Adjusted Cumulative Incidence Curves for Competing Risks Data
Hage, P. van

Citation

Version: Not Applicable (or Unknown)
License: License to inclusion and publication of a Bachelor or Master thesis in the Leiden University Student Repository
Downloaded from: https://hdl.handle.net/1887/3676760

Note: To cite this publication please use the final published version (if applicable).
Mathematical Institute

Master Thesis

Statistical Science for the Life and Behavioural Sciences

Adjusted Cumulative Incidence Curves for Competing Risks Data

Author: Patrick van Hage
s1546732

Thesis Advisor: Prof. Dr. Hein Putter
Leiden University Medical Center

Thesis Co-Advisor: Dr. Ir. Nan van Geloven
Leiden University Medical Center

Submitted on October 4th, 2021
Foreword

First, the author of this thesis would like to express their immense gratitude to Prof. dr. Hein Putter & Dr. ir. Nan van Geloven for their supervision of this project. Their invaluable guidance and support enabled the author to develop covariate adjustment methods for competing risks in R. In addition to that, the weekly discussion sessions greatly helped to refine the execution and evaluation process of these methods. These methods may be applied or expanded upon in future clinical or methodological papers, where they can be used to visually examine and report cumulative incidences in the presence of covariate imbalance, such as present in observational data.

Furthermore, the author would like to thank the members of the Netherlands Comprehensive Cancer Organisation for being granted access to their collection of data from the Netherlands Cancer Registry. Observational data from patient records have been anonymized for the purposes of this study. This data carries a high degree of confidentiality, and will not be further distributed or used for other purposes without their consent. In particular, the author would like to thank Marissa van Maaren and Sabine Siesling for their involvement in this project, sharing of their ideas, and their help with the submission of the application.
## Contents

1 **Introduction** ..................................................................................................................... 5
   1.1 Competing risks ............................................................................................................ 5
   1.2 Observational studies . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
   1.3 Adjustment methods ..................................................................................................... 7
   1.4 Project description ....................................................................................................... 9

2 **Methods** ....................................................................................................................... 10
   2.1 Competing risk models . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10
   2.2 Simulation study .......................................................................................................... 12
   2.3 Adjustment methods ................................................................................................... 14
   2.4 Validation .................................................................................................................... 17
   2.5 Statistical software and packages .............................................................................. 18

3 **Results** ......................................................................................................................... 19
   3.1 Sampling schemes ....................................................................................................... 19
   3.2 Performance on the cause-specific hazard model . . . . . . . . . . . . . . . . . . . . . . . . 20
   3.3 Performance on subdistribution hazards .................................................................... 26
   3.4 Computation time of conditional methods .................................................................. 31

4 **Application: breast cancer survival** ............................................................................. 33
   4.1 Background .................................................................................................................. 33
   4.2 Data characteristics ..................................................................................................... 34
   4.3 Covariate-adjusted cumulative incidence curves ....................................................... 38

5 **Discussion** .................................................................................................................... 43
   5.1 Performance and limitations of adjustment methods .................................................. 43
   5.2 Perspective on breast cancer survival . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44
   5.3 Future directions ......................................................................................................... 45
   5.4 Conclusion .................................................................................................................... 46

**Appendices** ...................................................................................................................... 52

A **R-code** .......................................................................................................................... 52
   A.1 Packages and seed ....................................................................................................... 52
   A.2 Supporting functions ................................................................................................... 53
   A.3 Data generation and biased sampling . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54
   A.4 Adjustment methods ................................................................................................... 57
   A.5 True curve computation .............................................................................................. 73
   A.6 Validation of adjustment methods .............................................................................. 75
   A.7 Adjustment of competing risks in breast cancer data .............................................. 91

B **Breast cancer model coefficients** .................................................................................. 99
   B.1 Covariate effects on treatment allocation (IPW) ....................................................... 99
   B.2 Covariate effects on competing outcomes ................................................................. 101
Abstract

Patient survival in biomedical studies is often subject to multiple clinical endpoints, all of which compete for the first and possibly only opportunity of occurrence. As a result, the occurrence of competing events may preclude the observation of a specific clinical outcome of interest. To gain further insight into specific outcomes in the presence of competing events, a special type of survival analysis is required, known as competing risks analysis. The presence of treatment effects in competing risk models can be visually examined by constructing a cumulative incidence curves. These curves illustrate the probability of first occurrence for each event over a series of time points, and thereby avoid the bias that is introduced by competing events in classic survival curves.

In randomized controlled trials, cumulative incidence curves are unaffected by confounding from patient-specific covariates, which is the result of strict random assignment of patients between treatment cohorts. However, observational studies may often introduce imbalance of covariates between treatment cohorts, as certain groups of patients may be overrepresented within a particular treatment strategy. Covariate imbalance between cohorts results in a biased comparison of cumulative incidence curves, since they reflect the average failure probability within each cohort. This may discourage researchers from using cumulative incidence curves to report findings in the presence of competing risks in the presence of covariate imbalance. Fortunately, strategies have already been well-documented to address covariate imbalance for survival analysis, which has led to covariate-adjusted survival functions. However, these methods have yet to be further expanded upon to provide covariate adjustment for the cumulative incidence curves used to report competing risk models.

In this study, we have developed and examined various adjustment methods to produce covariate-adjusted cumulative incidence curves in the presence of covariate imbalance between cohorts. A simulation study was carried out to compare the accuracy and precision of these methods, and the best-performing method was applied on real-world breast cancer survival data. Covariate adjustment in breast cancer survival data allowed us to shed light on the role of covariate imbalance between patients treated with mastectomy and those treated with breast-conserving therapy for each of the competing outcomes.
1 Introduction

1.1 Competing risks

To examine the long-term benefits of treatment strategies in biomedical studies, researchers have to consider a multitude of clinical endpoints (defined as events), all of which may affect the well-being of a patient in different ways. For instance, patients in breast cancer studies may experience tumour persistence, tumour recurrence, second primary tumours, metastasis, or death (Saad 2011). For this type of data, composite endpoints are often used, such as metastasis-free survival, in order to gain a more summarized evaluation of the treatment efficacy (Bonnetain et al. 2014). By doing so, researchers may miss out on the changes in the occurrence of individual events that underlie these composite measures. For example, a reduction in mortality may be masked by an increase in the occurrence of metastasis as those patients may live longer. Competing risks analysis is a special type of survival analysis which allows researchers to more closely examine treatment effects on the first occurrence of several competing outcomes.

Survival data is uniquely characterized by the presence of censoring, which holds an important role in the selection of a competing risk model. Traditionally, subjects are censored if the event of interest occurs outside of the follow-up time or study duration (Austin et al. 2016). In such cases, it is assumed that the event of interest would eventually occur for a censored subject. In the case of competing risks, patients can become ineligible to encounter the event of interest whenever a competing event has occurred first. This conflicts with the aforementioned definition of censoring, as patients may become unable to experience the event of interest after having already encountered a competing event. To avoid this dilemma, conventional statistical methods often assume non-informative censoring. By definition, non-informative censoring indicates that censoring does not hold any information on the prognosis of a subject (Cox and Oakes 1984), and that the mechanism of censoring and occurrence of the event are independent. As a result, whenever the event of interest is precluded by a competing event, then the subject would simply be treated as a dropout. Alternatively, researchers may also feign ignorance for the occurrence of competing events, and have the subject remain at risk until the end of the study duration without experiencing the event.

Each of these approaches has led to the development of two important subtypes of hazard-based regression models in the presence of competing risks: causes-specific hazard models and subdistribution hazard models. By classifying competing events as censored observations, cause-specific hazard models base their hazard function only on the subjects that have remained in an event-free state up until a certain point in time (Putter et al. 2007). To obtain a natural interpretation of treatment effects on the occurrence of a specific event, the cause-specific model can then be used to construct the cumulative incidence function. This function is defined as the instantaneous probability for event-free subjects to fail from a specific event. This definition carries a more applicable interpretation for clinical use, as it is a more straightforward indicator for the prognosis of individual patients. Furthermore, Austin et al. (2016) showed that the cumulative incidence function has more desirable properties in the presence of compet-
ing risks when compared to the complement of the crude Kaplan-Meier estimator. For instance, the all-cause failure probability, the probability to experience any event, can be computed as the sum of event-specific cumulative incidences. In contrast, this does not hold for the complement of the crude Kaplan-Meier estimator. Like others, Austin et al. (2016) illustrated that the naive Kaplan-Meier estimator has an upwards bias on the event-specific failure probability, as it does not take failure from competing events into account.

The subdistribution hazard model was first described by Fine and Gray (1999). In this type of model, all observations remain at risk until the event of interest has occurred, or when they would be truly censored. Consequently, observations remain at risk regardless whether the event of interest has been precluded by the occurrence of a competing event. Hence, competing events in the subdistribution hazard model are not able to violate assumption of independence between censoring and event occurrence. As a result, covariate effects modeled by subdistribution hazards have a direct association with the cumulative incidence of the event of interest, which is in contrast with the aforementioned cause-specific hazard model. Importantly, it has been illustrated that the subdistribution hazards approach results in proper partial likelihood estimates and valid inference (Putter et al., 2020). Nevertheless, the concept of subdistribution hazard has been widely regarded as “unnatural” in literature.

The differences between the two hazard models also affects the interpretation of each model. In competing risk analysis, one is often interested in the cumulative incidence of a certain event of interest by a given time unit. In other words, one would like to know the probability for the event of interest to occur before a given moment in time. The probability for the event of interest to occur is well-reflected by the subdistribution hazards approach, as its covariate effects can be directly related to cumulative incidence function for the event of interest. In contrast, coefficients from the cause-specific hazards model affect the relative degree at which the rate of occurrence for the event of interest increases over time in event-free individuals, which does not hold a direct relationship with the cumulative incidence function (Austin et al., 2016). As a result, the hazard rates of the subdistribution hazard models lend themselves well for the prediction of individual risk, whereas the focus on event-free observations in cause-specific hazard models may be more suitable to study epidemiological topics, such as the origins of certain diseases (Lau et al., 2009).

1.2 Observational studies

Many samples may be required to obtain a proper record of individuals that have experienced the event of interest in the presence of competing events. Observational studies can offer an abundant source of valuable information for competing risks by combining data from a multitude of available sources. This may also contribute to a better understanding of the application and efficacy of treatment strategies within the current-world population (Dreyer et al., 2010). Unfortunately, these advantages come at the cost of higher complexity due to the uncontrolled treatment allocation to subjects, which is prone to sampling bias (Nørgaard et al., 2017). For instance, it has been found that
treatment assignment in breast cancer is often influenced by patient characteristics, tumour properties, and personal preferences from either patients and/or clinicians (Morrow et al., 2009). The association between treatment and outcome may become spurious in the presence of such covariates if they affect both treatment allocation and outcome, which has been widely described as confounding. Consequently, covariate imbalance between treatment cohorts may imply the presence of bias in the functions used to report survival models, such as the cumulative incidence functions that are used to report competing risks (Jiang et al., 2011; Hu et al., 2020).

To overcome the issue of confounding when working with observational data, multivariate regression models have been developed to model any confounding effects as covariates, such as the multivariate Cox proportional hazards model (Scheike and Zhang, 2008). The disadvantage of this approach is that the addition of multiple covariates require stratification of survival curves to clarify their relationship both with treatment allocation and outcome, resulting in multiple covariate-specific curves. These curves may become too complex to properly visualize or interpret in response to multiple covariates and/or continuous covariates (Cole and Hernán, 2004). Alternatively, one can formulate a curve solely based on the marginal cohort estimates from the multivariate Cox proportional hazards model. The major downside of this approach is that survival curves represented by these marginal estimates will not reflect the survival of the patient population. Covariate-adjustment methods can be used address the imbalance of covariates between cohorts, thereby resulting in fair visual comparison between treatment cohorts in the patient population. Such adjustment methods have been established to correct the marginal Kaplan-Meier survival estimates used to report Cox regression models (Therneau et al., 2015; Jiang et al., 2011). However, the effects of covariate adjustment have yet to be evaluated for the cumulative incidence curves produced by competing risk models.

1.3 Adjustment methods

Several methods have been proposed by Therneau et al. (2015) to address covariate imbalance when comparing treatment cohorts using standard Kaplan-Meier survival curves. These methods have been subdivided into two main classes, marginal and conditional adjustments, based on the step at which the adjustment is performed during the modeling process. Marginal adjustment methods apply covariate-adjustment directly on the estimation of the survival curve, whereas conditional methods perform covariate-adjustment by first extracting covariate-specific predicted curves from the model and combining these in a weighted average of curves.

The marginal approach as described by Therneau et al. (2015) may involve subgroup selection to obtain a balanced subset between cohorts, or the use of weighted observations in order to alter their influence on the Kaplan-Meier estimate. The latter method is more commonly used, as it prevents the loss of valuable observations. To establish sampling weights, the most popular method is to calculate the probability for a set of covariates associated with a specific cohort, known as the propensity score, and take the inverse of this value. Based on this process, the method is commonly referred to
as inverse probability weighting, and can often be very easily implemented in statistical software by using logistic regression methods to estimate these probabilities (Cao et al., 2009; Austin and Stuart, 2015; Cole and Hernán, 2004). This method ensures a balanced comparison between treatment cohorts, in the presence of covariate imbalance, by counteracting on the differences in the covariate distribution between cohorts that originate from unbalanced sampling. This method has been validated by Cole et al. (2003) to produce adjusted survival curves to graphically examine the long-term survival benefits of highly active antiretroviral therapy against HIV. The main disadvantages of this method, as described by Austin and Stuart (2015), is that the accuracy of the adjustment is dependent on the validity of the logistic regression model. Hence, the method may perform poorly in the presence of a non-linear predictor function, or when certain covariates and/or interactions have not been properly included into the model. Therefore, the diagnostics for a logistic regression model should be verified carefully before use, which is often overlooked in practice. Another disadvantage mentioned by Austin and Stuart (2015) is that rare combinations of covariate values may result in extremely large weights. This can be overcome by using thresholds based on the distribution of the weights (Cole and Hernán, 2008).

As described by Therneau et al. (2015), the conditional approaches perform covariate-adjustment on intermediate predicted survival curves. First, a multivariate model is produced to account for confounders by including them as covariates. After that, the risk set of the entire sample or a desired subgroup is duplicated to each cohort, and a separate survival curve is predicted for each observation within the risk set of each cohort. Then the average survival probability is computed per cohort for each time point within a range of event times. This approach was used by Jiang et al. (2011) to construct adjusted survival curves for an oncology trial in which the survival benefits of pemetrexed administration was evaluated in patients exhibiting malignant pleural mesothelioma, a type of asbestos-related lung cancer. Researchers constructed a semi-parametric proportional hazards regression model, which was then used to predict the individual survival functions. The predictions were made on the set of covariates from each subject, which was replicated for both the treatment and control cohort. Lastly, the individual curves were averaged for each time unit for each cohort. To assess the performance of this method, Jiang et al. (2011) performed a simulation study with varying degrees of imbalance in covariates between cohorts, and showed that covariate-adjustment led to a higher statistical power between treatment cohorts when compared to the unadjusted survival method. Furthermore, they illustrated that the difference between the unadjusted and adjusted curves correlated with the degree of covariate imbalance between cohorts. A preliminary study by Hu et al. (2020) applied this method to adjust cumulative incidence curves derived from a Fine-Gray model. They did so by using competing risk analysis, which compared the occurrence of relapse or treatment-related mortality in leukemia patients treated with one of two blood donor groups. Within this data, it was found that there was an imbalance in leukemia stage of patients between blood donor groups. The efficacy of the adjustment was evaluated by simply comparing the covariate-adjusted cumulative incidence curves to the unadjusted curves,
which were shown to differ according to their prior expectations. Unfortunately, Hu et al. (2020) provided no further validation of the adjusted cumulative incidence curves despite having constructed simulated confidence bands to evaluate adjusted survival curves. As a result, the accuracy and variability of the adjustment methods have yet to be evaluated for the cumulative incidence curves obtained from the cause-specific hazard model. Furthermore, Therneau et al. (2015) also noted that the estimation of standard errors from an average of predicted curves is complex, and requires further investigation.

While these studies show that averaging over all observations may provide a highly accurate adjustment of survival curves in the presence of covariate imbalance, Therneau et al. (2015) noted that it can be computationally exhaustive in the presence of many observations, and proposed a modification to make this method more computationally efficient. They did so by only predicting the cumulative incidence curve for each unique combination of covariates. This approach reduces the total number of predictions, which decreases the overall computation time. This method was then demonstrated by Therneau et al. (2015) to correct for covariate imbalance to compare survival curves between free light-chain groups in patients with multiple myeloma. The accuracy and precision of the covariate-adjustment methods was evaluated by comparing these curves to the survival curves from a balanced reference population, which showed that the modification resulted in a similarly effective adjustment as the previously discussed methods. However, the downside of this modification is that does not work well in the presence of continuous covariates, which can take on many unique values. This may be partially solved by discretization of continuous covariates. However, this additional step may result in discretization error, thereby reducing the overall adjustment accuracy of this method.

