to 0.95 when the sample size increases to 200. When there is heavy censoring (50%), the estimation on the boundary becomes more problematic. Both estimators have low coverage rates at $G(\cdot; 10)$. However, our proposed Dabrowska estimator seems to be more robust to extremely small number of subjects than the IPW estimator (e.g., coverages 0.64 vs. 0.31 for $G(0.1; 10)$ with $n = 100$ and 0.73 vs. 0.58 for n = 200).

Table 3.1: Simulation results comparing the inversely weighted estimators (IPW) to the Dabrowska estimators (Dab) of the conditional distributions for the outcome with independent censoring

10%		$F(-0.1;8)$		$F(0;8)$		$F(0.1;8)$		$F(-0.1;10)$		$F(0;10)$		$F(0.1;10)$	
DIM	VAR	IPW	Dab	IPW	Dab	IPW	Dab	IPW	Dab	IPW	Dab	IPW	Dab
100	True	0.12		0.17		0.22		0.12		0.16		0.21	
	AVE	0.12	0.11	0.17	0.16	0.22	0.21	0.12	0.11	0.16	0.15	0.22	0.21
	BSE	0.05	0.05	0.06	0.06	0.07	0.07	0.04	0.04	0.05	0.04	0.05	0.05
	ESE	0.04	0.04	0.05	0.05	0.06	0.06	0.04	0.04	0.04	0.04	0.06	0.06
	Cov	0.95	0.93	0.96	0.94	0.96	0.95	0.95	0.92	0.96	0.93	0.93	0.90
200	True	0.12		0.17		0.22		0.12		0.16		0.21	
	AVE	0.12	0.11	0.17	0.16	0.22	0.21	0.12	0.11	0.16	0.15	0.21	0.21
	BSE	0.04	0.04	0.04	0.04	0.05	0.05	0.03	0.03	0.03	0.03	0.04	0.04
	ESE	0.03	0.03	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.04	0.04
	Cov	0.95	0.94	0.97	0.96	0.97	0.96	0.96	0.94	0.96	0.94	0.94	0.93

10%		$G(-0.1;8)$		$G(0;8)$		$G(0.1;8)$		$G(-0.1;10)$		$G(0;10)$		$G(0.1;10)$	
DIM	VAR	IPW	Dab	IPW	Dab	IPW	Dab	IPW	Dab	IPW	Dab	IPW	Dab
100	True	0.81		0.87		0.92		0.87		0.92		0.95	
	AVE	0.81	0.81	0.87	0.87	0.92	0.92	0.87	0.87	0.92	0.92	0.96	0.96
	BSE	0.05	0.05	0.04	0.04	0.04	0.03	0.06	0.06	0.05	0.05	0.03	0.03
	ESE	0.05	0.05	0.04	0.04	0.03	0.03	0.06	0.06	0.05	0.05	0.04	0.04
	Cov	0.94	0.94	0.96	0.95	0.93	0.94	0.93	0.93	0.92	0.91	0.72	0.73
200	True	0.81		0.87		0.92		0.87		0.92		0.95	
	AVE	0.81	0.81	0.87	0.87	0.92	0.92	0.87	0.87	0.92	0.92	0.95	0.96
	BSE	0.04	0.04	0.03	0.03	0.03	0.03	0.05	0.05	0.04	0.04	0.03	0.03
	ESE	0.03	0.03	0.03	0.03	0.02	0.02	0.04	0.04	0.03	0.03	0.02	0.02
	Cov	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.95	0.95	0.92	0.92

50%		$F(-0.1;8)$		$F(0;8)$		$F(0.1;8)$		$F(-0.1;10)$		$F(0;10)$		$F(0.1;10)$	
DIM	VAR	IPW	Dab	IPW	Dab	IPW	Dab	IPW	Dab	IPW	Dab	IPW	Dab
100	True	0.12		0.17		0.22		0.12		0.16		0.21	
	AVE	0.11	0.12	0.17	0.16	0.22	0.21	0.12	0.12	0.16	0.15	0.24	0.22
	BSE	0.06	0.06	0.07	0.07	0.07	0.07	0.06	0.05	0.06	0.06	0.07	0.07
	ESE	0.05	0.05	0.06	0.06	0.07	0.07	0.05	0.06	0.06	0.07	0.09	0.10
	Cov	0.93	0.92	0.96	0.92	0.96	0.94	0.92	0.87	0.95	0.89	0.88	0.82
200	True	0.12		0.17		0.22		0.12		0.16		0.21	
	AVE	0.12	0.12	0.17	0.16	0.22	0.21	0.12	0.12	0.16	0.15	0.22	0.22
	BSE	0.04	0.04	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.05	0.05	0.05
	ESE	0.04	0.04	0.04	0.05	0.05	0.05	0.04	0.04	0.04	0.05	0.06	0.06
	Cov	0.95	0.92	0.97	0.94	0.97	0.95	0.94	0.91	0.94	0.92	0.90	0.88

50%		$G(-0.1;8)$		$G(0;8)$		$G(0.1;8)$		$G(-0.1;10)$		$G(0;10)$		$G(0.1;10)$	
DIM	VAR	IPW	Dab	IPW	Dab	IPW	Dab	IPW	Dab	IPW	Dab	IPW	Dab
100	True	0.81		0.87		0.92		0.87		0.92		0.95	
	AVE	0.81	0.80	0.88	0.88	0.92	0.92	0.87	0.87	0.92	0.92	0.96	0.96
	BSE	0.06	0.06	0.05	0.05	0.04	0.04	0.10	0.08	0.07	0.06	0.04	0.04
	ESE	0.06	0.06	0.05	0.04	0.04	0.03	0.11	0.09	0.09	0.08	0.06	0.05
	Cov	0.94	0.92	0.93	0.95	0.89	0.92	0.72	0.83	0.53	0.74	0.31	0.64
200	True	0.81		0.87		0.92		0.87		0.92		0.95	
	AVE	0.81	0.81	0.87	0.87	0.92	0.92	0.87	0.87	0.92	0.92	0.96	0.96
	BSE	0.04	0.04	0.04	0.03	0.03	0.03	0.08	0.07	0.06	0.05	0.04	0.04
	ESE	0.04	0.04	0.03	0.03	0.03	0.02	0.07	0.06	0.06	0.05	0.04	0.04
	Cov	0.96	0.95	0.95	0.96	0.94	0.94	0.92	0.94	0.80	0.87	0.58	0.73