1.4 Project description

The adjustment methods discussed so far have been well-documented for survival curves, but have yet to be validated for cumulative incidence curves used to report competing risk models. Furthermore, the functions required to produce covariate-adjusted cumulative incidence curves have yet to be developed for R statistical software. In this study, we produced methods in R to construct adjusted cumulative incidence curves for competing risk analysis. These methods have been developed for the cause-specific hazard and subdistribution hazard models. To integrate these methods into current analysis pipelines, they have been developed by using pre-existing functions from software packages available for competing risk analysis in R. Secondly, a simulation study was performed to evaluate the performance of the methods by comparing the adjusted curves to a hypothetical true curve. The best-performing method was further evaluated on real-world breast cancer survival data, which was done in collaboration with the Netherlands Comprehensive Cancer Organisation. This has allowed us to gain more insight on their applicability, and has also led to a better understanding of the competing risk aspects that underlie breast cancer survival.
2 Methods

2.1 Competing risk models

In this study, correction methods were developed to construct covariate-adjusted cumulative incidence functions under the cause-specific hazard and subdistribution hazard model. This study will assume the traditional case of right-censored data, with a continuous distribution of all failure times, and independence between the mechanism of censoring and occurrence of the events. Furthermore, proportional hazards of covariate effects were assumed in the presence of a Cox proportional hazards model. To understand how the conceptual differences between the two models, reflected by their hazard functions, one must first interpret the hazard function $\lambda(t)$, a function of time $t$, in the absence of competing risks (Putter et al., 2007; Austin et al., 2016).

$$\lambda_k(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

(1)

Using the definitions described by Austin et al. (2016), consider the hazard function to be the instantaneous rate of occurrence for a certain event of interest $k$, with $T$ indicating the time unit at which the event is observed. The hazard function considers the rate with which the event will occur between the current time $t$ and the very near future $T + \Delta t$, given that the event has to yet to occur, as indicated by $T \geq t$.

In the presence of covariates, summarized by a covariate matrix $X$, Putter et al. (2007) described that this hazard can be estimated using the Cox proportional hazards model $\lambda(t|X) = \lambda_0(t) \exp(X \beta)$, where $\lambda_0(t)$ represents the baseline hazard at $t$, and $\beta$ the regression coefficients. The hazard can be used to estimate the probability of survival, given by the survival function $S(t|X) = S_0(t)^{\lambda(t|X)}$, with $S_0(t) = \exp(-\int_0^t \lambda_0(t)dt)$ as the baseline survival function. As discussed by Austin et al. (2016), the association between hazard and failure probability is made by taking the complement of the survival function $F(t) = 1 - S(t)$. When there are a multitude of events (eg. clinical endpoints), the hazard function can represent the instantaneous rate of occurrence of any event, which summarizes all events into a single composite endpoint. We will get back to this in the application of competing risk on real-world data (Section 4).

Cause-specific hazard approach. In the presence of competing risks, one may be interested in the occurrence of one particular event. As a result, the all-cause hazard is narrowed down to the cause-specific hazard $\lambda_k$. To define the cause-specific hazard, consider multiple events that compete for first occurrence, with $D \in \{1, ..., K\}$ describing the event of first occurrence from a total of $K$ possible events, and $k$ representing the event of interest. This definition has been described by Putter et al. (2007), and can be expressed as

$$\lambda_k(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t, D = k | T \geq t)}{\Delta t}$$

(2)
Here, the cause-specific hazard function narrows the instantaneous rate of occurrence of any cause down to the event of interest \( k \). This hazard function only applies to individuals who have been event-free up to \( t \), and individuals who have encountered a competing event will be treated as censored observations. As discussed by Austin et al. (2016), this may conflict with the assumption of independent censoring, as competing events are also censored. Furthermore, as illustrated by Austin et al. (2016), the complement of the survival function has an upward bias in the presence of competing risk when based on a single cause, therefore making it unsuitable to estimate cause-specific failure probabilities. Hence, the cumulative incidence function is used instead, which represents the marginal probability to experience the event \( k \) at \( t \). To construct the cumulative function \( I_k(t) \), define \( \lambda_k(t) \) as the probability of experiencing the event of interest at \( t \), and \( S(t) \) as the probability to remain event-free till \( t \), where \( S(t) = \exp\left(-\sum_{k=1}^{K} \lambda_k(t)\right) \).

This results in the following expression

\[
I_k(t) = \Pr(T \leq t, D = k) = \int_0^t S(u^-)\lambda_k(u)du \tag{3}
\]

This function can be approximated by the Aalen-Johanson estimator. As illustrated by Putter et al. (2007), consider \( 0 < t_1 < t_2 < \ldots < t_N \) to be an ordered series of time points at which any event can occur. Define \( d_j \) to be the number of individuals that fail from any event at \( t_j \), \( d_{kj} \) as the number of individuals that failed from the event of interest at \( t_j \), and \( n_j \) the number of individuals that remain event-free at \( t_j \). These estimates can be used to estimate the cause-specific hazard \( \lambda_k(t_j) \) at \( t_j \), and the event-free survival \( S(t_j) \). The cumulative incidence function \( I_k(t_j) \) can then be estimated by

\[
\hat{I}_k(t_j) = \sum_{j : t_j \leq t} \hat{S}(t_{j-1})\hat{\lambda}_k(t_j), \quad \hat{S}(t_j) = \prod_{j : t_j \leq t} \left(1 - \frac{d_j}{n_j}\right), \quad \hat{\lambda}_k(t_j) = \frac{d_{kj}}{n_j} \tag{4}
\]

In the presence of covariates, as summarized by a covariate matrix \( X \), the Cox proportional regression can be similarly used as described before in the absence of competing risks. This was defined by Putter et al. (2007) as \( \lambda_k(t|X) = \lambda_{k,0}(t)\exp(\beta^T_k X) \).

**Subdistribution hazard approach.** The subdistribution hazard function was first described by Fine and Gray (1999). As discussed previously, the subdistribution hazard models assumes that all individuals will remain at risk until the event of interest has occurred or until they have been truly censored, even if the occurrence of the event of interest is precluded by that of a competing event. This relationship has been described as following

\[
\lambda_k^{sh} = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t, D = k|T \geq t \cup (T < t \cap K \neq k))}{\Delta t} \tag{5}
\]

Given that individuals have yet to encounter the event of interest, despite already having encountered a competing event, denoted by \( T < t \cap K \neq k \), the subdistribution hazard function is defined as the instantaneous risk of encountering the event of interest.
between now and the very near future [Austin et al., 2016]. That is to say, individuals will remain in the risk set until the event of interest eventually occurs. The relationship between subdistribution hazard \( \lambda_{k}^{sh} \) and cumulative incidence \( I_{k}(t) \) has been formulated by Fine & Gray as

\[
\lambda_{k}^{sh}(t) = -\frac{d\log(1 - I_{k}(t))}{dt} \quad (6)
\]

Similar to cause-specific hazard, the hazard function can be estimated in the presence of covariates using proportional hazards regression

\[
\lambda_{k}^{sh}(t|X) = \lambda_{k,0}^{sh}(t) \exp((\beta_{k}^{sh})^{T}X) \quad (7)
\]

### 2.2 Simulation study

**Generation of competing risk data.** A number of \( N \) subjects were generated, and divided over two cohorts. Consider \( Z_{i} \) a vector of length \( N \) that denotes the assigned cohort (control = 0, treatment = 1), which was determined by a series of independent and identically distributed Bernoulli trials, \( Z_{i} \sim \text{Bernoulli} \left( \frac{1}{2} \right) \). Each subject was simulated with their own set of covariates, one discrete covariate following a categorical distribution, \( X_{i1} \sim \text{Cat} \left( \{0, 1, 2\}, \{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\} \right) \), and one continuous covariate with a standard normal distribution, \( X_{i2} \sim \text{N}(0, 1) \). These two covariates were combined into a \((N \times 2)\) covariate design matrix \( X \).

To simulate competing risk data, let \( k \) denote a particular event from a total of \( K \) possible events, \( k \in \{1, ..., K\} \). For our simulation study, we assumed a total of \( K = 2 \) possible events. To model the event-specific effects of treatment and covariates on the hazard, the vector of cohort assignments \( Z \) was multiplied with an event-specific treatment effect coefficient \( \theta_{k} \), and the covariate design matrix \( X \) was multiplied with a \((1 \times 2)\) vector of event-specific covariate effect coefficients \( \beta_{k} \). We assumed a constant baseline hazard \( \lambda_{0}(t) = 1 \). This led to the following expression

\[
\lambda_{k} = \exp(Z\theta_{k} + X\beta_{k}) \quad (7)
\]

Latent event times were generated under the cause-specific hazard model as described by [Bender et al., 2005]. Individual event times for each event \( k \), \( T_{ik} \) were randomly generated by taking the logarithm of a random variable from a uniform distribution \( u_{ik} \sim \text{U}(0, 1) \), and dividing by the hazard of the event \( k \)

\[
T_{ik} = \frac{\log(u_{ik})}{\lambda_{k}} \quad (8)
\]

After simulating latent event times for each event \( k \), the observed event \( \delta_{i} \) was defined as the event that was associated with the smallest latent event time \( T_{i} \) from the set of \( T_{ik} \) generated latent event times

\[
T_{i} = \min(T_{i1}, ..., T_{iK}), \quad \delta_{i} = k : T_{ik} = \min(T_{i1}, ..., T_{iK}).
\]
Event times from the subdistribution hazards model were generated by a tailor-made method, based on the subdistribution hazard functions from Fine and Gray (1999). Consider \( K = 2 \) possible events, where \( k = 1 \) indicates the occurrence of the event of interest. To generate event times, an additional parameter \( p_1 = 1 - (1 - p_1^*) \) was added to express the probability for an individual to experience the event of interest

\[
T_{i1} = -\log \left( 1 - \frac{(1 - u_i^{1/\lambda_1})}{p_1} \right)
\]

\[
T_{i2} = -\log \left( 1 - \frac{u_i - p_1}{1 - p_1} \right)
\]

(9)

After generating the event times separately for both events, each subject experienced the event of interest if \( p_1 \) exceeded a randomly generated value from the uniform distribution \( u_i \sim U(0, 1) \), and experienced \( k = 2 \) otherwise. Observations were assigned to the first event if this value was smaller or equal to the baseline probability for encountering the event of interest

\[
T_i = \begin{cases} 
T_{i1} & \text{if } u_i \leq p_1 \\
T_{i2} & \text{otherwise}
\end{cases}
\]

\[\delta_i = k : T_i = T_{ik}\]

Lastly, censoring was added to both models by sampling censoring times \( C_i \) from a uniform distribution. Consider \( C_i \sim U(P_{.20}(T), P_{.95}(T)) \), where \( P_{.20}(T) \) and \( P_{.95}(T) \) respectively refer to the 20th and 95th percentile from the empirical distribution of the observed observation times \( T \). If the observation time \( T_i \) exceeded the simulated censoring time \( C_i \), then the event time \( T_i \) was set to the censoring time \( C_i \), and the observed status \( \delta_i \) was set to 0.

**Unbalanced selection.** To evaluate the efficacy of the correction methods on covariate imbalance between cohorts, two sampling procedures were used to generate an imbalanced subset from the simulated competing risks data. To determine the selection probability \( P_i \) for each observation \( i \) within the simulated population, the vector of individual covariate values \( X_{i,*} \) was multiplied with a vector of logistic regression coefficients \( b \), a vector of length two, where \( b_1 \) alters the selection probability \( P_i \) based on the discrete covariate value \( X_{i1} \), and where \( b_2 \) alters the selection probability \( P_i \) based on the continuous covariate value \( X_{i2} \). The value \( b \) determines the direction of the selection bias, where \( b = 0 \) represented strict random selection, \( b > 0 \) an increased selection probability for observations with higher values for \( X_{i1} \) and/or \( X_{i2} \), and \( b < 0 \) increased selection probability for observations with lower values for \( X_{i1} \) and/or \( X_{i2} \).

The first sampling scheme simulated opposing selection between the treatment and control cohort. Here, the covariate distribution of the entire sample would reflect that of the original population. To express this type of selection, the selection probabilities in the control cohort were calculated as the opposite of the selection probabilities determined under the treatment cohort.
\[ P_i(Z_i, X_i) = \begin{cases} \exp(X_i^\top b) / (1 + \exp(X_i^\top b)) & \text{if } Z_i = 1 \\ 1 - \exp(X_i^\top b) / (1 + \exp(X_i^\top b)) & \text{if } Z_i = 0 \end{cases} \] (10)

For the second sampling scheme, we wanted to retain strict random sampling of individuals from the simulated control cohort \( Z = 0 \). This was done by using a constant selection probability for all observations in the control cohort \( P_i = 0.5 \). As a result, the distribution of covariates from the simulated control cohort would reflect that of the entire target population. Similar to the previous scheme, we induced selection bias in the sampling process of the treatment cohort \( Z = 1 \), where the selection probability was affected by with the individual covariate values \( X_i \). The selection probability \( P_i \) for each individual was then calculated as

\[ P_i(Z_i, X_i) = \begin{cases} \exp(X_i^\top b) / (1 + \exp(X_i^\top b)) & \text{if } Z_i = 1 \\ 0.5 & \text{if } Z_i = 0 \end{cases} \] (11)

Based on the selection probabilities obtained from either sampling procedure, a subset of \( n \) samples were obtained per cohort by using an iterative sampling method. First, observations were ordered by their selection probability \( P_i \), starting from the observation with the highest value for \( P_i \) to the lowest value. For either sampling procedure, samples were obtained by performing a biased coin flip on the sampling probability of each observation, relative to the combined sampling probabilities of the remaining observations. If an observation was selected, it was removed from the pool of observations, and the process was repeated until a subset of \( n \) observations was achieved. The selection procedure was performed independently for each of the two simulated cohorts, in order to ensure an equal sample size for both cohorts.

2.3 Adjustment methods

Marginal approaches. The marginal adjustment methods use observations weights to perform correction on the estimation of the cumulative incidence function. Recall the aforementioned cumulative incidence estimator \( I_k(t) \), where the \( d_{kj} \) represents the number of subjects that experienced event \( k \) at \( t_j \), \( d_j \) the total number of individuals that experienced any event, and \( n_j \) the number of subjects that remain at risk at \( t_j \) (Therneau et al., 2021). We can substitute the indicators functions used to define these parameters by a weighted indicator function, in which we sum the weights from the collection of subjects in \( d_j \) and \( n_j \) at \( t_j \). This would result in the weighted survival function \( \hat{S}^W(t) \), and weighted cause-specific hazard functions \( \hat{\lambda}^W_k(t) \) that can then be used to estimate the covariate-adjusted cumulative incidence function \( \hat{I}^W_k(t) \)

\[
\hat{I}^W_k(t|X) = \sum_{j:t_j \leq t} \hat{S}^W(t_{j-1}) \hat{\lambda}^W_k(t_j),
\]

\[
\hat{S}^W(t_j) = \prod_{j:t_j \leq t} \left( 1 - \frac{\sum_{i \in d_j} w_i}{\sum_{i \in n_j} w_i} \right), \quad \hat{\lambda}^W_k(t_j) = \frac{\sum_{i \in d_{kj}} w_{ki}}{\sum_{i \in n_j} w_{i}}
\] (12)
Two methods were developed to construct the weights used for the marginal adjustment of cumulative incidence curves. The first method being the inverse probability weights, which were estimated as described by [Austin and Stuart (2015)]. In this method, the weights \( w_i \) are based on the inverse of the probability for an individual with a given set of covariates \( x_i \) to have been assigned their own cohort \( Z_i \). To formulate these weights, consider \( \Pr(Z = z_i, X = x_i) \) the joint probability of a set of covariates \( x \) to occur in combination with a cohort \( z \), and \( \Pr(Z = z) \) the marginal probability for an individual to be assigned the \( z \)th cohort. We can apply Bayes rule to obtain the conditional probability for an individual to be assigned their own cohort, given their set of covariates

\[
\Pr(Z = z_i | X = x_i) = \frac{\Pr(Z = z_i, X = x_i)}{\Pr(Z = z_i)}
\]

To obtain the weights, we can simply take the inverse of the conditional probability for an individual to be assigned their own strata.

\[
w_i = \frac{1}{\Pr(Z = z_i | X = x_i)} \tag{13}
\]

This formulation indirectly assumes that the covariate distribution of the entire sample reflects that of the population, and that unbalanced assignment can be explained by the underlying correlation between covariates and treatment allocation. In the case where subjects from one cohort were assigned to that cohort by a strict random draw from the target population, these weights may behave poorly. In this scenario, [Austin and Stuart (2015)] used an altered formulation of the inverse probability weights, in which one particular cohort can be used as a reference, and where the weights are dependent on the ratio between the reference cohort and an individual cohort. Consider \( z = 0 \) to be the reference cohort with balanced covariate distributions. As a result, the weights for observations in the reference cohort become equal to 1, which prevents the introduction of bias in that cohort. Then the weights become relative to the conditional probabilities of this cohort

\[
w_i = \frac{\Pr(Z = 0 | X = x_i)}{\Pr(Z = z_i | X = x_i)}
\]

These conditional probabilities can be easily obtained by using standard logistic regression methods. As a result, inverse probability weighting can be considered a rather flexible correction method, as one can freely decide which covariates and interactions are required to ensure an unbiased comparison between treatment. However, it also requires researchers to properly identify which covariates to include during model selection, and to identify possible interactions between covariates on the sampling scheme. As [Austin and Stuart (2015)] points out, this can be particularly difficult in the presence of many covariates. Furthermore [Austin and Stuart (2015)] indicated that rare combinations of covariates may result in extremely large weights. To circumvent this, they proposed to apply a threshold based on the 1th and 99th percentile of the distribution of weights.
By extending on the concepts discussed for inverse probability weighting, one can also consider incorporating external information on the distribution of covariates from a hypothetical target population into the weights.

\[ w_i = \frac{\Pr(Z = z_i, X = x_i|i \in target)}{\Pr(Z = z_i, X = x_i|i \in sample)} \quad (14) \]

In this method, weights are determined based on how far the joint probability of cohort assignment and covariate distribution within the sample deviates from that of a hypothetical target distribution. In theory, this approach should allow researchers to standardize the curves based on demographics that have been obtained from external studies. These probabilities can be obtained by calculating the joint probability for each combination of covariates both within the sample and within individual patient data from external cohorts. This may require discretization of continuous covariates such that they may be represented by both the sample and external data.

**Conditional approaches.** While the marginal approach performs adjustment during the estimation of the cumulative incidence itself, the conditional methods perform the adjustment on the predictions from the intermediate model by including \(X\) as a covariate. This process has been described by Jiang et al. (2011) for survival curves, and Hu et al. (2020) for cumulative incidence curves from subdistribution hazard models. To construct the weighted average, cumulative incidence functions are constructed for all subjects based on the cause-specific hazard or subdistribution hazard model. Thereafter, each subject \(i\) with their covariate values \(X_i\) is duplicated to each cohort, after which predicted cumulative incidence curves \(I_{ki}(t|Z = z_i)\) are extracted for each observation. Lastly, an average is taken over all individuals per cohort to obtain the event-specific cumulative incidence curve \(I_k(t|Z)\) for each cohort.

\[ I_k(t|Z) = \mathbb{E}_X \left[ \tilde{I}_k(t|Z, X = x_i) \right] = \frac{1}{N} \sum_{i=1}^{N} \tilde{I}_k(t|Z, X = x_i) \quad (15) \]

As discussed with the marginal methods, some cohorts may already represent the target distribution of covariates, and thus should be used as a reference to perform the weighted average. This can be simply done by only duplicating the risk set of the reference cohort (here \(Z = 0\)) to the other cohorts (here \(Z = 1\)). Consider, the control cohort from the simulation to represent the reference cohort, then the weighted average can be rewritten as \(E_{\{X|Z=0\}} \left[ \tilde{I}_k(t|Z, X = x) \right] \), in which \(Z = 0\) restrains the predicted curves to only subjects that were originally from the control cohort, and have been duplicated to each cohort. After that, the procedure continues as described before, with an average for each cause-specific cumulative incidences per cohort at \(t\).

While the weighted average may prove to be very effective to ensure an unbiased comparison between cohorts, Therneau et al. (2015) argued that it can become a quite computationally exhaustive method if curves are to be predicted over a large number of subjects and time points. The computation time can be reduced by performing...
a weighted average over all unique combinations of covariates, instead. To optimize computation efficiency, one can consider to discretize continuous covariates. Otherwise, every continuous value might be seen as a distinct value, which would then still result in a distinct cumulative incidence function for every single observation. Consider $M$ the total number of distinct combinations of (binned) covariates. A cause-specific hazard or subdistribution hazard model is constructed from the entire sample. Given that an individual was assigned the $z$th cohort in combination with a unique set of covariate values $x$, a cumulative incidence curve for the event $k$ is extracted. We can obtain a weighted cumulative incidence with a desired covariate distribution by multiplying each distinct cumulative incidence $I_k(t|Z)$ at $t$ with the relative frequency of that set of covariates to occur within the entire sample, and sum over all unique combinations of covariate values.

$$I_k(t|Z) = \sum_{x=1}^{M} p_x I_k(t|Z, X = x), \quad p_x = \Pr(X = x) \quad (16)$$

If one would like to account for a reference cohort, such as the control cohort ($Z = 0$) then the conditional probability $p_x = \Pr(X = x|Z = 0)$ can be used instead.

### 2.4 Validation

The accuracy and precision of the correction methods were evaluated for the cause-specific hazard and subdistribution hazard model by examining their ability to approximate the true cumulative incidence curve. This was done by calculating the true cumulative incidence for each cohort $I_k(t|Z)$ at $t$, given the distribution of covariates $X$. To gain a measure of accuracy, the unadjusted and covariate-adjusted cumulative incidence curves were subtracted from the true curve. The adjusted cumulative incidence curve $\tilde{I}_k^{adj}(t|Z)$ was constructed by generating event-time data using the same set of coefficients as the true curve, which was followed by biased sampling and covariate adjustment. To obtain cohort-specific estimates of the true curve, the true curve was constructed for every unique combination of covariates, and averaged over discrete and continuous covariates. Due to the non-symmetrical properties of the continuous covariate $X_2$, a moment-generating function was used to average over a function with random variable $x_2 \sim N(0, 1)$, where $f_{x_2}$ denotes the point density function of a standard normal distribution. The average over the continuous covariate was determined in combination with each of the categories from the discrete covariate $X_1$, and an average was taken over all categories. Here, $f_{x_1}$ represents the probability mass function of a categorical random variable. As the integral was not of a closed-form expression, numerical integration was used to solve this integral.

$$I_k(t|Z) = \mathbb{E}_{x_1} [\mathbb{E}_{x_2} [I_k(t, x_1, x_2|Z)]]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_k(t, x_1, x_2|Z)f_{x_1}(x_1)f_{x_2}(x_2) \, dx_1 dx_2 \quad (17)$$
These curves were separately constructed based on the cumulative incidence function of the cause-specific hazard and subdistribution hazard model. To evaluate the performance of the adjustment method, the simulation was repeated $B = 1000$ times, after which the average deviation of the adjusted cumulative incidence $\hat{I}_{k|Z}^{adj}(t)$ from the true cumulative incidence $I_{k|Z}(t)$ was calculated, which we can refer to as the bias of our adjustment method. The deviation from the true curve was determined over a range of time points within the distribution of observation times $T$, and examined more closely at the central point of simulated event times, where the deviation of the unadjusted curve from the true curve is well observable.

$$E \left[ \hat{I}_{k|Z}^{adj}(t) - I_{k|Z}(t) \right] = \frac{1}{B} \sum_{b=1}^{B} \left( \hat{I}_{k|Z}^{adj}(t) - I_{k|Z}(t) \right) \quad (18)$$

The degree and direction of the bias can serve as an overall measure of adjustment accuracy. In addition to that, the reproducibility of the results (i.e. the degree of variation between results for similar sample configurations) can also be incorporated in the evaluation of the overall performance by calculating the root-mean-squared error (RMSE). This value was obtained by squaring the differences before averaging, and taking the squared root over the average of squared errors.

$$\text{RMSE} \left[ \hat{I}_{k|Z}^{adj}(t) \right] = \sqrt{\frac{1}{B} \sum_{b=1}^{B} \left( \hat{I}_{k|Z}^{adj}(t) - I_{k|Z}(t) \right)^2} \quad (19)$$

Lastly, graphical comparisons were made between the true cumulative incidence curve, and the cumulative incidence curves of the simulated population, unadjusted sample, and the adjusted sample. This was done by subtracting the true cumulative incidence curve from the latter curves, thereby providing a visual impression of their performance at a single point in time or over a range of time points.

2.5 Statistical software and packages

R statistical software was used to develop the covariate-adjustment methods for competing risks, and to apply these methods to analyze breast cancer survival data from van Maaren et al. (2019) (R Core Team, 2021). The adjustment methods were developed using functions from the survival and riskRegression packages (Therneau, 2020; Gerds and Ozenne, 2020). Appendix A contains the R-code used to produce the adjustment methods, perform the evaluation study, and to construct the figures and tables used in the results section. This also includes a more concise description of the packages used during this study.
3 Results

3.1 Sampling schemes

To evaluate whether the biased sampling procedures were successful, the distribution of covariates within each cohort was examined. To ensure a fair comparison between sampling schemes, the coefficients that influenced the degree of biased sampling on the treatment cohort (treatment sampling coefficients $b_1 = 2, b_2 = 2$) were made twice as large as the coefficients to perform biased sampling on both cohorts in opposing direction (opposite sampling coefficients $b_1 = 1, b_2 = 1$). Figure 2 shows the effects from each sampling procedure on the resulting covariate distribution of the sample ($n = 1000$), which was compared to the distribution of the covariates from the original population ($N = 10000$) on which the biased sampling procedures were applied.

![Diagram showing the effects of each sampling procedure on the resulting covariate distribution of the sample.](image)

As discussed in section 2.2, using positive regression coefficient for the discrete covariate in combination with the opposing biased sampling would result in a opposite shift in cohort-specific group proportions of the discrete covariate (controlled by $b_1$). Furthermore, the covariate distributions from the treatment and control cohorts are shown to equally shift in opposite directions. A similar effect is seen by using a positive regression coefficient for the continuous covariate (controlled by $b_2$).

Performing biased sampling on solely the treatment cohort resulted in a distribution of covariates similar to the treatment cohort in the opposing biased sampling scheme. Furthermore, these results illustrate that by using a constant selection probability for the control cohort, the covariate distribution of the control cohort remains in line with original population. These results were in line with our expectations from this particular sampling procedure.

Therefore, these results demonstrate that the biased sampling procedures were successful in creating a desired covariate imbalance between cohorts, either on a particular cohort or counteractive on both cohorts.
3.2 Performance on the cause-specific hazard model

To evaluate the performance of the adjustment methods, a total of 1000 replications of population simulation ($N = 10000$), biased sampling, and covariate-adjustment were performed. As described in section 2.2, two biased sampling schemes were considered, one with opposing correlation between treatment allocation and covariate distribution ($n = 1000$, $b_1 = 1$, $b_2 = 1$), and an alternative scheme were the sampling probability solely correlated with the covariate distribution in the treatment cohort ($n = 1000$, $b_1 = 2$, $b_2 = 2$). Covariate adjustment was performed on each biased sampling scheme, using the marginal or conditional approaches discussed in section 2.3. As described previously, the marginal approaches comprised of inverse probability weighting (weights inverse prob) and weighting based on external reference data (weights external ref). For the first conditional approach, predicted curves were obtained for each observation from the cause-specific Cox proportional hazard regression model of the riskRegression package, after which the curves were averaged per cohort (average CI riskReg pred). The cause-specific Cox regression model from the riskRegression was used as it has been described to be more computationally efficient compared to the cause-specific Cox regression model from the survival package (Gerds and Ozenne, 2020). Additionally, conditional stratification methods were also examined. These were described as alternative methods by Therneau et al. (2015) to increase the computation efficiency of the models from the survival package. For these methods, predictions were performed on stratified sets of covariates rather than individual observations. The predicted cumulative incidence curves would then be either obtained from the Cox proportional hazard regression model (average CI coxph strata), or by using a cumulative incidence estimator (average CI survfit strata). To evaluate the adjustment accuracy and reproducibility for each of the adjustment methods, the difference from the hypothetical true curve was computed. A visual representation of the deviation from the true curve is shown in Figures 3 and 4. Furthermore, to provide a numerical comparison of the adjustment accuracy and reproducibility, bias and RMSE values were determined at $t = 0.3$, which represents a center point within the range of event times with a noticeable deviation from the true curve. These results have been illustrated by Figures 5 and 6, and Tables 1 and 2.

The effects of covariate-adjustment in the opposing sampling scheme are illustrated by Figure 3 over the entire range of time points [0, 0.06]. Following biased sampling, there is a clear deviation from the true curve in the unadjusted treatment and control cohorts, which is shown to drive the cumulative incidence curves of both cohorts in opposite directions. At first glance, these results show that performing covariate adjustment over the entire curve, with any of the adjustment methods, is well able resolve the bias introduced by the opposing sampling scheme across all events. Furthermore, the results of the adjustment is shown to be similarly reproducible between methods. The degree of variation that is present in the adjusted curves is shown to be comparable to the level of variation found within the initial sample, which can therefore likely be attributed to sampling error. Closer examination of the adjustment accuracy (bias values) indicates that the methods that require a degree of stratification retain a slight deviation from
the true curve. These methods include the external reference weights, the stratified Cox regression model, and stratified cumulative incidence estimator. In comparison, there is no apparent bias left following inverse probability weighting or by averaging over predicted individual curves using Cox regression from the riskRegression package. This indicates that a loss in adjustment accuracy in the stratification methods is likely the result of the required discretization of the continuous covariate. However, closer examination at $t = 0.3$, illustrated by Figure 5 and Table 1, emphasizes that the differences in bias and RMSE between the methods remains close to negligible. However, it should be considered that the adjustment accuracy may further decline upon stratification in the presence of multiple continuous covariates. Nonetheless, these results show that all methods are able to resolve bias introduced by the opposing selection, and provide highly reproducible results.

As discussed in the section 2.3, covariate-adjustment may require the incorporation of a reference group in the adjustment procedure. Therefore, an alternative sampling scheme was introduced where biased sampling solely affected the treatment cohort, in order to evaluate the performance of adjustment methods on a peculiar sampling scheme. The results for this sampling scheme are shown in Figure 4 and also further demonstrate that the constant selection probability for the control cohort successfully prevented its deviation from the hypothetical true curve. In contrast to the previous sampling scheme, the inverse probability weighted curve of the treatment cohort noticeably deviates from the hypothetical true curve. This is further emphasized by a relatively large value for the bias at $t = 0.3$ in the inverse probability weighted treatment cohort compared to other methods, as illustrated by Figure 6 and Table 2. Despite the low adjustment accuracy of the treatment cohort, the curve from the control cohort does not show noticeable deviation from the true curve. These results indicate that the multiplication with the propensity of the control cohort was able to prevent the introduction of bias in the control cohort. However, the resulting weights were not capable enough to resolve the bias present in the treatment cohort. Comparison of the variation in adjustment accuracy between methods indicates a noticeable difference in reproducibility of the results for the adjustment of the treatment cohort. Interestingly, the methods that are based on the cumulative incidence estimator are shown to be less reproducible (i.e. can deviate more heavily from the true curve) on the treatment cohort. These methods include the inverse probability weights, the external reference weights, and the average of stratified curves from the cumulative incidence estimator. Each of these methods show a relatively higher deviation in the treatment cohort compared to the other methods. These results are further emphasized by the bias and RMSE values at $t = 0.3$, illustrated by Figure 6 and Table 2, which show that their RMSE values in the treatment cohort are almost twice as large as that RMSE values from their control cohort. Interestingly, each of these methods are based on the cumulative incidence estimator rather than the Cox regression model. Thereby indicating that the intermediate step of Cox regression modeling may result in a more reproducible outcome of the adjustment. This is in contrast with the results from the previous sampling scheme. Therefore, these results indicate that the modeling approach is able to adapt better to this particular sampling
scheme when compared to the marginal approach.