Next, we focus on the outcome that is subject to competing-risk censoring. We used the same Weibull model to generate the event times and the covariates were the same as those in the one-event setting. The cause indicators, k, were generated by randomly assigning to one of the two causes with probability 0.5. The other simulation settings remained the same. For each simulated dataset, we first applied the multistate model to predict CIFs without or with Z_3, $\hat{P}_1$ and $\hat{P}_2$, and computed $\hat{D}^* = \hat{P}_2 - \hat{P}_1$. Then we obtained the modified inversely weighed estimates (mIPW) $\hat{F}_1(s;t_0)$ and $\hat{G}_1(s;t_0)$ and our proposed alternative estimates (CFK) $\hat{F}_1'(s;t_0)$ and $\hat{G}_1'(s;t_0)$, as we described in Section 3.2.2. The simulation results are summarized in Table 3.2.

In Table 3.2, we add the subscript 1 to all quantities to emphasize that we are only interested in the cause 1 event. Recall that $F_1(s;t_0) = \Pr(\hat{D}^* \leq s | T \leq t_0, k = 1)$ and $G_1(s;t_0) = \Pr\{\hat{D}^* \leq s | (T \leq t_0, k = 1)^c\}$. The estimation of $\hat{P}_1$ and $\hat{P}_2$ involves running the Cox model twice for the two causes. Due to the memory limitation in R, the true values in Table 3.2 were computed based on a sample of size $500,000$. Like Table 3.1, the top two panels in Table 3.2 show the estimates of $F_1(s;t_0)$ and $G_1(s;t_0)$ at $s = -0.1, 0, 0.1$ and $t_0 = 8, 10$ for a sample of 100 or 200 with 10% censoring. The bottom two panels are for the 50% censoring. Since we have roughly half the cause 1 events as compared to the overall number of events, in some cells the coverage rates are not satisfactory. However, the coverage rates generally increase with the increase in the sample size and the decrease in the censoring proportion, except for $F_1(0.1;10)$ and $G_1(0.1;10)$, where the "true" values may not be accurate due to the relatively small sample in computing the true values. In general, the two sets of estimators perform well with moderate sample size and light censoring.

3.4 A BREAST CANCER STUDY

We applied the estimators of the conditional distribution functions and summary measures to the breast cancer study data used in Uno et al. (2009). The original gene expression

Table 3.2: Simulation results comparing the modified inverse weighted estimators (mIPW) and the alternative estimators (CFK) in a competing-risk setting

10%		$F_1(-0.1;8)$		$F_1(0;8)$		$F_1(0.1;8)$		$F_1(-0.1;10)$		$F_1(0;10)$		$F_1(0.1;10)$	
DIM	VAR	mIPW	CFK	mIPW	CFK	mIPW	CFK	mIPW	CFK	mIPW	CFK	mIPW	CFK
100	True	0.07		0.17		0.28		0.09		0.16		0.32	
	AVE	0.07	0.07	0.17	0.17	0.28	0.28	0.09	0.09	0.17	0.17	0.37	0.37
	BSE	0.05	0.05	0.08	0.08	0.11	0.11	0.05	0.05	0.07	0.07	0.08	0.08
	ESE	0.05	0.05	0.08	0.08	0.10	0.10	0.05	0.05	0.06	0.06	0.13	0.13
	Cov	0.73	0.73	0.93	0.93	0.94	0.95	0.92	0.92	0.94	0.94	0.80	0.81
200	True	0.07		0.17		0.28		0.09		0.16		0.32	
	AVE	0.07	0.07	0.16	0.16	0.28	0.28	0.09	0.09	0.16	0.16	0.36	0.36
	BSE	0.04	0.04	0.06	0.06	0.07	0.07	0.03	0.04	0.05	0.05	0.06	0.06
	ESE	0.04	0.04	0.05	0.05	0.07	0.07	0.03	0.03	0.04	0.04	0.10	0.10
	Cov	0.92	0.92	0.95	0.95	0.96	0.96	0.94	0.95	0.97	0.97	0.74	0.74

10%		$G_1(-0.1;8)$		$G_1(0;8)$		$G_1(0.1;8)$		$G_1(-0.1;10)$		$G_1(0;10)$		$G_1(0.1;10)$	
DIM	VAR	mIPW	CFK	mIPW	CFK	mIPW	CFK	mIPW	CFK	mIPW	CFK	mIPW	CFK
100	True	0.55		0.70		0.79		0.42		0.51		0.63	
	AVE	0.51	0.51	0.71	0.71	0.80	0.80	0.43	0.43	0.54	0.54	0.70	0.70
	BSE	0.06	0.05	0.05	0.05	0.05	0.04	0.06	0.06	0.06	0.06	0.06	0.06
	ESE	0.12	0.12	0.05	0.05	0.04	0.04	0.07	0.08	0.07	0.07	0.06	0.07
	Cov	0.70	0.69	0.94	0.94	0.94	0.94	0.89	0.88	0.90	0.90	0.77	0.75
200	True	0.55		0.70		0.79		0.42		0.51		0.63	
	AVE	0.53	0.53	0.71	0.71	0.79	0.79	0.43	0.43	0.52	0.52	0.68	0.68
	BSE	0.04	0.04	0.04	0.04	0.03	0.03	0.05	0.04	0.05	0.04	0.04	0.04
	ESE	0.07	0.08	0.04	0.04	0.03	0.03	0.06	0.06	0.05	0.05	0.05	0.05
	Cov	0.72	0.72	0.92	0.93	0.94	0.94	0.89	0.88	0.91	0.89	0.74	0.74