To summarize, these results show that all methods are able to resolve the bias introduced by opposing covariate selection during sampling, and that any decrease in adjustment accuracy may be the result from stratification in the presence of continuous covariates. Out of all methods, the conditional adjustment methods, involving Cox regression modeling, are able to consistently demonstrate a high adjustment accuracy against each biased sampling scheme. These results therefore indicate that averaging of the predicted curves via Cox modeling is both an accurate and versatile adjustment method to provide a fair comparison between cohorts, which can be easily tailored to adjust for specific cases of covariate imbalance. Interestingly, the adjustment accuracy of methods that involve the cumulative incidence estimator were shown to be less reproducible compared to the methods that involve Cox modeling. This may either correspond to poor handling of the weights by the cumulative incidence estimator, or may be the result from a lack of pooling of information to infer over missing combinations of covariates. Consequently, the conditional modeling methods have demonstrated to be more versatile in general compared to the marginal methods, as they were easily adaptable to perform accurate and reproducible adjustment in the presence of a reference group.
Figure 3: Deviation from the true cumulative incidence (CI) over a range of simulated time points [0, 0.6] from the cause-specific hazard model over 1000 replications, shown per cohort for each event. Covariates influenced the sampling probability for each cohort in opposite directions. The ribbon indicates the standard deviation per time point.

Figure 4: Deviation from the true cumulative incidence (CI) over a range of simulated time points [0, 0.6] from the cause-specific hazard model over 1000 replications, shown per cohort for each event. Covariates solely shifted the covariate distribution from the treatment cohort. The ribbon indicates the standard deviation per time point.
Figure 5: Deviation from the true cumulative incidence (CI) of the cause-specific hazard model over 1000 replications at $t = 0.3$, illustrated per cohort and per event. Covariates influenced the sampling probability for each cohort in opposite directions. Each violin denotes the mean (●) and 95% quantile of the bias (|).

Table 1: Results from the cause-specific hazards simulation study following covariate adjustment on the opposing sampling bias between cohorts. Bias and root-mean-square error (RMSE) at $t = 0.3$ are shown for all events following 1000 replications of population generation, unbiased sampling, and covariate-adjustment.

<table>
<thead>
<tr>
<th></th>
<th>Event 1</th>
<th>Event 2</th>
<th>Event free</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias  RMSE</td>
<td>Bias  RMSE</td>
<td>Bias  RMSE</td>
</tr>
<tr>
<td>Target population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>-0.00017  0.08755</td>
<td>0.00010  0.08081</td>
<td>0.00006  0.07785</td>
</tr>
<tr>
<td>Treatment</td>
<td>-0.00021  0.08705</td>
<td>0.00015  0.08319</td>
<td>0.00006  0.07944</td>
</tr>
<tr>
<td>Sample opposing bias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>-0.11936  0.34727</td>
<td>0.05192  0.23251</td>
<td>0.06744  0.26295</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.11970  0.34759</td>
<td>-0.05750  0.24374</td>
<td>-0.06220  0.25158</td>
</tr>
<tr>
<td>Wts inverse prob</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>-0.00067  0.12746</td>
<td>0.00102  0.11817</td>
<td>-0.00036  0.10854</td>
</tr>
<tr>
<td>Treatment</td>
<td>-0.00017  0.13082</td>
<td>-0.00011  0.14033</td>
<td>0.00029  0.13241</td>
</tr>
<tr>
<td>Wts external ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>-0.00530  0.12769</td>
<td>0.00447  0.12056</td>
<td>0.00083  0.10549</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.00486  0.13003</td>
<td>-0.00600  0.14323</td>
<td>0.00114  0.13161</td>
</tr>
<tr>
<td>Avg CI riskReg obs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>-0.00018  0.11668</td>
<td>0.00065  0.11188</td>
<td>-0.00046  0.09910</td>
</tr>
<tr>
<td>Treatment</td>
<td>-0.00038  0.11847</td>
<td>0.00042  0.12630</td>
<td>-0.00005  0.10955</td>
</tr>
<tr>
<td>Avg CI coxph strata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>-0.00492  0.12101</td>
<td>0.00467  0.11630</td>
<td>0.00025  0.08910</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.00431  0.12184</td>
<td>-0.00534  0.12992</td>
<td>0.00103  0.10948</td>
</tr>
<tr>
<td>Avg CI survfit strata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>-0.00502  0.12819</td>
<td>0.00445  0.12080</td>
<td>0.00057  0.10671</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.00445  0.13111</td>
<td>-0.00560  0.14316</td>
<td>0.00115  0.13191</td>
</tr>
</tbody>
</table>
Figure 6: Deviation from the true cumulative incidence (CI) of the cause-specific hazard model over 1000 replications at $t = 0.3$, illustrated per cohort and per event. Covariates solely influenced the sampling probability of the sample drawn for the treatment cohort. Each violin denotes the mean (●) and 95% quantile of the bias (|).

Table 2: Results from the cause-specific hazards simulation study following covariate adjustment on the sampling bias solely present in the treatment cohort. Bias and root-mean-square error (RMSE) at $t = 0.3$ are shown for all events following 1000 replications of population generation, unbiased sampling, and covariate-adjustment.

<table>
<thead>
<tr>
<th></th>
<th>Event 1</th>
<th>Event 2</th>
<th>Event free</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
</tr>
<tr>
<td>Target population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>−0.00009</td>
<td>0.00721</td>
<td>−0.00009</td>
</tr>
<tr>
<td>Treatment</td>
<td>−0.00004</td>
<td>0.00745</td>
<td>0.00001</td>
</tr>
<tr>
<td>Sample treatment bias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>−0.00061</td>
<td>0.01621</td>
<td>−0.00011</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.17309</td>
<td>0.17382</td>
<td>−0.08207</td>
</tr>
<tr>
<td>Wts inverse prob</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>−0.00061</td>
<td>0.01621</td>
<td>−0.00011</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.03029</td>
<td>0.03758</td>
<td>−0.02546</td>
</tr>
<tr>
<td>Wts external ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>−0.00013</td>
<td>0.01398</td>
<td>−0.00018</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.01322</td>
<td>0.02952</td>
<td>−0.01366</td>
</tr>
<tr>
<td>Avg CI riskReg obs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>−0.00056</td>
<td>0.01500</td>
<td>−0.00027</td>
</tr>
<tr>
<td>Treatment</td>
<td>−0.00047</td>
<td>0.01540</td>
<td>−0.00007</td>
</tr>
<tr>
<td>Avg CI coxph strata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>−0.00001</td>
<td>0.01500</td>
<td>−0.00072</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.00708</td>
<td>0.01715</td>
<td>−0.01026</td>
</tr>
<tr>
<td>Avg CI survfit strata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>−0.00064</td>
<td>0.01601</td>
<td>−0.00018</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.01036</td>
<td>0.02701</td>
<td>−0.01654</td>
</tr>
</tbody>
</table>
3.3 Performance on subdistribution hazards

Similar to the evaluation approach of the cause-specific hazard model, the adjustment accuracy was determined for the subdistribution hazard model by performing 1000 replications of population simulation ($N = 10000$), biased sampling, and covariate-adjustment. Since the simulation of subdistribution hazards data had a particular focus towards the occurrence of the first event, the effects of covariate adjustment were examined for this particular event. The accuracy and precision of covariate-adjustment were evaluated on two sampling scheme described in section 2.2, one with opposing selection probabilities between cohorts in response to covariate imbalance ($n = 1000$, $b_1 = 1$, $b_2 = 1$), and another sampling scheme where covariate imbalance solely affected sampling in the treatment cohort ($n = 1000$, $b_1 = 2$, $b_2 = 2$). The results from the cause-specific hazard model showed that the specific combination of stratification and the cumulative incidence generally resulted in a poor adjustment accuracy and reproducibility. Therefore, external reference weighting (weights external ref) and averaging of stratified predicted curves from the cumulative incidence estimator (average CI survfit strata) were excluded from the subdistribution hazard evaluation study. Furthermore, during preliminary application it was found that performing cumulative incidence predictions for the subdistribution hazard model was not as computationally intensive compared to the cause-specific hazard model. Therefore, predictions for individual observations and stratified sets of covariates were both performed using the subdistribution hazards Cox regression model from the `survival` package.

The deviation from the hypothetical true curve in response to covariate adjustment is illustrated by Figures 7 and 8, respectively for the opposing bias and treatment bias sampling schemes. Similar to the approach for the cause-specific hazards evaluation, bias and RMSE values were computed for the subdistribution hazard model at a single point in time ($t = 0.7$), which represented a center point within the range of event times with noticeable deviation from the true curve in response to biased sampling. The results are illustrated by Figures 9 and 10, and Tables 3 and 4. Overall, the adjustment accuracy and reproducibility of each method is in line with the results from the cause-specific hazard model. Each adjustment was able to resolve the biased introduced by the opposing bias scheme, illustrated by Figure 7, and the observed degree of variation can attributed to sampling error already present in the unadjusted sample (see RMSE). This is further emphasized by the negligible differences in bias and RMSE at $t = 0.7$ between adjustment methods, as shown by Figure 9 and Table 3.

As discussed previously, the covariate adjustment methods were also evaluated on their ability to incorporate a reference group in their adjustment procedure. This was evaluated by solely introducing bias in the treatment cohort, as shown in Figure 8. Similar to the results for the cause-specific hazard model, the entire cumulative incidence curve for the treatment cohort is shown to deviate from the hypothetical true curve. In comparison, the curve for the sampled control cohort is observed to be unaffected. Furthermore, the results for inverse probability weighting again demonstrated that this method was not able resolve the bias present in the the treatment cohort. In contrast, both individual and stratified conditional approaches were able to resolve this specific
type of bias. Closer examination at $t = 0.7$, illustrated by [10] and [4] confirms these observations. Their output shows a negligible difference in bias for the control cohort between adjustment methods, but a noticeable difference in bias for the inverse probability weighted treatment cohort relative to the conditional methods. As mentioned previously, these observations are in line with the effects observed for the cause-specific hazard model.

To summarize, these results show that inverse probability weighting can be considered a simple and effective method to perform covariate adjustment when the distribution of the entire sample reflects that of a hypothetical target distribution. This was reflected by high adjustment accuracy and reproducibility of each method in response to the opposing bias sampling scheme. However, in the presence of a reference group, the inverse probability weights demonstrated poor adjustment accuracy, which indicates that the incorporation of the reference group has failed, and may require a more complex strategy than simply multiplying the weights with the propensity score of the reference group. Conditional averaging of predicted cumulative incidences has demonstrated the highest and most reproducible adjustment accuracy in each sampling scheme. Furthermore, these results indicate that the ability of the conditional approach to modify the risk set is a simple and effective tool to address specific types of covariate imbalance between cohorts. In line with the results from the cause-specific hazard model, the conditional method has also demonstrated to be the most accurate and versatile method for the subdistribution hazard model to provide covariate adjustment in response to multiple types of biased sampling procedures.
Figure 7: Deviation from the true cumulative incidence (CI) over a range of simulated time points [0, 1.4] from the subdistribution hazard model over 1000 replications, shown per cohort for the 1st event. Covariates influenced the sampling probability for each cohort in opposite directions. The ribbon indicates the standard deviation per time point.

Figure 8: Deviation from the true cumulative incidence (CI) over a range of simulated time points [0, 1.4] from the subdistribution hazard model over 1000 replications, shown per cohort for the 1st event. Covariates solely influenced the sampling probability of the sample drawn for the treatment cohort. The ribbon indicates the standard deviation per time point.
Figure 9: Deviation from the true cumulative incidence (CI) of the subdistribution hazard model over 1000 replications at $t = 0.7$, illustrated per cohort for the 1st event. Covariates influenced the sampling probability for each cohort in opposite directions. Each violin denotes the mean (●) and 95% quantile of the bias (|).

Table 3: Results from the subdistribution hazards simulation study following covariate adjustment on the opposing sampling bias between cohorts. Bias and root-mean-square error (RMSE) at $t = 0.7$ are shown for event 1 following 1000 replications of population generation, unbiased sampling, and covariate-adjustment.

<table>
<thead>
<tr>
<th></th>
<th>Event 1</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>RMSE</td>
<td></td>
</tr>
<tr>
<td>Target population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>$-0.00007$</td>
<td>0.00651</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>$0.00045$</td>
<td>0.00723</td>
<td></td>
</tr>
<tr>
<td>Sample opposing bias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>$-0.09851$</td>
<td>0.09977</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>$0.12072$</td>
<td>0.12146</td>
<td></td>
</tr>
<tr>
<td>Weights inverse prob</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>$0.00204$</td>
<td>0.01215</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>$0.00299$</td>
<td>0.01906</td>
<td></td>
</tr>
<tr>
<td>Avg CI coxph obs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>$0.00224$</td>
<td>0.01011</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>$0.00171$</td>
<td>0.01178</td>
<td></td>
</tr>
<tr>
<td>Avg CI coxph strata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>$0.00010$</td>
<td>0.00996</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>$0.00510$</td>
<td>0.01273</td>
<td></td>
</tr>
</tbody>
</table>
Figure 10: Deviation from the true cumulative incidence (CI) of the subdistribution hazard model over 1000 replications at $t = 0.7$, illustrated per cohort for the 1st event. Covariates solely influenced the sampling probability of the sample drawn for the treatment cohort. Each violin denotes the mean (●) and 95% quantile of the bias (‖).

Table 4: Results from the subdistribution hazards simulation study following covariate adjustment on the sampling bias present in the treatment cohort. Bias and root-mean-square error (RMSE) at $t = 0.7$ are shown for event 1 following 1000 replications of population generation, unbiased sampling, and covariate-adjustment.

<table>
<thead>
<tr>
<th></th>
<th>Event 1</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td><strong>Target population</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>0.00018</td>
<td>0.00648</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.00041</td>
<td>0.00715</td>
<td></td>
</tr>
<tr>
<td><strong>Sample treatment bias</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>0.00054</td>
<td>0.01413</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.17250</td>
<td>0.17292</td>
<td></td>
</tr>
<tr>
<td><strong>Wts inverse prob</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>0.00054</td>
<td>0.01413</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.06223</td>
<td>0.06500</td>
<td></td>
</tr>
<tr>
<td><strong>Avg CI coxph obs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>0.00022</td>
<td>0.01255</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.00021</td>
<td>0.01414</td>
<td></td>
</tr>
<tr>
<td><strong>Avg CI coxph strata</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>0.00105</td>
<td>0.01260</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.00047</td>
<td>0.01486</td>
<td></td>
</tr>
</tbody>
</table>
3.4 Computation time of conditional methods

As discussed in previous sections, the main disadvantage of the conditional methods is that they often suffer from a large computation time, as they often have to predict cumulative incidence curves for a large number of observations. This issue was already noticeable during preliminary development of the conditional adjustments methods for the cause-specific hazard model. As a result, the computation time of various computational methods was more closely examined, and their results are illustrated in Figure 11. These methods either predict over each individual observation ($n = 500$) or each unique subpopulation within a set of covariates ($n = 15$). Furthermore, the following prediction models were examined during this study: the cause-specific Cox regression model from the `survival` package (`coxph`), the cumulative incidence estimator from the `survival` package (`survfit`), and the computationally efficient cause-specific Cox regression model from `riskRegression` package.

![Computation time comparison](image)

Figure 11: Computation time in milliseconds (ms) over 100 replications of conditional covariate adjustment for the cause-specific hazard model ($n = 500$). The results are compared between the `survival` package and `riskRegression` package, and CI predictions were performed either over all observations or over stratified sets of covariates. Computation times are described on the logarithmic scale.

The computation times are shown to vary greatly between the methods used to perform predictions for the cause-specific hazard model. In particular, the methods associated with the cause-specific hazards Cox regression from the `survival` package were generally more associated with relatively large computation time when compared to the other methods. Stratification of covariates with the Cox regression model from the `survival` package was shown to be able to reduce some of the computational burden. However, previous results in section 3.1 have indicated that this approach could potentially come at the cost of a reduced adjustment accuracy in the presence of continuous covariates. Alternatively, one could abolish the Cox regression by performing predictions...
on stratified sets of covariates using the cumulative incidence estimator from the same package \texttt{(survfit)}. The downside of this approach is that the absence of the modeling step would prevent pooling of information to predict over missing combinations of covariates. The lack of information pooling could further reduce the adjustment accuracy in these methods. Lastly, the alternative package \texttt{riskRegression} was developed to provide efficient computation for cause-specific hazards Cox regression. These results show that this method is able to drastically reduce computation time, while also retaining the ability to perform predictions for each observation. Therefore, the prediction method from the \texttt{riskRegression} is considered the most preferable method to optimize computation efficiency in conditional covariate adjustment.