50%		$F_1(-0.1;8)$		$F_1(0;8)$		$F_1(0.1;8)$		$F_1(-0.1;10)$		$F_1(0;10)$		$F_1(0.1;10)$	
DIM	VAR	mIPW	CFK	mIPW	CFK	mIPW	CFK	mIPW	CFK	mIPW	CFK	mIPW	CFK
100	True	0.07		0.17		0.28		0.09		0.16		0.32	
	AVE	0.07	0.07	0.17	0.16	0.28	0.28	0.09	0.08	0.17	0.17	0.36	0.35
	BSE	0.05	0.05	0.09	0.09	0.12	0.12	0.06	0.06	0.09	0.09	0.10	0.11
	ESE	0.06	0.06	0.08	0.09	0.11	0.11	0.06	0.06	0.08	0.08	0.15	0.14
	Cov	0.65	0.65	0.92	0.92	0.95	0.95	0.78	0.79	0.91	0.90	0.83	0.85
200	True	0.07		0.17		0.28		0.09		0.16		0.32	
	AVE	0.07	0.07	0.17	0.17	0.28	0.28	0.09	0.09	0.17	0.16	0.37	0.37
	BSE	0.04	0.04	0.07	0.07	0.08	0.08	0.05	0.05	0.06	0.06	0.08	0.08
	ESE	0.04	0.04	0.06	0.06	0.08	0.08	0.04	0.04	0.06	0.06	0.12	0.12
	Cov	0.88	0.88	0.95	0.95	0.96	0.96	0.91	0.90	0.96	0.96	0.76	0.80

50%		$G_1(-0.1;8)$		$G_1(0;8)$		$G_1(0.1;8)$		$G_1(-0.1;10)$		$G_1(0;10)$		$G_1(0.1;10)$	
DIM	VAR	mIPW	CFK	mIPW	CFK	mIPW	CFK	mIPW	CFK	mIPW	CFK	mIPW	CFK
100	True	0.55		0.70		0.79		0.42		0.51		0.63	
	AVE	0.49	0.49	0.71	0.71	0.80	0.81	0.42	0.43	0.55	0.55	0.71	0.71
	BSE	0.06	0.06	0.06	0.05	0.05	0.05	0.10	0.07	0.09	0.08	0.07	0.08
	ESE	0.13	0.13	0.06	0.06	0.05	0.05	0.11	0.11	0.09	0.10	0.08	0.10
	Cov	0.66	0.59	0.92	0.90	0.90	0.90	0.89	0.79	0.89	0.83	0.76	0.74
200	True	0.55		0.70		0.79		0.42		0.51		0.63	
	AVE	0.52	0.52	0.71	0.70	0.79	0.79	0.43	0.43	0.53	0.53	0.69	0.69
	BSE	0.05	0.04	0.04	0.04	0.04	0.03	0.07	0.05	0.06	0.05	0.05	0.05
	ESE	0.09	0.09	0.05	0.04	0.04	0.03	0.08	0.07	0.07	0.07	0.06	0.08
	Cov	0.70	0.63	0.91	0.90	0.92	0.93	0.89	0.82	0.90	0.85	0.71	0.70

data were collected at the Netherlands Cancer Institute by van't Veer et al. (2002) and van de Vijver et al. (2002). The data contain the clinical and demographic information of 295 breast cancer patients including age, tumor diameter, number of positive lymph nodes, tumor grade, vascular invasion, estrogen receptor status, chemo/hormonal therapy or not, and mastectomy. These conventional clinical markers were collected at baseline and used to predict patients' risk of metastasis or death. Chang et al. (2005) developed a new biomarker, gene score, from the original microarray gene expression data. Uno et al. (2009) evaluated the added value of the gene score in predicting risks of metastasis or death. Here the outcome is time to either metastasis or death which is subject to random censoring.

We first replicated Uno et al. (2009)'s results on the breast cancer data by using their estimators of the conditional distribution functions, $\hat{F}(s;10)$ and $\hat{G}(s;10)$. Then we applied our alternative estimators, $\hat{F}'(s;10)$ and $\hat{G}'(s;10)$, based on the Dabrowska estimator for bivariate survival function. These two sets of estimates for the conditional distributions are given in Figure 3.1. The left panel shows the Uno estimates of the distribution functions. The dark solid line represents $\hat{F}(s;10)$, the distribution function of $\hat{D}$ conditional on that patients have experienced the outcome event by 10 years, either metastasis or death. The dashed line is for $\hat{G}(s;10)$, the distribution function conditional on that patients have not experienced any event. The added value of gene score was evaluated using the three criteria, the IDI, IAUC, and the difference in the medians of the two distributions. In the left panel of Figure 3.1, the area between the two distributions 0.05 is the estimated IDI, the vertical distance between the two gray dots is the estimated IAUC, which is 0.27, and the horizontal distance between the two dark dots (medians) is 0.07. We also ran 1,000 bootstrap samples to obtain 95% bootstrap confidence intervals (0.02, 0.08) for the IDI, (0.12, 0.42) for the IAUC and (0.02, 0.09) for the difference in the medians. The right panel of Figure 3.1 gives our alternative estimates for the conditional distribution functions. Upon close inspection, the two estimates of $F(s;10)$ are almost identical, and our alternative estimate of $G(s;10)$ is slightly lower and more smoother than Uno's estimate. Therefore, we obtained sightly lower values for the three criteria. The IDI index is 0.04 with a 95% bootstrap confidence interval (0.02, 0.07), the IAUC is 0.20 with a 95% confidence interval (0.07, 0.35) and the difference in