Interestingly, the differences in computation time were found to be less dissimilar between marginal and conditional methods using the \texttt{survival} package (results not shown). While some optimization was required for the cause-specific hazard model, the Fine-Gray weights applied by the subdistribution hazard model were found to be not as computationally intensive. As a result, no further optimization was necessary for any adjustment method within the subdistribution hazard model or the marginal adjustment methods in the cause-specific hazard model.
4 Application: breast cancer survival

4.1 Background

To evaluate the performance of covariate adjustment on real-world data, a collaboration was formed with the Netherlands Comprehensive Cancer Organisation. The inverse probability weights and conditional averaging over individual predicted cumulative incidence curves were applied as covariate adjustment methods to more closely examine the competing risk aspects of breast cancer survival of an observational study.

Although many treatments against breast cancer have been developed over the years, mastectomy and breast conserving therapies have been among the most prominent treatment strategies up till this day. These methods differ in the degree to which they physically remove breast tumour tissue. Mastectomy involves the removal of all breast tissue to prevent the development of new tumours, whereas breast-conserving therapy is restricted to the removal of tumour tissue followed by a dose of radiation. Between these two procedures, breast-conserving therapy is often only recommended during early-stage diagnosis of breast cancer. It is considered to be less surgically invasive for the patient (Rahman, 2011). Early randomized controlled trials were performed to examine the efficacy of these surgical procedures on the long-term survival of breast cancer patients, and observational studies were carried out to generalize these results to the current world population (Fisher et al., 2002; Veronesi et al., 2002). This allowed researchers to examine the efficacy of treatment strategies on long-term survival in patients that were underrepresented during randomized controlled trials, such as elderly patients. In contrast to clinical expectations, observational studies indicated an increased survival for patients who had undergone breast-conserving therapy, despite it being considered to be less efficient at tumour removal (Hartmann-Johnsen et al., 2015; van Maaren et al., 2016).

A recent study by van Maaren et al. (2019) aimed to address the role of confounders in the findings previously reported by observational studies. This study aimed to control for covariate balance between the cohorts by modeling covariates using a multivariate Cox regression model, and by using inverse probability weighted Cox regression. The results from this study remained inconclusive on the potential differences in long-term survival between the procedures. For this study, researchers made use of the 10-year distant metastasis free endpoint, which is a commonly used measure to compare treatment efficacy on survival in cancer research. This composite endpoint is used to represent failure from either metastasis or death, depending on whichever occurs first. Although the measure itself is a convenient representation of overall therapeutic efficacy, it fails to encompass the competition between metastasis and death for first occurrence. This has important consequences for the interpretation of both treatment effects and confounders. For instance, differences between treatment cohorts may result from disproportional difference between the event types. In composite endpoints, a reduction in one of the outcomes may be obscured by an increase in occurrence of a competing event. Furthermore, the covariates described by van Maaren et al. (2019), including age, tumour characteristics, therapeutic application, and patient origin may be specifically associated with
a certain cause of failure. For example, based on the general knowledge of age-specific survival rates, it is safe to assume that death may be a more common clinical endpoint in older cancer patients compared to younger cancer patients. Therefore, adjusted cumulative incidence curves for competing risk analysis might allow us to more closely examine the underlying competition between endpoints, while retaining an unbiased comparison between treatment cohorts.

4.2 Data characteristics

Breast cancer survival data was obtained from a collaboration with the Netherlands Comprehensive Cancer Organisation. The same data set was previously used in an earlier study by van Maaren et al. (2019). The data consisted of 8879 patient records, which were acquired from the Netherlands Cancer Registry. All women included in this study received a diagnosis of primary invasive breast cancer in 2005 (T1-2N0-1). Patients were treated by a hospital in the Netherlands, during which they underwent either mastectomy or breast conserving surgery with adjuvant radiation therapy. Patient-, tumour-, treatment and hospital-related characteristics have been summarized in Table 5. Characteristics related to tumour staging were based on the pathological tumour characters, and described in accordance with the 6th edition of the tumour size, node, and metastasis classification system (TMN) for breast cancer (Singletary and Greene, 2003). Tumour grade, topology, and morphology were classified based on the 3rd edition of the International Classification of Diseases for Oncology (Fritz et al., 2000). The patients social economic status (SES) was determined at the time of diagnosis, and obtained per zip code area from the publicly available economic status records of The Netherlands Institute for Social Research.

Table 5: Tumour-, patient-, treatment-, and hospital-related characteristics specified per treatment cohort. A logistic regression model was used to predict treatment allocation, including all characteristics as covariate effects. A $\chi^2$-test statistic was computed for each covariate within the complete model, and converted to a $p$ value. Covariate effects on treatment allocation were considered statistically significant at $p < 0.05$.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Breast-conserving therapy ($n = 5417$)</th>
<th>Mastectomy ($n = 3462$)</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 40</td>
<td>328 (6.1)</td>
<td>221 (6.4)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>40-49</td>
<td>978 (18.0)</td>
<td>534 (15.4)</td>
<td></td>
</tr>
<tr>
<td>50-59</td>
<td>1690 (31.2)</td>
<td>783 (22.6)</td>
<td></td>
</tr>
<tr>
<td>60-69</td>
<td>1506 (27.8)</td>
<td>690 (19.9)</td>
<td></td>
</tr>
<tr>
<td>70-79</td>
<td>787 (14.5)</td>
<td>694 (20.1)</td>
<td></td>
</tr>
<tr>
<td>≥ 80</td>
<td>128 (2.4)</td>
<td>540 (15.6)</td>
<td></td>
</tr>
<tr>
<td>Mean tumour size in mm (IQR)</td>
<td>16 (11-20)</td>
<td>20 (13-25)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Breast-conserving therapy $(n = 5417)$</td>
<td>Mastectomy $(n = 3462)$</td>
<td>p value</td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>---------------------------------------</td>
<td>-------------------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>T stage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>4128 (76.2)</td>
<td>1922 (55.5)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>T2</td>
<td>1289 (23.8)</td>
<td>1540 (44.5)</td>
<td></td>
</tr>
<tr>
<td><strong>N stage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>4126 (76.2)</td>
<td>2283 (65.9)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>N1</td>
<td>1291 (23.8)</td>
<td>1179 (34.1)</td>
<td></td>
</tr>
<tr>
<td>Mean number of positive lymph nodes (IQR)</td>
<td>0 (0-0)</td>
<td>1 (0-1)</td>
<td>0.329</td>
</tr>
<tr>
<td><strong>Grade</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1460 (26.9)</td>
<td>655 (18.9)</td>
<td>0.087</td>
</tr>
<tr>
<td>2</td>
<td>2340 (43.2)</td>
<td>1557 (45.0)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1407 (26.0)</td>
<td>1037 (30.0)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>210 (3.9)</td>
<td>213 (6.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Multifocality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>4963 (91.6)</td>
<td>2480 (71.6)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Yes</td>
<td>343 (6.3)</td>
<td>881 (25.5)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>111 (2.1)</td>
<td>101 (2.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Histological tumour type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ductal</td>
<td>4513 (83.3)</td>
<td>2664 (76.9)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Lobular</td>
<td>448 (8.3)</td>
<td>483 (14.0)</td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>174 (3.2)</td>
<td>182 (5.3)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>282 (5.2)</td>
<td>133 (3.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Sublocalisation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer quadrants</td>
<td>2717 (50.2)</td>
<td>1525 (44.0)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Inner quadrants</td>
<td>1164 (21.5)</td>
<td>581 (16.8)</td>
<td></td>
</tr>
<tr>
<td>Central parts</td>
<td>316 (5.8)</td>
<td>318 (9.2)</td>
<td></td>
</tr>
<tr>
<td>Overlapping lesions</td>
<td>1125 (20.8)</td>
<td>949 (27.4)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>95 (1.7)</td>
<td>89 (2.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Laterisation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>2810 (51.9)</td>
<td>1777 (51.3)</td>
<td>0.091</td>
</tr>
<tr>
<td>Right</td>
<td>2607 (48.1)</td>
<td>1685 (48.7)</td>
<td></td>
</tr>
<tr>
<td><strong>Estrogen status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>883 (16.3)</td>
<td>600 (17.3)</td>
<td>0.775</td>
</tr>
<tr>
<td>Positive</td>
<td>4441 (82.0)</td>
<td>2792 (80.7)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>93 (1.7)</td>
<td>70 (2.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Progesteron status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>1581 (29.2)</td>
<td>1074 (31.0)</td>
<td>0.268</td>
</tr>
<tr>
<td>Positive</td>
<td>3524 (65.0)</td>
<td>2125 (61.4)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>312 (5.8)</td>
<td>263 (7.6)</td>
<td></td>
</tr>
<tr>
<td>Characteristics</td>
<td>Breast-conserving therapy ($n = 5417$)</td>
<td>Mastectomy ($n = 3462$)</td>
<td>$p$ value</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>----------------------------------------</td>
<td>--------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>HER2 status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>3875 (71.5)</td>
<td>2318 (66.9)</td>
<td>0.116</td>
</tr>
<tr>
<td>Unclear</td>
<td>297 (5.5)</td>
<td>196 (5.7)</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>597 (11.0)</td>
<td>454 (13.1)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>648 (12.0)</td>
<td>494 (14.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Targeted therapy (trastuzumab)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>5170 (95.4)</td>
<td>3302 (95.4)</td>
<td>0.253</td>
</tr>
<tr>
<td>Yes</td>
<td>247 (4.6)</td>
<td>160 (4.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Adjuvant systematic therapy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>2978 (55.0)</td>
<td>1463 (42.3)</td>
<td>0.071</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>555 (10.2)</td>
<td>305 (8.8)</td>
<td></td>
</tr>
<tr>
<td>Endocrine therapy</td>
<td>879 (16.2)</td>
<td>965 (27.9)</td>
<td></td>
</tr>
<tr>
<td>Both</td>
<td>1005 (18.6)</td>
<td>729 (21.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Axillary lymph node dissection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3809 (70.3)</td>
<td>1501 (43.4)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Yes</td>
<td>1608 (29.7)</td>
<td>1961 (56.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Social economic status (score)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1482 (27.4)</td>
<td>1052 (30.4)</td>
<td>0.051</td>
</tr>
<tr>
<td>Medium</td>
<td>2141 (39.5)</td>
<td>1389 (40.1)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>1794 (33.1)</td>
<td>1021 (29.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Hospital type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>1977 (36.5)</td>
<td>1407 (40.7)</td>
<td>0.015</td>
</tr>
<tr>
<td>Top-clinical</td>
<td>2952 (54.5)</td>
<td>1742 (50.3)</td>
<td></td>
</tr>
<tr>
<td>Academic</td>
<td>488 (9.0)</td>
<td>313 (9.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Hospital region</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1078 (19.9)</td>
<td>578 (16.7)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>B</td>
<td>320 (5.9)</td>
<td>218 (6.3)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>409 (7.6)</td>
<td>203 (5.8)</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>793 (14.6)</td>
<td>436 (12.6)</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>487 (9.0)</td>
<td>370 (10.7)</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>631 (11.6)</td>
<td>634 (18.3)</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>355 (6.6)</td>
<td>255 (7.4)</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>556 (10.3)</td>
<td>345 (10.0)</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>788 (14.5)</td>
<td>423 (12.2)</td>
<td></td>
</tr>
</tbody>
</table>

Brackets indicate the percentage (%) of observations that correspond to a certain group, $IQR$ denotes the interquartile range for continuous covariates.
In this study, a total of seven competing outcomes of breast cancer were considered, illustrated by Figure 12, which further extends on the event types that were previously included by [van Maaren et al. (2019)] to obtain the 10-year distant metastasis free survival. These endpoints have been classified in three major outcomes: mortality, recurrence, and second primary tumour development. Data on patient outcomes were obtained by combining the patient medical records (recurrence and development of second primary tumour), and the personal records database (vital state). Patients were followed from the time of diagnosis up until the first of February 2017, after which they were censored if no event had taken place. To evaluate the competition for first occurrence, the event with the smallest event time was chosen as the observed event, and the observation time was defined as the time in days between the moment of diagnosis up until the occurrence of the observed event. Furthermore, if patients encountered multiple events within a close time proximity of three months, then the most clinically severe event was indicated as the observed event. This was only the case for a small fraction of patients within the sample (<100).

Figure 12: Stage diagram describing the possible endpoints in breast cancer survival considered for the cause-specific hazard model.
4.3 Covariate-adjusted cumulative incidence curves

To replicate the analysis conditions as described by van Maaren et al. (2019), incomplete cases were removed prior to covariate adjustment, which resulted in a remainder of 6680 observations. Both the marginal and conditional approaches used for covariate adjustment, assumed that the covariate distribution from the entire sample population represented the target distribution. This setting was similar to the opposing selection scheme described in section 3.1. To gain an understanding of adjustment effects on all outcomes, the cause-specific hazard model was applied to estimate the cumulative incidence curve for each individual event. The estimated effects of tumour-, patient-, treatment-, and hospital-related characteristics on treatment allocation and cause-specific hazard can be found in Appendix B.

Inverse probability weights were constructed as described previously by performing logistic regression on the cohort assignment as a function of the set of characteristics illustrated in Table 5. The distribution of propensity scores is shown in Figure 13. The propensity scores for individuals within mastectomy cohort appear to be uniformly distributed. In contrast, the distribution of propensity scores for individuals assigned to breast-conserving therapy shows a higher frequency of larger propensity scores. Comparing the distribution of propensity scores between mastectomy and breast-conserving therapy reveals a noticeable difference in the shape between distributions. More specifically, certain subgroups appear to be highly represented within the breast-conserving therapy cohort, but are rather impartially assigned to the mastectomy cohort. This is further emphasized by the results from the $\chi^2$-tests, described by Table 5, which shows that significant covariate effects on treatment allocation were mostly related to characteristics related to tumour development. For instance, within the breast-conserving cohort, patients in the early stages of tumour development (T1) are shown to appear more frequently (76.2%) than patients in later stages (T2; 23.8%). In contrast, the distribution of tumour stages in patients assigned to the mastectomy cohort, appears to more uniform (55.5% to 45.5%). Therefore, these results indicate the presence of covariate imbalance between cohorts which may confound the relationship between treatment and outcome if left unaddressed.

To perform adjustment via inverse probability weighting, the inverse of individual propensity scores were taken and used as weights to compute the weighted cumulative incidence estimator. The inverse probability weighted estimator was compared against the unweighted estimator for all endpoints, and the results are illustrated by Figure 14. These results show that local and regional recurrence, along with DCIS and miscellaneous tumour development, all have a noticeably small probability of first occurrence. Furthermore, their probability of occurrence is shown to be indifferent to the type of treatment procedure or adjustment method applied on the data set. Closer examination of the remaining endpoints indicates a small difference between treatment procedures on the first occurrence of metastasis and the development of invasive second primary tumours. Both of these endpoints show a negligible response to covariate adjustment by inverse probability weighting. In contrast, the results for mortality show that intervention by mastectomy results in a higher probability of death when compared to breast-conserving
therapy. Interestingly, covariate-adjustment via inverse probability weighting reduces the cumulative incidence of death in the mastectomy cohort, while the cumulative incidence for the breast-conserving cohort seems to be largely unaffected. This difference may be explained from the differences in the distribution of propensity scores between the two cohorts, as illustrated by Figure 13. As discussed previously, there is a noticeable peak in the propensity scores of the breast-conserving cohort. As a result, many individuals within this cohort will receive the same weight, which would then correspond with a small shift in the cumulative incidence curve for this cohort. In contrast, the propensity scores of the treatment cohort are more uniformly distributed, which resulted in a higher variability of the weights between individuals within this cohort, corresponding to a larger shift in the observed cumulative incidence curve.

To compare the differences between marginal and conditional adjustment methods, the unadjusted cumulative incidence curve was also compared to an average over individual predicted cumulative incidence curves by the cause-specific Cox regression model. These results are illustrated by Figure 15. This figure shows that covariate adjustment in tumour recurrence and second primary tumour development remain in line with the previous findings. Furthermore, conditional covariate-adjustment within the mastectomy cohort resulted in a decrease in mortality as the event of first occurrence. This effect of the conditional covariate adjustment corresponds to the results found after marginal covariate adjustment with inverse probability weights. Interestingly, in contrast to the
previous findings for the inverse probability weighted curve, conditional adjustment on the breast-conserving cohort led to a larger increase in the probability of death as first occurrence. This may indicate a difference in the goodness of fit between the two methods, leading up to slight differences in their results. This is supported by the occurrence convergence issues and violated proportional hazards assumption for some of the covariates within the conditional method, thereby indicating a poor fit. In contrast, no issues were encountered during the construction nor application of the inverse probability weights. Therefore, the inverse probability weighted curves may be considered more reliable compared to those obtained from the conditional adjustment method.