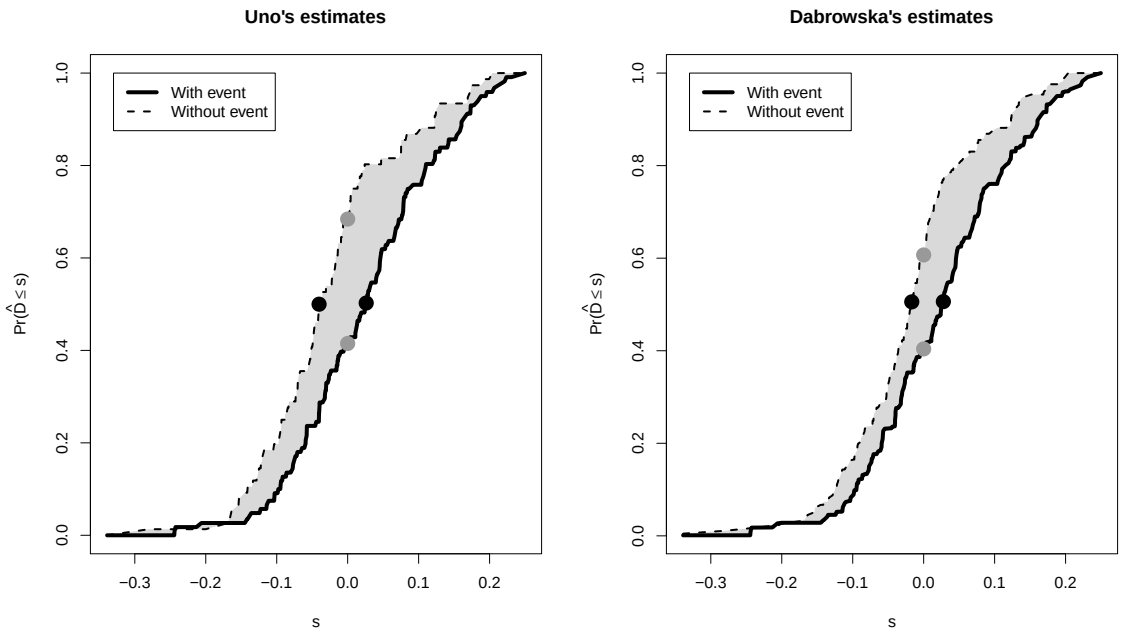

Figure 3.1: Breast Cancer Study, one-event setting

two medians is 0.04 with a 95% confidence interval (0.01, 0.07). However, none of the three confidence intervals contains zero, indicating that the gene score does significantly improve the predictive ability of the model.

Metastasis and death are actually two competing events. In the study, 101 patients had metastasis, 5 patients died without metastasis, and the remaining 189 patients survived by the end of study without metastasis. When we focused our interest on the event of metastasis, which was subject to dependent censoring by death, we applied our modified Uno estimators and the CFK estimators to evaluate the improvement on predicting the risk of metastasis using the gene score. However, in this application, the multistate model did not provide good estimates for P_1 and P_2, since the Cox model for death did not fit well with only five events and multiple covariates. To address this issue, we actually modeled the cause 1 CIF directly using the semiparametric model of Fine and Gray (1999). The R function "crr" from the package **cmprsk** was used and the regression coefficients from the old and new models are summarized in Table 3.3. The values of these estimators are quite

Table 3.3: Estimates of regression coefficients for Fine and Gray (1999)'s models with breast cancer study

	Model without gene score			Model with gene score		
	Est.	SE	p	Est.	SE	p
Age/10[yrs]	-0.50	0.19	0.01	-0.59	0.19	0.002
Diameter of tumor [cm]	0.20	0.12	0.09	0.20	0.12	0.10
Lymph nodes	0.01	0.08	0.88	0.00	0.09	0.99
Grade = 2 vs 1	0.91	0.33	0.01	0.69	0.35	0.05
Grade = 3 vs 1	1.13	0.34	<0.001	0.73	0.38	0.06
Vascular invasion 1-3 vs 0	0.13	0.43	0.77	-0.02	0.43	0.96
Vascular invasion >3 vs 0	0.81	0.67	0.23	0.68	0.68	0.32
Estrogen Status=Positive	-0.20	0.27	0.45	0.01	0.27	0.96
Chemo or Hormonal=Yes	-0.55	0.38	0.15	-0.51	0.39	0.19
Mastectomy=Yes	0.11	0.23	0.61	0.17	0.23	0.45
Gene score	-	-	-	2.17	0.70	0.002

similar to those from the Cox regression model in Uno et al. (2009).

The estimates of the conditional distributional functions of $\hat{D}^*$ given metastasis or free of metastasis are showed in Figure 3.2. The left panel represents the modified Uno estimates, $\hat{F}_1(s;10)$ and $\hat{G}_1(s;10)$. The area between two estimated distribution functions, i.e. IDI, is 0.04 with a 95% bootstrap confidence interval (0.02, 0.07), the IAUC is 0.24 with a 95% confidence interval (0.10, 0.37), and the difference in two medians is 0.06 with a 95% confidence interval (0.01, 0.08). The right plot is for the CFK estimates of the conditional distribution functions, $\hat{F}_1'(s;10)$ and $\hat{G}_1'(s;10)$. The IDI here is 0.03 with a 95% confidence interval (0.002, 0.06), the IAUC is 0.15 with a 95% confidence interval (-0.005, 0.29), and the difference in medians is 0.04 with a 95% confidence interval (-0.007, 0.06). Compared to modified Uno's estimates, our alternative estimates are slightly closer to zero. However, the IDI is still significantly greater than zero, and the lower bounds of the other two are very close to zero. In summary, the three summary criteria consistently show that adding the gene score into the model along with conventional markers would improve the prediction of risks of metastasis for breast cancer patients.

It is more clearly seen in Figure 3.3. For those subjects who developed metastasis before

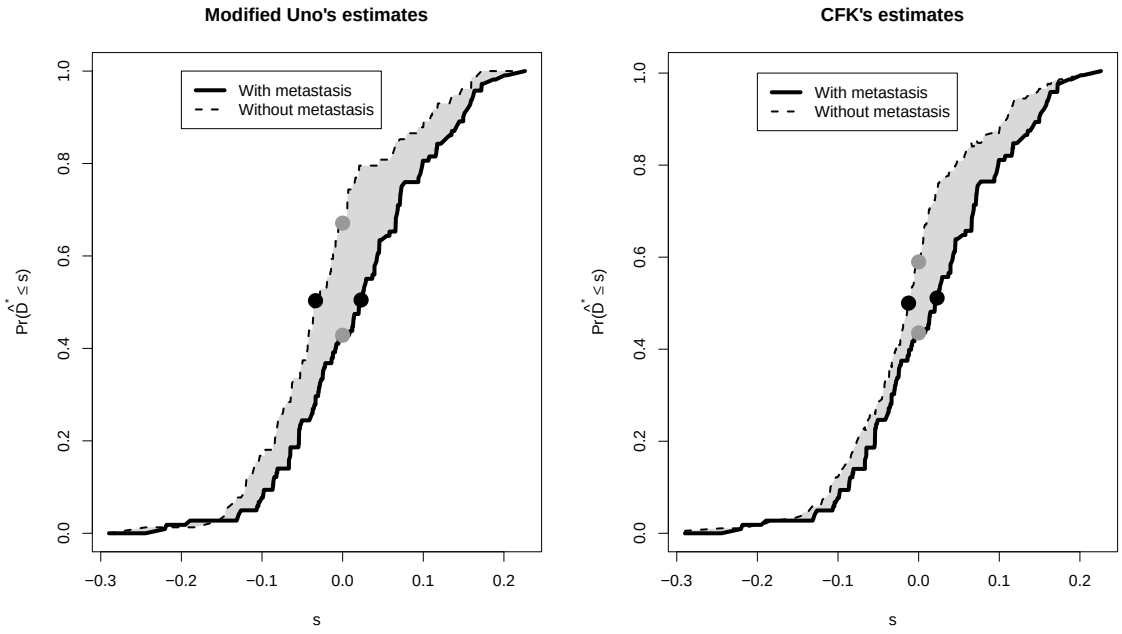

Figure 3.2: Breast Cancer Study, competing-risk setting

or at 10 years, the density of the difference in the predicted risk of metastasis with and without the gene score is more concentrated on $s > 0$. Hence, the inclusion of the gene score in the prediction model improves the prediction of metastasis. In contrast, the density of the difference is more clustered for $s < 0$ among those who did not have metastasis. That is, the inclusion of the gene score in the prediction model improves the prediction of no metastasis in this subpopulation. The two density curves are reasonably separated out.

3.5 REMARKS

For simulation studies, we used the multistate model to estimate the cause 1 CIFs given conventional predictors and new biomarkers P_1 and P_2. However, this model may not work when there are only few events from the competing cause as in the breast cancer study.

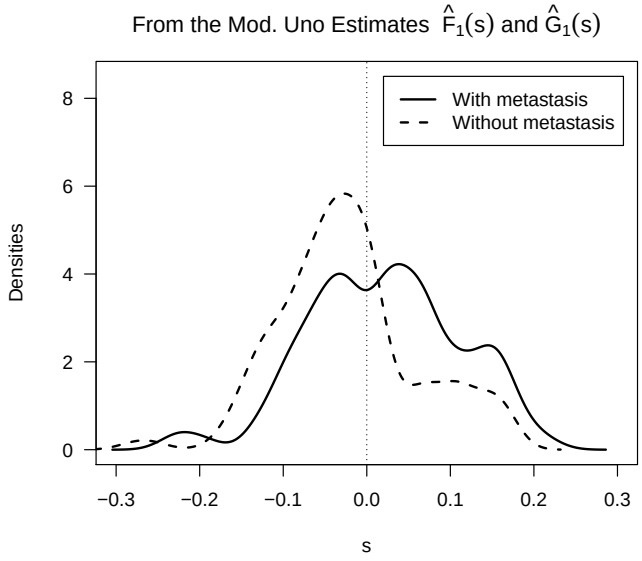

Figure 3.3: Density plots generated from sampling from the estimated F_1 and G_1 functions

In contrast, the Fine and Gray (1999) model focuses on the cause 1 event only and avoids modeling the competing event. However, our experience suggests that the Fine and Gray (1999) model may not be computationally efficient for a large sample due to the need of constructing an n by n covariance matrix. Therefore, we rely on the Cox regression model to obtain the predicted risks, especially for obtaining the "true" values reported in Tables 3.1 and 3.2 from very large datasets. There are other regression models on CIFs such as Fine (2001), Klein and Andersen (2005), Jeong and Fine (2007), Scheike et al. (2008) and Shi et al. (2012) among others. The estimations of P_1 and P_2 will definitely affect the evaluations of added values of the new biomarkers. However, they are not the focus of our research. Similarly, for the composite outcome, we have been using the Cox model which may not be proper if the proportionality assumption does not hold. Stratified Cox models or piecewise Cox models may be used instead. We suggest using a well-calibrated regression model to estimate P_1 and P_2, or P_{10} and P_{20}, and performing model diagnosis before proceeding to the proposed discrimination analysis.

It is worth pointing out that the discrimination ability of the new biomarkers may differ at various time points. In our data analysis, we have been focusing on predicting metastasis

or composite event at ten years. It may be of interest to evaluate the accuracy improvement at some different time points or across a range of time points. How to combine accuracy improvement measures across times may be a topic of future research.

Finally we wish to caution practitioners about the use of the proposed accuracy measures when the increase of AUC is small. Even though AUC increment may be limited by its construction to detect further accuracy improvement beyond certain accuracy level, there may be other practical reasons. For example, if the model has already contained some good predictors, additional markers may only provide better prediction for a sub-population. Requiring more tests than are needed may lead to major increase in medical testing costs, and more iatrogenic damage, since many of these tests (e.g., arteriography or biopsy) involve risks to the patients. Under such situations, the crucial issue is to define what tests are essential for which patients, not to seek methods that simply increase the number of tests for minimal benefit.

APPENDIX

Table A1 summarizes the simulation results for the four simulation scenarios when the sample size is n = 500 and the censoring rate is 40%.

Table A2 summarizes the simulation results of how the censoring distribution affects the prediction of CIFs. We focused on our parametric model and the Fine and Gray method, and simulated censoring times from the log logistic model. That is, $\log(C) = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \sigma\epsilon$, where γ_0 is a constant and chosen so that the censoring rate is around 20%, $\gamma_1 = \gamma_2 = 0.5$, $\sigma = 0.5$, and $\epsilon \sim$ standard logistic distribution. Therefore, the censoring time depends on covariates in a nonproportional subdistribution hazard fashion.