To summarize, the application of either adjustment method led to a reduction in the difference in mortality between intervention procedures, but showed little effect on the occurrence of the other endpoints. This may indicate that the covariate imbalance present between these two cohorts mainly affects mortality, but not so much the occurrence or recurrence of the disease. Therefore, if a difference between cohorts was expected on disease progression, then these results indicate that the current collection of covariates may not be able to resolve any confounding effects that may have resulted in a similar efficacy between treatment procedures.
Figure 14: Comparison of unweighted (dotted line) and inverse probability weighted (solid line) cumulative incidence curves, separately shown for mastectomy (blue) and breast-conserving therapy (red). Cumulative incidences curves have been developed using the cause-specific hazard model for each competing endpoint within (A) mortality, (B) recurrence, and (C) second primary tumours.
Figure 15: Comparison between the unweighted cumulative incidence curves (dotted line) and Cox-predicted average of individual cumulative incidences curves (solid line). The effect of covariate-adjustment is shown for mastectomy (blue) and breast-conserving therapy (red). Cumulative incidences curves have been developed using the cause-specific hazard model for each competing endpoint within (A) mortality, (B) recurrence, and (C) second primary tumours.
5 Discussion

5.1 Performance and limitations of adjustment methods

In this project, we have demonstrated the performance of marginal and conditional adjustment methods for cumulative incidence curves that originate from competing risk models. We have found that the adjustment methods performed well for both types of competing risk models. From the methods evaluated in the simulation study, the highest adjustment accuracy and most reproducible results were achieved by conditional averaging over individual predicted cumulative incidence curves from a Cox regression model. Furthermore, this method was also proven to be a flexible adjustment method, as it was able to handle each sampling scheme by simply modifying the set of individuals over which the predicted curves were constructed. Unfortunately, we have also confirmed that this method may suffer from a large computation time in the presence of many observations, which may hinder practical application. These results were in line with the advantages and limitations mentioned by Therneau et al. (2015). As a result, practical use of this method may largely depend on the availability of computationally efficient methods in order to make the computation time more manageable. Furthermore, as Therneau et al. (2015) pointed out, the construction of standard errors for averaged curves is complex, and has yet to be resolved by future studies.

Similar to the conditional approach, inverse probability weighting displayed a high adjustment accuracy when there is counteractive selection on covariates between cohorts. Furthermore, this method was found to be easily executable by simply adding the computed inverse probability weights as sampling weights to the weighted cumulative incidence estimator, thereby making it a faster method to accurately perform covariate-adjustment when compared to the conditional approach. Importantly, standard errors for weighted cumulative incidence curves have already been developed by the author of the survival package in R (Therneau, 2020). Furthermore, it has been mentioned previously by Austin and Stuart (2015) that the inverse probability weights require modifications to handle reference groups (i.e. groups with a covariate distribution that already reflects that of the target distribution). If left unaddressed, the weights would introduce bias in the reference group while demonstrating severely compromised adjustment accuracy in the remaining groups. To overcome this issue, stabilized weights were used as proposed by Austin and Stuart (2015) to prevent the introduction of bias in the reference group. Our findings show that these weights indeed proved to be successful in fully resolving the deviation from the true curve in the remaining groups. Attempts were made to improve these weights, for instance by multiplying them with the ratio of the marginal probabilities between the reference group and biased group, as described by Goetghebeur et al. (2020). However, preliminary attempts showed that this improvement did not affect the outcome of the adjustment since the individuals within the same group would be multiplied by the same constant (results not shown). Therefore, more work is required to further examine new approaches to increase the flexibility of inverse probability weights to address particular sampling schemes.
Lastly, the simulation study indicated that the stratification of data, often applied to reduce computation time, may be detrimental to the accuracy of the adjustment methods. The loss in adjustment accuracy may likely be attributed to the discretization of continuous covariates, which introduces discretization error. Furthermore, attempts to apply stratification on the large number of covariates from the breast cancer survival data set led to computational issues, which originated from the profusion of unique subpopulations produced by this method. Consequently, this would also debilitate the intended purpose of stratification, which was to reduce the number of predicted curves in effort to reduce computation time. Hence, careful consideration should be taken when using methods that involve stratification in the presence of continuous covariates or a large number of covariates. Therefore, our study found that the adjustment methods related to stratification were to be considered the least favourable approach out of all adjustment methods examined in this study.

5.2 Perspective on breast cancer survival

Covariate adjustment was performed on the breast cancer survival data that was previously analyzed in a study by van Maaren et al. (2019). The aim of this application was to observe the performance on the adjustment methods on real-world data, and to provide deeper insights on the competing risks that underlie the 10-year distant metastasis free survival previously described by van Maaren et al. (2019). It was found that the treatment procedures and covariate adjustment methods mainly affected the cumulative incidence of mortality. It was shown by the unadjusted cumulative incidence curve for mortality that mastectomy would more often result in death as the event of first occurrence. Covariate adjustment, either by inverse probability weighting or conditional averaging of predicted curves, led to a reduction in the cumulative incidence of mortality for patients within the mastectomy cohort. Furthermore, conditional averaging of predicted curves also increased the probability of death as the event of first occurrence for patients that had undergone breast-conserving therapy. In contrast, inverse probability weighting resulted in a negligible increase in mortality as the event of first occurrence in the breast-conserving cohort. As we have previously discussed, this is likely the result from having only a few patients with a larger weight in this particular cohort. The differences in outcome between marginal and conditional covariate adjustment emphasize the importance of the conceptual differences between these two methods. As previously discussed by Therneau et al. (2015), inverse probability weighting aims to explain the differences in treatment allocation as a function of the covariates. In contrast, conditional methods aim to explain the outcome as a function of the covariates. Consequently, the choice between each adjustment method relies on which model is able to better represent the underlying mechanisms that drive sampling and outcome. In our analysis, we found no issues in the logistic regression modeling of treatment allocation to construct the inverse probability weights. In contrast, the high complexity of the cause-specific hazards Cox regression model led to convergence issues, and indicated violation of the proportional hazards assumption for some of the covariates included in the model. Therefore, based on these observations we assume that the inverse probability weighted cumulative
incidence curve provided the most reliable results for this particular data set.

Regardless of the conceptual differences between the two methods, both methods indicated that adjustment, based on the current set of characteristics, only affected the cumulative incidence curve for mortality, and did not seem to affect the occurrence of outcomes related to recurrence or development of the disease. This is important since the previous study by van Maaren et al. (2019) indicated the potential presence of unmeasured confounding, which may underlie the inexplicable differences between the findings of observational studies compared to randomized controlled trials. Our results indicate that the adjustment procedure only affected the outcome of mortality within the 10-year distant metastasis free survival composite endpoint, which was used by van Maaren et al. (2019). Therefore, our results demonstrate that a large portion of unmeasured confounding may be particularly related to further progression into invasive breast cancer. Future work will be required to expand on the individual endpoints by comparing them with previous results from randomized controlled trials.

5.3 Future directions

Our study has shown that the adjustment methods used to address covariate imbalance between survival curves, as described by Therneau et al. (2015), can also be applied to produce covariate-adjusted cumulative incidence curves for competing risks. These findings offer a promising perspective to further expand on covariate adjustment in higher complexity settings, such as the modeling of time-dependent covariates and multi-state frameworks. As we have shown, the application of the adjustment methods will only require the model to be able to integrate weights in its estimation procedure for marginal adjustment, or is able to produce the predicted curves required for conditional adjustment. Fortunately, these features are largely available for most statistical methods. Furthermore, as we have discussed during the application of the adjustment methods on real-world data, it is important that either adjustment method is chosen based on whether one is more able to explain the covariate effects directly on treatment allocation, which could be derived from information about the sampling procedure, or whether covariate effects explain the outcome well. This may also bring into question the potential role of correct model specification, as the performance of either adjustment method may be prone to different aspects of model misspecification. Future work may be done to shed more light on this aspect. Lastly, future research may also be dedicated to combine these adjustment methods into a single doubly robust estimator, which may be able to overcome the current limitations from either model by combining the two. A recent study by Zhang and Schaubell (2012) has already provided promising results when using the doubly robust estimator to correct for covariate imbalance between cohorts in regular Cox regression models. Hence, future studies may aim to further expand on these results by producing doubly robust cumulative incidence curves.
5.4 Conclusion

To conclude, we have demonstrated that the covariate adjustment methods described by [Therneau et al., 2015] and others can be used to construct covariate-adjusted cumulative incidence curves for competing risk models. Functions have been created in R to construct covariate-adjusted cumulative incidence curves, both for the cause-specific hazard and subdistribution hazard model. The conditional approach has most notably shown promising results during the simulation study by providing high adjustment accuracy and well reproducible results for each specific type of covariate imbalance. Furthermore, we would like to stress the conceptual differences between marginal and conditional methods that may heavily impact the results of their application. Careful consideration should be made when choosing either method with regards to the relationship between the covariates, treatment allocation, and outcome. We have demonstrated that the covariate adjustment methods were able to address the covariate imbalance present in breast cancer survival data, and may have been able to more shed light on the role of covariate imbalance between the various endpoints that underlie the 10-year distant metastasis free endpoint. Our results may provide new insights on the unresolved causal relationship between treatment procedures and breast cancer outcomes.
References


Ganiyu A. Rahman. Editorial: Breast conserving therapy: A surgical technique where little can mean more, 2011. ISSN 20068808.


49


Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi,

A R-code

A.1 Packages and seed

All R libraries used in this study were publicly available through the Comprehensive R Archive Network (CRAN). Below is an overview of the most important packages required to run the code used in this study, and their corresponding authors:

- **survival** - Therneau (2020)
  Cause-specific hazard and subdistribution hazard modeling, performing model diagnostics, and prediction of cumulative incidence curves

- **riskRegression** - Gerds and Ozenne (2020)
  Efficient computation and cumulative incidence curve prediction for the cause-specific hazard model

- **microbenchmark** - Mersmann (2019)
  Providing execution time measurements and comparisons between R functions

- **tidyverse** - Wickham et al. (2019)
  A collection of the data manipulation and visualization packages, most notably tidyr and ggplot2

- **ggpubr** - Kassambara (2020)
  Producing arrangements of plots constructed by ggplot2

- **scales** - Wickham and Seidel (2020)
  Additional functions to enable complex graphical scaling in ggplot2

- **openxlsx** - Schauberger and Walker (2020)
  Loading and storage of output in Excel .xlsx files

- **xtable** - Dahl et al. (2019)
  Exportation of simple table output to a LaTeX

- **tables** - Murdoch (2020)
  Construction of complex table outputs and exportation to LaTeX
A.2 Supporting functions

long_sfc(): extract cumulative incidence curves from a survfit object

This helper function can be used to quickly extract the strata, times, and cause-specific probabilities from a `survfit` object for each event in a long format, such that it can be easily plotted (e.g. by ggplot).

- **object**: a `survfit` object from which one would like to extract the cumulative incidence estimator.
- **stack**: a boolean operator (`TRUE` / `FALSE`), which determines whether to return the marginal probabilities per event. If `TRUE`, which is by default, the cause-specific probabilities are returned as stacked probabilities for subsequent events.

```r
long_sfc <- function(object, stack = F) {
  # Extract components from survfit
  sf_CI <- object$pstate[, -1]
  time <- object$time
  strata <- rep(names(object$strata), object$strata)
  if(is.null(strata)){
    strata <- rep(1, length(time))
  }
  tmp <- data.frame(time, strata = factor(strata), sf_CI)
  colnames(tmp)[3:ncol(tmp)] <- object$states[-1]
  # Return subsequent events as a stacked probability
  if(stack == T){
    tmp2 <- tmp
    for(i in 4:ncol(tmp)){
      tmp2[i] <- rowSums(tmp[3:i])
    }
    tmp <- tmp2
  }
  # Construct long format and return
  tmp_long <- gather(tmp, event, prob, 3:ncol(tmp))
  return(tmp_long)
}
```
A.3 Data generation and biased sampling

`sim_comp()`: simulation of data under the cause-specific hazard model

The purpose of this function is to simulate a competing risks data set from the cause-specific hazard model, with the following set of predictors: a strata (cohort), a discrete covariate (mutation), and a continuous covariate (age). The observation times are simulated according to the latent competing risk approach, where the first event can censor the event of interest.

- \( N \): the number of observations to be generated.
- \( k \): the number of competing outcomes to be generated.
- `base_haz`: a baseline hazard for each event.
- `beta_coef`: a list where each listed element is a vector of covariate coefficients associated with event \( k \). The first element in the vector is the effect size of the cohort, the second element the effect of the discrete covariate, and the third element the effect of the continuous covariate.

```r
sim_comp <- function(N, k, base_haz, beta_coef) {
  # covariate simulation
  age <- rnorm(N)
  mutant_type <- sample(c(0, 1, 2), N, c(1/3, 1/3, 1/3), replace = T)
  cohort <- rbinom(N, 1, 0.5)
  X <- as.matrix(data.frame(cohort, mutant_type, age))

  # simulate event times per event k
  event_times <- matrix(NA, nrow = N, ncol = k)
  for(i in 1:k) {
    haz_fun <- base_haz[i] * exp(X %*% beta_coef[[i]])
    event_times[, i] <- rexp(length(haz_fun), rate = haz_fun)
  }

  # simulate competition for first occurrence between k events
  obs_time <- apply(event_times, 1, min)
  status <- apply(event_times, 1, function(x) which(x == min(x)))

  # Add censoring
  censor_time <- runif(N, quantile(obs_time, 0.20), quantile(obs_time, 0.95))
  status <- ifelse(censor_time < obs_time, 0, status)
  obs_time <- pmin(censor_time, obs_time)

  # construct output
  sim_data <- data.frame(obs_time, status = factor(status),
                          cohort = factor(ifelse(cohort == 1, "treat", "control")),
                          mutation = factor(paste("stage", mutant_type)),
                          age)

  return(sim_data)
}
```

54
**sim_sh()**: simulation of data under the subdistribution hazard model

This function allows us to simulate event times under the subdistribution hazard model with a static number of two competing events. This function was made based on an algorithm from Prof. dr. H. Putter to simulate subdistribution hazard data with a single continuous covariate.

- **N**: the number of observations to be generated.
- **shapes**: a vector of length 2 with shape parameters for each event.
- **rates**: a vector with of length 2 with rate parameters for each event.
- **beta_coef**: a list of length two where each listed element is a vector of covariate coefficients for each event. The first element in the vector is the effect size of the cohort, the second element the effect of the discrete covariate, and the third element the effect of the continuous covariate.
- **p**: the baseline probability of occurrence for the first event.

```r
sim_sh <- function(N, shapes, rates, beta_coef, p) {
  # covariate matrix X simulation
  age <- rnorm(N)
  cohort <- rbinom(N, 1, 0.5)
  mutant_type <- sample(c(0, 1, 2), N, c(1/3, 1/3, 1/3), replace = T)
  X <- as.matrix(data.frame(cohort, mutant_type, age))

  # Generate a random sample of size n for each event
  HR1 <- exp(X %*% beta_coef[[1]])
  HR2 <- exp(X %*% beta_coef[[2]])
  p1 <- 1 - (1-p)^HR1
  u <- runif(N)

  # Cause 1
  wh1 <- which(u <= p1)
  u1 <- ( (1 - (1 - u)^(1/HR1))/p )[wh1]
  t1 <- ( -log(1-u1) / rates[1])^(1/shapes[1])

  # Cause 2
  wh2 <- which(u > p1)
  u2 <- ( (u - p1) / (1 - p1) )[wh2]
  t2 <- ( -log(1-u2) / rates[2])^(1/shapes[2])

  # Return resulting data frame
  t12 <- d12 <- rep(NA, N)
  t12[wh1] <- t1; t12[wh2] <- t2
d12[wh1] <- 1; d12[wh2] <- 2

  # censoring
  c12 <- runif(N, quantile(t12, 0.2), quantile(t12, 0.95))
t12 <- pmin(t12, c12)
d12 <- ifelse(t12 < c12, d12, 0)

  res <- data.frame(id = 1:N,
                   time = t12,
                   status = factor(d12),
                   cohort = factor(ifelse(cohort == 1, "treat", "control")),
                   mutation = factor(paste("stage", mutant_type)),
                   )
}
```
sample_cov(): biased sampling from simulated competing risk data

This function was created to perform biased sampling on the data generated from either competing risk model. Adjustment of the regression coefficients in beta_coef determines the direction of selection, where beta_coef = 0 results in strict random selection on that covariate, beta_coef > 0 results in positive selection for higher covariate values, and beta_coef < 0 results in positive selection for lower covariate values. The effects of the regression coefficients are reversed for the control cohort.

data: a generated competing risk data set.
n: the number of observations to be sampled per cohort.
beta_coef: a vector of coefficients that tune the non-random selection in the treatment cohort. The first element affects selection of the discrete covariate, and the second element affects the selection of the continuous covariate.
bias: a character string indicating whether bias should only occur in the simulated treatment cohort "treat", the default, or in either cohort "both".

```r
sample_cov <- function(data, n, beta_coef, bias = "treat"){
  # construct design matrix for age and mutation
  mutation_risk <- ifelse(as.numeric(data$mutation) - 1 == 0, -1,
                          ifelse(as.numeric(data$mutation) - 1 == 1, 0, 1))
  age_risk <- data$age
  # construct linear predictor and calculate probability selection per subject
  prob_select <- exp(lin_pred) / (1 + exp(lin_pred))
  # perform selection (separate cohorts to ensure equal n samples)
  control_obs <- which(data$cohort == "control")
  treat_obs <- which(data$cohort == "treat")
  sample_rows_treat <- sample(treat_obs, n, replace = F, prob = prob_select[treat_obs])
  if(bias == "treat"){
    sample_rows_control <- sample(control_obs, n, replace = F)
  } else if(bias == "both") {
    sample_rows_control <- sample(control_obs, n, replace = F, prob = 1 - prob_select[control_obs])
  }
  # combine selected rows in a subset
  select_subset <- data[sort(c(sample_rows_treat, sample_rows_control)), ]
  return(select_subset)
}
```
A.4 Adjustment methods

**est_ipw()**: inverse probability weighting

This function constructs inverse probability weights for the cause-specific hazard model by modeling group assignment in `strata` as a function of a set of covariates. Including between the covariates may allow for more specific fine-tuning of the weights.