Table A1: Simulation results where the data were simulated from our proposed modified logistic (panel LOG + PSH) or Gompertz model (panel GOM + PSH) with complimentary log-log transformation or with generalized-odds rate transformation (panels LOG + GOR and GOM + GOR) with sample size n=500 and 40% censoring rate; referring to Table 2.1 for the definition of AVE, MSE, ESE, and Cov

LOG + PH		$\hat{\beta}_{11}$			$\hat{\beta}_{12}$			$\hat{F}_1(3)$				$\hat{F}_1(5)$			
DIM	VAR	Log	Gom	FG	Log	Gom	FG	Log	Gom	FG	Sch	Log	Gom	FG	Sch
500	True	0.50	0.50	0.50	0.50	0.50	0.50	0.32	0.32	0.32	0.32	0.42	0.42	0.42	0.42
	AVE	0.50	-	0.50	0.50	-	0.50	0.32	-	0.32	0.32	0.43	-	0.42	0.43
	MSE	0.09	-	0.09	0.09	-	0.09	0.05	-	0.05	0.15	0.06	-	0.07	0.15
	ESE	0.09	-	0.09	0.09	-	0.09	0.05	-	0.06	0.06	0.06	-	0.07	0.09
	Cov	0.95	-	0.95	0.95	-	0.95	0.95	-	0.94	1.00	0.94	-	0.94	0.99

GOM + PH		$\hat{\beta}_{11}$			$\hat{\beta}_{12}$			$\hat{F}_1(1)$				$\hat{F}_1(5)$			
DIM	VAR	Log	Gom	FG	Log	Gom	FG	Log	Gom	FG	Sch	Log	Gom	FG	Sch
500	True	0.50	0.50	0.50	0.50	0.50	0.50	0.51	0.51	0.51	0.51	0.56	0.56	0.56	0.56
	AVE	0.52	0.52	0.50	0.52	0.52	0.50	0.53	0.52	0.51	0.51	0.60	0.58	0.53	0.53
	MSE	0.08	0.08	0.09	0.08	0.08	0.09	0.07	0.06	0.08	0.13	0.08	0.07	0.08	0.14
	ESE	0.08	0.11	0.09	0.09	0.11	0.09	0.07	0.09	0.08	0.09	0.08	0.09	0.08	0.11
	Cov	0.94	0.88	0.95	0.93	0.88	0.94	0.92	0.87	0.94	0.95	0.89	0.86	0.90	0.97

LOG + GOR		$\hat{\beta}_{11}$			$\hat{\beta}_{12}$			$\hat{F}_1(3)$				$\hat{F}_1(5)$			
DIM	VAR	Log	Gom	FG	Log	Gom	FG	Log	Gom	FG	Sch	Log	Gom	FG	Sch
500	True	0.50	0.50	0.50	0.50	0.50	0.50	0.26	0.26	0.26	0.26	0.34	0.34	0.34	0.34
	AVE	0.52	-	0.27	0.52	-	0.26	0.27	-	0.26	0.26	0.34	-	0.34	0.34
	MSE	0.21	-	0.09	0.21	-	0.09	0.05	-	0.05	0.18	0.05	-	0.06	0.18
	ESE	0.21	-	0.09	0.21	-	0.09	0.05	-	0.05	0.05	0.05	-	0.06	0.08
	Cov	0.94	-	0.26	0.95	-	0.25	0.94	-	0.93	1.00	0.95	-	0.94	1.00

GOM + GOR		$\hat{\beta}_{11}$			$\hat{\beta}_{12}$			$\hat{F}_1(1)$				$\hat{F}_1(5)$			
DIM	VAR	Log	Gom	FG	Log	Gom	FG	Log	Gom	FG	Sch	Log	Gom	FG	Sch
500	True	0.50	0.50	0.50	0.50	0.50	0.50	0.41	0.41	0.41	0.41	0.45	0.45	0.45	0.45
	AVE	0.52	0.51	0.24	0.52	0.51	0.24	0.41	0.41	0.41	0.41	0.45	0.45	0.43	0.42
	MSE	0.20	0.20	0.09	0.20	0.20	0.09	0.05	0.05	0.07	0.14	0.06	0.06	0.07	0.15
	ESE	0.20	0.20	0.09	0.20	0.19	0.09	0.05	0.05	0.07	0.07	0.06	0.06	0.07	0.09
	Cov	0.94	0.93	0.18	0.94	0.94	0.17	0.94	0.94	0.95	0.99	0.94	0.94	0.92	0.99

Table A2: Simulation results on censoring time following a proportional odds model; the data were simulated from the modified logistic or Gompertz base with a proportional subdistribution hazard transformation (LOG+PSH and GOM+PSH) or with a generalized odds rate transformation (LOG + GOR and GOM + GOR); referring to Table 2.1 for the definition of AVE, MSE, ESE, and Cov

		$\hat{F}_1(3)$		$\hat{F}_1(5)$		$\hat{F}_2(3)$		$\hat{F}_2(5)$	
LOG+PSH	VAR	Log	FG	Log	FG	Log	FG	Log	FG
DIM	True	0.32	0.32	0.42	0.42	0.07	0.07	0.49	0.49
100	AVE	0.33	0.33	0.44	0.43	0.06	0.06	0.48	0.49
	MSE	0.11	0.12	0.13	0.14	0.02	0.04	0.12	0.15
	ESE	0.12	0.14	0.14	0.16	0.02	0.05	0.12	0.14
	Cov	0.92	0.89	0.91	0.88	0.91	0.82	0.91	0.93
200	AVE	0.33	0.33	0.43	0.43	0.06	0.06	0.48	0.49
	MSE	0.08	0.09	0.09	0.10	0.02	0.03	0.09	0.11
	ESE	0.08	0.10	0.10	0.11	0.02	0.04	0.09	0.10
	Cov	0.93	0.91	0.93	0.92	0.92	0.82	0.94	0.96
		$\hat{F}_1(1)$		$\hat{F}_1(5)$		$\hat{F}_2(1)$		$\hat{F}_2(5)$	
GOM+PSH	VAR	Log	FG	Log	FG	Log	FG	Log	FG
DIM	True	0.51	0.51	0.56	0.56	0.31	0.31	0.44	0.44
100	AVE	0.52	0.52	0.57	0.57	0.32	0.32	0.43	0.44
	MSE	0.13	0.14	0.14	0.14	0.10	0.13	0.14	0.15
	ESE	0.13	0.15	0.14	0.15	0.11	0.13	0.14	0.17
	Cov	0.91	0.91	0.90	0.90	0.91	0.93	0.90	0.89
200	AVE	0.52	0.52	0.57	0.57	0.33	0.32	0.43	0.44
	MSE	0.09	0.10	0.10	0.10	0.07	0.09	0.10	0.12
	ESE	0.10	0.10	0.10	0.11	0.08	0.11	0.10	0.17
	Cov	0.93	0.93	0.92	0.92	0.93	0.94	0.92	0.91
		$\hat{F}_1(3)$		$\hat{F}_1(5)$		$\hat{F}_2(3)$		$\hat{F}_2(5)$	
LOG+GOR	VAR	Log	FG	Log	FG	Log	FG	Log	FG
DIM	True	0.26	0.26	0.34	0.34	0.08	0.08	0.57	0.57
100	AVE	0.27	0.27	0.35	0.35	0.07	0.08	0.56	0.57
	MSE	0.10	0.11	0.11	0.13	0.02	0.04	0.10	0.15
	ESE	0.11	0.13	0.12	0.15	0.02	0.05	0.11	0.14
	Cov	0.89	0.89	0.91	0.90	0.91	0.87	0.92	0.93
200	AVE	0.27	0.27	0.35	0.35	0.08	0.08	0.56	0.57
	MSE	0.07	0.08	0.08	0.09	0.02	0.03	0.07	0.11
	ESE	0.08	0.09	0.08	0.11	0.02	0.06	0.07	0.11
	Cov	0.93	0.91	0.93	0.91	0.93	0.90	0.93	0.95
		$\hat{F}_1(1)$		$\hat{F}_1(5)$		$\hat{F}_2(1)$		$\hat{F}_2(5)$	
GOM+GOR	VAR	Log	FG	Log	FG	Log	FG	Log	FG
DIM	True	0.41	0.41	0.45	0.45	0.40	0.40	0.55	0.55
100	AVE	0.42	0.41	0.45	0.45	0.40	0.42	0.54	0.54
	MSE	0.11	0.13	0.12	0.14	0.09	0.13	0.12	0.16
	ESE	0.11	0.14	0.11	0.15	0.08	0.14	0.11	0.17
	Cov	0.95	0.90	0.95	0.89	0.95	0.93	0.95	0.90
200	AVE	0.41	0.41	0.45	0.45	0.41	0.41	0.55	0.56
	MSE	0.08	0.09	0.08	0.10	0.06	0.10	0.08	0.12
	ESE	0.08	0.10	0.08	0.11	0.06	0.12	0.08	0.15
	Cov	0.95	0.93	0.95	0.93	0.95	0.94	0.95	0.91

BIBLIOGRAPHY

Andersen, P. K., S. Rosthoj, S. Rosthoej, S. Z. Abildstrom, and S. Rosthøj (2002). Competing risks as a multi-state model. *Statistical Methods in Medical Research 11*(2), 203–215.

Beyersmann, J., A. Allignol, and M. Schumacher (2012). *Competing Risks and Multistate Models with R.* New York: Springer.

Chang, H., D. Nuyten, J. Sneddon, T. Hastie, R. Tibshirani, T. Sorlie, H. Dai, Y. He, L. van't Veer, H. Bartelink, M. van de Rijn, P. Brown, and M. van de Vijver (2005). Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *PNAS 102*(10), 3738–3743.

Cheng, S. C., J. P. Fine, and L. J. Wei (1998). Prediction of cumulative incidence function under the proportional hazards model. *Biometrics 54*, 219–228.

Cheng, Y. (2009). Modeling cumulative incidences of dementia and dementia-free death using a novel three-parameter logistic function. *The International Journal of Biostatistics 5*.

Cheng, Y., J. P. Fine, and M. R. Kosorok (2007). Nonparametric analysis of bivariate competing risks data. *Journal of the American Statistical Association 102*, 1407–1416.

Chi, Y.-Y. and X.-H. Zhou (2008). The need for reorientation toward cost-effective prediction: Comments on "Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond". *Statistics in Medicine 27*(2), 182–184.

Cook, N. R. (2008). Comments on "Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond". *Statistics in Medicine 27*(2), 191–195.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B: Methodological 34*, 187–220.

Dabrowska, D. M. (1988). Kaplan Meier estimate on the plane. *The Annals of Statistics 16*, 1475–1489.

Dabrowska, D. M. and A. Doksum, K (1998). Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association 83*, 744–749.

Demler, O., M. Pencina, et al. (2012). Misuse of delong test to compare aucs for nested models. *Statistics in Medicine 31*, 2577–2587.

Efron, B. (2010). Correlated z-values and the accuracy of large-scale statistical estimates. *Journal of the American Statistical Association 105*, 1042–1055.

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

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

Foucher, Y., M. Giral, J. P. Soulilloua, and J. P. Dauresb (2011). Time-dependent roc analysis for a three-class prognostic with application to kidney transplantation. *Statistics in Medicine*, DOI: 10.1002/sim.4052.

Gaynor, J. J., E. J. Feuer, C. C. Tan, D. H. Wu, C. R. Little, D. J. Straus, B. D. Clarkson, and M. F. Brennan (1993). On the use of cause-specific failure and conditional failure probabilities: Examples from clinical oncology data. *Journal of the American Statistical Association 88*, 400–409.

Grambauer, N., M. Schumacher, and J. Beyersmann (2010). Proportional subdistribution hazards modeling offers a summary analysis, even if misspecified. *Statistics in Medicine 29*(7–8), 875–884.

Gray, R. J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics 16*, 1140–1154.

Greenland, P. (2008). Comments on "Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond". *Statistics in Medicine 27*(2), 188–190.