- **formula**: a `formula` object in the form of `y ~ x`, where `x` represents the collection of covariates which should be balanced between the levels of `y`.
- **data**: an optional `data.frame` used to interpret the variables described by `formula` argument.
- **ref**: a character input to specify a reference group from `x`. If this argument is `NULL`, which is by default, the entire sample is used as a reference instead.

```r
est_ipw <- function(formula, data = NULL, ref = NULL) {
  # Deconstruct formula object
  resp <- paste(attr(terms(formula, data = data), which = "variables")[[2]])
  resp_obs <- factor(model.frame(formula, data = data)[, 1])
  pred <- paste(attr(terms(formula, data = data), which = "term.labels"),
                 collapse = " + ")

  # Calculate weights
  ipw_ref <- matrix(NA, nrow = length(resp_obs), ncol = nlevels(resp_obs))
  for(i in 1:nlevels(resp_obs)) {
    ipw_formula <- formula(paste0("I(" , resp, " == " , levels(resp_obs)[i], "\"", " ~ " , pred))
    ipw_ref[, i] <- 1 / predict(glm(ipw_formula, data = data, family = "binomial"),
                              type = "response")
  }
  if(!is.null(ref)){
    ref_formula <- formula(paste0("I(" , resp, " == " , ref, "\"", " ~ " , pred))
    ref_prob <- predict(glm(ref_formula, data = data, family = "binomial"),
                        type = "response")
    ipw_ref <- apply(ipw_ref, 2, function(x) ref_prob * x)
  }

  # Match observed cohort with ipw_ref matrix
  ipw <- numeric(nrow(ipw_ref))
  for(j in 1:nrow(ipw_ref)){
    ipw[j] <- ipw_ref[j, which(resp_obs[j] == levels(resp_obs))]
  }
  return(ipw)
}
```

57
est_wt(): external reference group weighting

This function assigns weights based on the covariate distribution of an external data set. Continuous covariates are automatically binned for this approach, which is required to obtain the joint probabilities for each set of covariates. The degree of binning can be controlled by the n_qntl argument, which determines the number of quantiles used to split a continuous covariate.

**tbl**

A matrix or data.frame with the covariates on which the weights should be computed.

**ref**

A matrix or data.frame with the same structure as tbl. In this external data set, covariates are in line with a desired target distribution.

**n_qntl**

An integer input used to define the number of quantiles used to automatically discretize continuous covariates. The default value is set to 5.

```r
est_wt <- function(tbl, ref, n_qntl = 5) {
    # Convert numeric variables to quantiles (discretize)
    if(is.null(ref)) {ref_tbl <- tbl} else {ref_tbl <- ref}
    if(is.atomic(tbl)){
        tbl <- data.frame(dummy = factor(rep("dummy", length(tbl))), tbl)
        ref_tbl <- data.frame(dummy = factor(rep("dummy", length(ref_tbl))), ref_tbl)
    }
    for(i in 1:ncol(tbl)){
        if(is.numeric(tbl[, i])) {
            tbl[, i] <- cut(tbl[, i],
                breaks = c(quantile(ref_tbl[, i],
                    probs = seq(0, 1, by = 1/n_qntl)),
                labels = apply(cbind(seq(0, 100, 100/n_qntl)[-n_qntl + 1]),
                    prob = seq(0, 100, 100/n_qntl)[-1]),
                include.lowest = T),
                ref_tbl[, i] <- cut(ref_tbl[, i],
                breaks = c(quantile(ref_tbl[, i],
                    probs = seq(0, 1, by = 1/n_qntl)),
                labels = apply(cbind(seq(0, 100, 100/n_qntl)[-n_qntl + 1]),
                    prob = seq(0, 100, 100/n_qntl)[-1]),
                include.lowest = T),
                include.lowest = T)
        }
    }
    # Subdivide covariates and calculate proportions
    obs_covs <- factor(apply(tbl, 1, paste, collapse = "|"))
    obs_props <- as.data.frame(table(obs_covs) / length(obs_covs))
    # Reference proportions and weight division
    ref_covs <- factor(apply(ref_tbl, 1, paste, collapse = "|"))
    ref_props <- as.data.frame(table(ref_covs) / length(ref_covs))
    match_levels <- match(obs_props$obs_covs, ref_props$ref_covs)
    ref_wts <- data.frame(group = levels(obs_props$obs_covs),
        wt = ref_props$Freq[match_levels] / obs_props$Freq)
}
```
# Assign weights to each observation
wts <- numeric(nrow(tbl))
for(j in 1:nrow(tbl)){
  wts[j] <- ref_wts[which(obs_covs[j] == ref_wts[, 1]), 2]
}
return(wts)

---

**cox_wt()**: average of individual predicted curves for cause-specific hazards

In this conditional approach, we duplicate the risk set, used to construct a cause-specific Cox regression model, to each group within the specified strata, and calculate the average of individual predicted cumulative incidence curves from the risk set. A reference group within the strata variable may be specified to only duplicate individuals from the risk set of this group to other groups.

- **formula**: a **formula** object containing a `Surv()` object as the response, and the strata and covariates as the predictors.
- **strata**: a character string used to identify the strata within the **formula** object.
- **data**: an optional **data.frame** used to interpret the variables described by **formula** argument.
- **ref**: a character input to specify a reference group from **x**. If this argument is **NULL**, which is by default, the entire sample is used as a reference instead.
- **times**: a series of user-defined time points for which the cumulative incidence should be computed per strata. If **NULL**, which is by default, the unique observation times from the data are used instead.

```r
cox_wt <- function(formula, strata, ref = NULL, data = NULL, times = NULL){
  # Extract components from formula object
  tmp_frame <- model.frame(formula, data)
  status_levels <- levels(tmp[, 1])
  tmp <- cbind(data.frame(id = 1:nrow(tmp),
                         time = tmp[, 1][, 1],
                         status = factor(tmp[, 1][, 2]),
                         tmp[2:ncol(tmp)]))
  names(tmp)[which(names(tmp) == strata)] <- "strata"
  tmp <- tmp[, c(which(colnames(tmp) %in% c("id", "time", "status", "strata")),
                  which(!colnames(tmp) %in% c("id", "time", "status", "strata")))]
  names(tmp) <- gsub("\$", "", names(tmp))

  # Construct Cox model
  pred <- paste(names(tmp)[4:ncol(tmp)], collapse = "\$\$")
  cox_formula <- formula(paste("Surv(time, status) ~ " , pred))
  df_cox <- coxph(cox_formula, id = id, data = tmp)

  # Construct dummy data w/ every observation for each cohort
  if(is.null(ref)){
    #...
  }
```
dummy <- do.call("rbind", 
    replicate(nlevels(tmp$strata), 
    tmp[, c(-1, -4)], 
    simplify = F))
dummy <- cbind(id = factor(1:nrow(dummy)), 
    strata = rep(levels(tmp$strata), 
    each = nrow(tmp)), 
    dummy)
} else {
    dummy <- do.call("rbind", 
        replicate(nlevels(tmp$strata), 
        tmp[tmp$strata == paste0(strata, "=" ref), c(-1, -4)], 
        simplify = F))
dummy <- cbind(id = factor(1:nrow(dummy)), 
    strata = rep(levels(tmp$strata), 
    each = nrow(tmp[tmp$strata == paste0(strata, "=" ref)])) 
    dummy)
}
dummy <- dummy[, c(-3:-4)]

# Extract predicted timepoints with dummy data
sf_cox <- survfit(df_cox, newdata = dummy, se.fit = F)
if(is.null(times)){
xtime <- sort(unique(c(0, tmp$time)))
} else {
    xtime <- times
}

sf_pred <- summary(sf_cox, times = xtime, extend = T)

# Take mean curve per strata for each state
curves <- vector("list", length(sf_pred$states))
for(i in 1:length(sf_pred$states)){
    curves_i <- vector("list", nlevels(tmp$strata))
    for(j in 1:nlevels(tmp$strata)){
        index <- 1:(nrow(dummy) / nlevels(tmp$strata)) + 
            (j - 1) * (nrow(dummy) / nlevels(tmp$strata))
        curves_i[[j]] <- rowMeans(sf_pred$pstate[, index, i])
    }
    curves[[i]] <- do.call("c", curves_i)
}

# Combine states and construct output
curves_output <- do.call("cbind", curves)
colnames(curves_output) <- c("(s0)", status_levels)
output <- cbind(data.frame(time = rep(xtime, times = nlevels(tmp$strata))), 
    strata = factor(rep(levels(tmp$strata), 
    each = length(xtime))))
    curves_output)

return(output)
rr_wt(): alternative average of individual predicted curves by riskRegression

This function is a computation-efficient version of the cox_wt() function, which is used to perform an average of individual predicted curves from the cause-specific hazard model. This function runs on the computationally efficient cause-specific Cox regression model CSC() from the ‘riskRegression’ package, in effort to derive predicted cumulative incidence curves for each observation.

formula: a formula object containing a Surv() object as the response, and the strata and covariates as the predictors.

strata: a character string used to identify the strata within the formula object.

data: an optional data.frame used to interpret the variables described by formula argument.

ref: a character input to specify a reference group from x. If this argument is NULL, which is by default, the entire sample is used as a reference instead.

times: a series of user-defined time points for which the cumulative incidence should be computed per strata. If NULL, which is by default, the unique observation times from the data are used instead.

rr_wt <- function(formula, strata, ref = NULL, data = NULL, times = NULL){
  # Extract components from formula object and convert to model matrix
  tmp <- model.frame(formula, data)
  status_levels <- levels(tmp[, 1])
  tmp <- cbind(data.frame(time = tmp[, 1],
                           status = factor(tmp[, 1],
                                         levels = levels(tmp[, 1]),
                                         labels = as.character(tmp[, 2]))), tmp[2:ncol(tmp)])
  names(tmp)[which(names(tmp) == strata)] <- "strata"
  tmp$strata <- factor(paste0(strata, "=",
                                 levels(tmp$strata)))
  tmp <- tmp[, c(which(colnames(tmp) %in% c("time", "status", "strata")),
                  which(!colnames(tmp) %in% c("time", "status", "strata")))]
  names(tmp) <- gsub(".*\$", "", names(tmp))
  # Produce csc model
  pred <- paste(names(tmp)[3:ncol(tmp)], collapse = "+"
  cox_formula <- formula(paste0("Hist(time, status) ~ "
                           , collapse = "+
                           , pred))
  df_csc <- CSC(cox_formula, data = tmp)
  # Prediction times
  if(is.null(times)){
    xtime <- sort(unique(c(0, tmp$time)))
  } else {
    xtime <- times
  }
  # Construct dummy data w/ every observation for each cohort
  if(is.null(ref)){
    dummy <- do.call("rbind",
                     replicate(nlevels(tmp$strata), tmp[, -3],
                               simplify = F))
    dummy <- cbind(strata = rep(nlevels(tmp$strata), each = nrow(tmp)),
                   dummy)
  } else {
    dummy <- do.call("rbind",
                     replicate(nlevels(tmp[strata == ref]) - 1,
                               tmp[, -3],
                               simplify = F))
    dummy <- cbind(strata = rep(nlevels(tmp[strata == ref]) - 1,
                               each = nrow(tmp)),
                   dummy)
  }
  return(df_csc, dummy, xtime)
}
replicate(nlevels(tmp$strata),
  tmp[tmp$strata == paste0(strata, "", ref), -3],
  simplify = F))
dummy <- cbind(strata = rep(levels(tmp$strata),
  each = nrow(tmp[tmp$strata == paste0(strata, "", ref)])),
  dummy)
}  
dummy <- dummy[, c(-2:-3)]

# Take mean curve per strata for each status
curves <- vector("list", nlevels(tmp$status) - 1)
for(i in 1:(nlevels(tmp$status) - 1)){
csc_pred <- predict(df_csc,
  newdata = dummy,
  times = xtime,
  cause = levels(tmp$status)[-1][i])
curves_i <- vector("list", nlevels(tmp$strata))
for(j in 1:nlevels(tmp$strata)){
  index <- 1:(nrow(dummy) / nlevels(tmp$strata)) +
    (j - 1) * (nrow(dummy) / nlevels(tmp$strata))
  curves_i[[j]] <- colMeans(csc_pred$absRisk[index, ])
}
curves[[i]] <- do.call("c", curves_i)
}

# Combine states and construct output
curves_output <- do.call("cbind", curves)
curves_output <- cbind(1 - rowSums(curves_output), curves_output)
output <- cbind(data.frame(time = rep(xtime, times = nlevels(tmp$strata)),
  strata = factor(rep(levels(tmp$strata),
    each = length(xtime)))),
  curves_output)
return(output)
}
coxstrat_wt(): average of strata-predicted curves for cause-specific hazards

This function is a computation efficient version of the \texttt{cox_wt()} function, which only performs predictions on the each unique combination of covariates. This is done by creating a risk set with every unique combination of covariates, and take a weighted average over these curves. For this method to optimally work, continuous covariates are discretized based on a user-defined number of quantiles.

\begin{verbatim}
coxstrat_wt <- function(formula, strata, ref = NULL, data = NULL, n_qntl = 5, times = NULL){
  # Extract components from formula object
  tmp <- model.frame(formula, data)
  status_levels <- levels(tmp[, 1])
  tmp <- cbind(data.frame(id = factor(1:nrow(tmp)),
    time = tmp[, 1][, 1],
    status = factor(tmp[, 1][, 2])), tmp[2:ncol(tmp)])
  names(tmp)[which(names(tmp) == strata)] <- "strata"
  tmp$strata <- factor(paste0(strata, "=", tmp$strata))
  tmp <- tmp[, c(which(colnames(tmp) %in% c("id", "time", "status", "strata")),
    which(!colnames(tmp) %in% c("id", "time", "status", "strata")))]
  names(tmp) <- gsub(".*\$", "", names(tmp))

  # Discretize continuous covariates
  for(n in 5:ncol(tmp)){
    if(is.numeric(tmp[, n])) {
      tmp[, n] <- cut(tmp[, n],
        breaks = quantile(tmp[, n],
          probs = seq(0, 1, by = 1/n_qntl)),
        labels = apply(cbind(seq(0, 100, 100/n_qntl)[-1],
          seq(0, 100, 100/n_qntl)[-n_qntl - 1]),
          1, paste, collapse = "-"),
        include.lowest = T)
    }
  }

  # Reference weights
  tmp2 <- tmp
}
\end{verbatim}

\texttt{formula}:

- a \texttt{formula} object containing a \texttt{Surv()} object as the response, and the strata and covariates as the predictors.

\texttt{strata}:

- a character string used to identify the strata within the \texttt{formula} object.

\texttt{data}:

- an optional \texttt{data.frame} used to interpret the variables described by \texttt{formula} argument.

\texttt{ref}:

- a character input to specify a reference group from \texttt{x}. If this argument is \texttt{NULL}, which is by default, the entire sample is used as a reference instead.

\texttt{n_qntl}:

- an \texttt{integer} input used to define the number of quantiles used to automatically discretize continuous covariates. The default value is set to 5.