Greenland, P. and P. G. O'Malley (2005). When is a new prediction marker useful? A consideration of lipoprotein-associated phospholipase a2 and c-reactive protein for stroke risk. *Archives of Internal Medicine 165*(21), 2454–2456.

Heagerty, P., T. Lumley, and M. S. Pepe (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics 56*, 337–344.

Heagerty, P. and Y. Zheng (2005). Survival model predictive accuracy and ROC curves. *Biometrics 61*, 92–105.

Hyun, S., Y. Sun, and R. Sundaram (2009). Assessing cumulative incidence functions under the semiparametric additive risk model. *Statistics in Medicine 28*(22), 2748–2768.

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

Jeong, J. and J. P. Fine (2007). Parametric regression on cumulative incidence function. *Biostatistics 55*, 184–200.

Jin, Z., D. Y. Lin, L. J. Wei, and Z. Ying (2003). Rank-based inference for the accelerated failure time model. *Biometrika 90*(2), 341–353.

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

Kerr, K., R. McClelland, E. Brown, and T. Lumley (2011). Evaluating the incremental value of new biomarkers with integrated discrimination improvement. *American Journal of Epidemiology 174*, 364–374.

Klein, J. P. (2006). Modelling competing risks in cancer stuties. *Statistics in Medicine 25*(6), 1015–1034.

Klein, J. P. and P. K. Andersen (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics 61*(1), 223–229.

Koller, M., H. Raatz, E. Steyerberg, and M. Wolbers (2012). Competing risks and the clinical community: irrelevance or ignorance? *Statistics in Medicine 31*, 1089–1097.

Korn, E. L. and F. J. Dorey (1992). Applications of crude incidence curves. *Statistics in Medicine 11*, 813–829.

Kraemer, H. C. (2008). Comments on "Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond". *Statistics in Medicine 27*(2), 196–198.

Latouche, A., V. Boisson, S. Chevret, and R. Porcher (2007). Misspecified regression model for the subdistribution hazard of a competing risk. *Statistics in Medicine 26*(5), 965–974.

Li, J. and S. Ma (2011). Time-dependent roc analysis under diverse censoring patterns. *Statistics in Medicine 30*, 1266–1277.

Pencina, M., R. D'Agostino Sr, and E. Steyerberg (2011). Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine 30*, 11–21.

Pencina, M. J., R. B. D'Agostino, Sr., R. B. D'Agostino, Jr., and R. S. Vasan (2008). Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine 27*(2), 157–172.

Pepe, M. S., Z. Feng, and J. W. Gu (2008). Comments on "Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond". *Statistics in Medicine 27*(2), 173–181.

Pepe, M. S., H. Janes, G. Longton, W. Leisenring, and P. Newcomb (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology 159*(9), 882–890.

Pepe, M. S. and M. Mori (1993). Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine 12*, 737–751.

Pepe, M. S., Y. Zheng, Y. Jin, Y. Huang, C. R. Parikh, and W. C. Levy (2008). Evaluating the ROC performance of markers for future events. *Lifetime Data Analysis 14*(1), 86–113.

Pfeiffer, R. and M. Gail (2011). Two criteria for evaluating risk prediction models. *Biometrics 67*, 1057–1065.

Prentice, R. L., J. D. Kalbfleisch, A. V. Peterson, N. Flournoy, V. T. Farewell, and N. E. Breslow (1978). The analysis of failure time data in the presence of competing risks. *Biometrics 12*, 737–751.

Saha, P. and P. J. Heagerty (2010). Time-dependent predictive accuracy in the presence of competing risks. *Biometrics 66*, 999 – 1011.

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

Scheike, T. H. and M.-J. Zhang (2011). Analyzing competing risk data using the R timereg package. *Journal of Statistical Software 38*(2), 1–15.

Scheike, T. H., M.-J. Zhang, and T. A. Gerds (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika 95*(1), 205–220.

Shi, H., Y. Cheng, and J.-H. Jeong (2012). Parametric joint modeling of covariate effects on all cause-specific cumulative incidence functions. *Biometrical Journal*. tentatively accepted.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *PNAS 72*, 20–22.

Uno, H., L. Tian, T. Cai, I. S. Kohane, and L. J. Wei (2009). Comparing risk scoring systems beyond the ROC paradigm in survival analysis. Harvard University Biostatistics Working Paper Series. Working Paper 107. http://biostats.bepress.com/harvardbiostat/paper107.

van de Vijver, M., Y. He, L. vant Veer, H. Dai, A. Hart, D. Voskuil, G. Schreiber, J. Peterse, C. Roberts, M. Marton, M. Parrish, D. Atsma, A. Witteveen, A. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E. Rutgers, S. Friend, and R. Bernards (2002). A gene-expression signature as a predictor of survival in breast cancer. *The New England Journal of Medicine 347*(25), 1999–2009.

van't Veer, L., H. Dai, M. van de Vijver, Y. He, A. Hart, M. Mao, H. Peterse, K. van der Kooy, M. Marton, A. Witteveen, G. Schreiber, R. Kerkhoven, C. Roberts, P. Linsley,

R. Bernards, and S. Friend (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature 415*(6871), 530–536.

Ware, J. H. (2006). The limitations of risk factors as prognostic tools. *New England Journal of Medicine 355*, 2615–2617.

Ware, J. H. and T. Cai (2008). Comments on "Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond". *Statistics in Medicine 27*(2), 185–187.

Wei, L. J., Z. Ying, and D. Y. Lin (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika 77*, 845–851.

Wei, Y. (2008). An Approach to Multivariate Covariate-Dependent Quantile Contours With Application to Bivariate Conditional Growth Charts. *Journal of the American Statistical Association 103*, 397–409.

Zheng, Y., T. Cai, Y. Jin, and Z. Feng (2011). Evaluating prognostic accuracy of biomarkers under competing risk. Biometrics, in press.

Zhou, B., A. Latouche, V. Rocha, and J. Fine (2011). Competing risks regression for stratified data. *Biometrics 67*(2), 661–670.