\texttt{times}:

- a series of user-defined time points for which the cumulative incidence should be computed per strata. If \texttt{NULL}, which is by default, the unique observation times from the data are used instead.
```r
tmp2$dummy <- factor(rep(1, nrow(tmp)))
if(is.null(ref)){
  ref_props <- as.data.frame(table(tmp2[, 5:ncol(tmp2)]) /
                            nrow(tmp2[, 5:ncol(tmp2)]))
  dummy <- do.call("rbind", 
                  replicate(nlevels(tmp2$strata),
                            ref_props[, -ncol(ref_props)], 
                            simplify=F))
  dummy <- cbind(id = 1:nrow(dummy), 
                  strata = factor(rep(levels(tmp2$strata),
                            each = nrow(dummy) / 
                            nlevels(tmp2$strata))), 
                  dummy)
} else {
  ref_props <- as.data.frame(table(
      tmp2[tmp2$strata == paste0(strata, 
        "=", ref), 5:ncol(tmp2)]) /
      nrow(tmp2[tmp2$strata == paste0(strata, 
       "=", ref), 5:ncol(tmp2)]))
  dummy <- do.call("rbind", 
                  replicate(nlevels(tmp2$strata),
                            ref_props[, -ncol(ref_props)],
                            simplify=F))
  dummy <- cbind(id = 1:nrow(dummy), 
                  strata = factor(rep(levels(tmp2$strata),
                            each = nrow(dummy) / 
                            nlevels(tmp2$strata))),
                  dummy)
}

dummy <- dummy[, -ncol(dummy)]

# Formulate cox model
pred <- paste(names(tmp)[4:ncol(tmp)], collapse = "+")
cox_formula <- formula(paste0("Surv(time, status) ~ ", pred))
df_cox <- coxph(cox_formula, id = id, data = tmp)
sf_cox <- survfit(df_cox, newdata = dummy, se.fit = F)

# Prediction times
if(is.null(times)){
  xtime <- sort(unique(c(0, tmp$time)))
} else {
  xtime <- times
}

# Extract curves per substrata and average
curves <- vector("list",length(sf_cox$states))
for(i in 1:length(sf_cox$states)){
  curves_i <- vector("list", nlevels(tmp$strata))
  for(j in 1:nlevels(tmp$strata)){
    curves_i[[j]] <- numeric(length(xtime))
    for(k in 1:nrow(ref_props)){
      index <- k + (j - 1) * nrow(ref_props)
      sf_pred <- summary(sf_cox, times = xtime, extend = T)$pstate[, index, i]
      curves_i[[j]] <- curves_i[[j]] + ref_props[k, ncol(ref_props)] * sf_pred
    }
  }
  curves[[i]] <- do.call("c", curves_i)
}

# Combine states and construct output
curves_output <- do.call("cbind", curves)
colnames(curves_output) <- c("(s0)", status_levels)
output <- cbind(data.frame(time = rep(xtime, times = nlevels(tmp$strata))), 
                 strata = factor(rep(levels(tmp$strata)),
                                each = nrow(dummy) / 
                                nlevels(tmp$strata))))
```

strata_wt(): average of strata-estimated curves for cause-specific hazards

Unlike the previous function, which used a cause-specific Cox regression model, this function uses a cumulative incidence estimator from the `survfit` to construct cumulative incidence curves per strata. After cumulative incidence estimation, a risk set with every unique combination of covariates is created for group within a user-defined strata, and a weighted average is taken within each strata. For this method to optimally work, continuous covariates are discretized based on a user-defined number of quantiles.

```r
strata_wt <- function(formula, strata, ref = NULL, data = NULL, n_qntl = 5, times = NULL){
  # Extract components from formula object
  tmp <- model.frame(formula, data)
  status_levels <- levels(tmp[, 1])
  tmp <- cbind(data.frame(time = tmp[, 1][, 1],
                         status = factor(tmp[, 1][, 2]),
                         tmp[2:ncol(tmp)]
                         names(tmp)[which(names(tmp) == strata)] <- "strata"
  tmp$strata <- factor(paste0(strata, "=",
                              tmp$strata))
  tmp <- tmp[, c(which(colnames(tmp) %in% c("time", "status", "strata")),
                 which(!colnames(tmp) %in% c("time", "status", "strata")))]
  names(tmp) <- gsub(".*\$", "", names(tmp))

  # Discretize numeric covariates
  for(n in 4:ncol(tmp)){
    if(is.numeric(tmp[, n])) {
      tmp[, n] <- cut(tmp[, n],
      breaks = c(quantile(tmp[, n],
                   probs = seq(0, 1, by = 1/n_qntl)))},
```
labels = apply(cbind(seq(0, 100, 100/n_qntl)[-(n_qntl + 1)],
seq(0, 100, 100/n_qntl)[[-1]],
1, paste, collapse = "-"),
include.lowest = T)
}

# Reference weights
tmp$dummy <- rep(1, nrow(dummy))
if(is.null(ref)) {
  ref_covs <- factor(apply(tmp[, 4:ncol(tmp)], 1, paste, collapse = "|"))
  ref Props <- as.data.frame(table(ref_covs) / length(ref_covs))
} else {
  ref_covs <- factor(apply(tmp[tmp$strata == paste0(strata, ",", ref), 4:ncol(tmp)],
1, paste, collapse = "|"))
  ref Props <- as.data.frame(table(ref_covs) / length(ref_covs))
}

# Construct survfit model
pred <- paste(names(tmp)[3:(ncol(tmp))], collapse = "+")
sf formula <- formula(paste0("Surv(time, status) ~ " , pred))
sf <- survfit(sf formula, data = tmp)

# predicted timepoints
if(is.null(times)){
xtime <- sort(unique(c(0, tmp$time)))
} else {
xtime <- times
}

# Extract curves per substrata and average
curves <- vector("list", nlevels(tmp$strata))
strata freq <- table(strata)
for(j in 1:nlevels(tmp$strata)){
  # Correct indices and frequency table for missing strata
  obs_covs <- levels(factor(apply(tmp[tmp$strata == levels(tmp$strata)[j]],
4:ncol(tmp)), 1, paste, collapse = "|"))
  common ref <- unname(sapply(
    intersect(obs covs, ref Props$ref_covs),
    function(x) which(x == levels(ref Props$ref_covs))))
  common obs <- unname(sapply(
    intersect(obs covs, ref Props$ref_covs),
    function(x) which(x == obs covs))))
  ind <- grep(paste0("strata=" , levels(tmp$strata)[j]),
names(sf$strata))[[common obs]]
  ref Props k <- ref Props[common ref, ]
  ref Props k$Freq <- ref Props k$Freq / sum(ref Props k$Freq)

  # Average curves using strata-specific frequency table
curves[[j]] <- matrix(0, nrow = length(xtime), ncol = length(sf$states))
  for(k in 1:length(ind)){
    ind k <- ind[k]
    extract substrata <- summary(sf[ind k, ], times = xtime, extend = T)
    curves[[j]] <- curves[[j]] + ref Props k[k, 2] * extract substrata$pstate
  }
}

# Construct output
curves long <- do.call("rbind", curves)
output <- data.frame(time = rep(xtime, times = nlevels(tmp$strata)),
strata = factor(rep(levels(tmp$strata),
202x730]seq(0, 100, 100/n_qntl)[[-1]],
1, paste, collapse = "-"),
include.lowest = T)
ipw_finegray(): inverse probability weighting for subdistribution hazards

The subdistribution hazard model from the survival package makes use of a special finegray data format, which can be obtained using the finegray() function. This function produces the finegray data format with the incorporation of inverse probability weights in the Fine-Gray weights.

**formula**
- a formula object containing a Surv() object as the response, and the strata and covariates as the predictors.

**strata**
- a character string used to identify the strata within the formula object.

**data**
- an optional data.frame used to interpret the variables described by formula argument.

**ref**
- a character input to specify a reference group from x. If this argument is NULL, which is by default, the entire sample is used as a reference instead.

**etype**
- a character string used to identify the event of interest among the status conditions provided in the ‘Surv()’ object.

```r
ipw_finegray <- function(formula, data = NULL, strata, ref = NULL, etype) {
  # Model frame
  tmp <- model.frame(formula, data = data)
  etype <- as.character(which(levels(tmp[, 1]) == etype))
  tmp <- cbind(data.frame(id = factor(1:nrow(tmp)),
    time = tmp[, 1][, 1],
    status = factor(tmp[, 1][, 2]),
    tmp[2:ncol(tmp)])
  names(tmp)[which(names(tmp) == strata) <- "strata"
  tmp$strata <- factor(paste0(strata, ",", tmp$strata))
  tmp <- tmp[, c(which(colnames(tmp) %in% c("id", "time", "status", "strata")),
    which(!colnames(tmp) %in% c("id", "time", "status", "strata")))]
  names(tmp) <- gsub(".*\$", "", names(tmp))

  # Construct ipw weights
  ipw_pred <- paste(names(tmp)[5:ncol(tmp)], collapse = "\n")
  ipw_ref <- matrix(NA, nrow = nrow(tmp), ncol = nlevels(tmp$strata))
  for(i in 1:nlevels(tmp$strata)) {
    ipw_formula <- formula(paste0("I(strata==", "', "levels(tmp$strata)[i] , "', "'))", "", ipw_pred))
    ipw_ref[, i] <- 1 / predict(glm(ipw_formula, data = tmp, family = "binomial"),
      type = "response")
  }

  # adjust weights for reference group
  if(!is.null(ref)){
```
sub_wt(): average of individual predicted curves for subdistribution hazards

In this conditional approach, we duplicate the risk set, used to construct a subdistribution Cox regression model, to each group within the specified strata, and calculate the average of individual predicted cumulative incidence curves from the risk set. A reference group within the strata variable may be specified to only duplicate individuals from the risk set of this group to other groups.

**formula**
a formula object containing a Surv() object as the response, and the strata and covariates as the predictors.

**strata**
a character string used to identify the strata within the formula object.

**data**
an optional data.frame used to interpret the variables described by formula argument.

**ref**
a character input to specify a reference group from x. If this argument is NULL, which is by default, the entire sample is used as a reference instead.

**etype**
a character string used to identify the event of interest among the status conditions provided in the ‘Surv()’ object.

**times**
a series of user-defined time points for which the cumulative incidence should be computed per strata. If NULL, which is by default, the unique observation times from the data are used instead.

```r
sub_wt <- function(formula, strata, data = NULL, ref = NULL, etype, times = NULL)
{
  # Extract components from formula object
  tmp <- model.frame(formula, data)
  etype_val <- as.character(which(levels(tmp[, 1]) == etype))
  tmp <- cbind(data.frame(time = tmp[, 1][, 1],
                            status = factor(tmp[, 1][, 2]),
                            tmp[2:ncol(tmp)])
  names(tmp)[which(names(tmp) == strata)] <- "strata"
  ref <- paste0(strata, "=" , ref)
  ref_formula <- formula(paste0("I(strata==" , ref, ", ref, "=")
  ref_prob <- predict(glm(ref_formula, data = tmp, family = "binomial"),
    type = "response")
  ipw_ref <- apply(ipw_ref, 2, function(x) ref_prob * x)

  # Subdivide weights over obs
  ipw_wt <- numeric(nrow(tmp))
  for(j in 1:nrow(tmp)){
    ipw_wt[j] <- ipw_ref[j, which(tmp[, strata[j] == levels(tmp$strata))]
  }

  # Construct finegray data.frame and multiply ipw weights w/ fg weights
  fg <- finegray(Surv(time, status) ~ strata + id, data = tmp, etype = etype)
  fg$fgwt <- fg$fgwt * ipw_wt[fg$id]
  fg <- fg[, -which(names(fg) == "id")]
  return(fg)
}
```
TMP <- factor(paste0(strata, "=" , tmp$strata))
tmp <- tmp[, c(which(colnames(tmp) %in% c("time", "status", "strata")),
   which(!colnames(tmp) %in% c("time", "status", "strata")))]
names(tmp) <- gsub(".*\$", "", names(tmp))

# Construct Cox model
pred <- paste(names(tmp)[3:ncol(tmp)], collapse = "+")
sub_formula <- formula(paste0("Surv(time, status) ~ " pred))
sub_fg <- finegray(sub_formula, data = tmp, etype = etype_val)
cox_formula <- formula(paste0("Surv(fgstart, fgstop, fgstatus) ~ " pred))
sub_cox <- coxph(cox_formula, data = sub_fg, weight = fgwt)

# Construct dummy data w/ every observation for each cohort
if(is.null(ref)){
dummy <- do.call("rbind", replicate(nlevels(tmp$strata),
   tmp[, -3],
   simplify = F))
dummy <- cbind(strata = rep(levels(tmp$strata),
   each = nrow(tmp)),
   dummy)
} else {
dummy <- do.call("rbind",
   replicate(nlevels(tmp$strata),
   tmp[tmp$strata == paste0(strata, "=", ref), -3],
   simplify = F))
dummy <- cbind(strata = rep(levels(tmp$strata),
   each = nrow(tmp[tmp$strata == paste0(strata, "=", ref)])),
   dummy)
}
dummy <- dummy[, c(-2:-3)]
sf_cox <- survfit(sub_cox, newdata = dummy, se.fit = F)

# predicted times
if(is.null(times)){
   xtime <- sort(unique(c(0, tmp$time)))
} else {
   xtime <- times
}
sub_pred <- summary(sf_cox, times = xtime, extend = T)

# Take mean curve per strata for each state
surv_curve <- numeric(length(xtime) * nlevels(tmp$strata))
for(i in 1:nlevels(tmp$strata)){
   ind_sf <- 1:(nrow(dummy) / nlevels(tmp$strata)) +
   (i - 1) * (nrow(dummy) / nlevels(tmp$strata))
   ind_surv <- 1:length(xtime) * length(xtime) + (i - 1)
   surv_curve[ind_surv] <- rowMeans(1 - sub_pred$surv[, ind_sf])
}

# Construct output
output <- data.frame(time = rep(xtime, times = nlevels(tmp$strata)),
   strata = factor(rep(levels(tmp$strata),
   each = length(xtime))),
   etype = rep(etype, length(surv_curve)),
   cuminc = surv_curve)

return(output)
**substrat_wt()**: average of strata-predicted curves for subdistribution hazards

This function is a computation efficient version of the `subwt()` function, which only performs predictions on the each unique combination of covariates. This is done by creating a risk set with every unique combination of covariates, and take a weighted average over these curves. For this method to optimally work, continuous covariates are discretized based on a user-defined number of quantiles.

- **formula**: a `formula` object containing a `Surv()` object as the response, and the strata and covariates as the predictors.
- **strata**: a character string used to identify the strata within the `formula` object.
- **data**: an optional `data.frame` used to interpret the variables described by `formula` argument.
- **ref**: a character input to specify a reference group from x. If this argument is NULL, which is by default, the entire sample is used as a reference instead.
- **n_quantile**: an integer input used to define the number of of quantiles used to automatically discretize continuous covariates. The default value is set to 5.
- **etype**: a character string used to identify the event of interest among the status conditions provided in the `Surv()` object.
- **times**: a series of user-defined time points for which the cumulative incidence should be computed per strata. If NULL, which is by default, the unique observation times from the data are used instead.

```r
substrat_wt <- function(formula, strata, ref = NULL, data = NULL, n_qntl = 5, etype, times = NULL){
  # Extract components from formula object
  tmp <- model.frame(formula, data)
  etype_val <- as.character(which(levels(tmp[, 1]) == etype))
  tmp <- cbind(data.frame(time = tmp[, 1][, 1],
    status = factor(tmp[, 1][, 2]),
    tmp[2:ncol(tmp)])
  names(tmp)[which(names(tmp) == strata)] <- "strata"
  tmp$strata <- factor(paste0(strata, "=",
    tmp$strata))
  tmp <- tmp[, c(which(colnames(tmp) %in% c("time", "status", "strata")),
    which(!colnames(tmp) %in% c("time", "status", "strata")))]
  names(tmp) <- gsub(".*\\$", "", names(tmp))
  # Bin numeric covariates
  for(n in 4:ncol(tmp)){
    if(is.numeric(tmp[, n])){
      if(tmp[, n] <= cut(tmp[, n],
        breaks = c(quantile(tmp[, n],
          probs = seq(0, 1, by = 1/n_qntl))),
        labels = apply(cbind(seq(0, 100, 100/n_qntl)[-(n_qntl + 1)],
          seq(0, 100, 100/n_qntl)[-1]),
        1, paste, collapse = "-"),
        include.lowest = T)
    }
  }
}
```
# Construct Cox model
pred <- paste(names(tmp)[3:ncol(tmp)], collapse = "+")
sub_formula <- formula(paste0("Surv(time, status) \sim ", pred))
sub_fg <- finegray(sub_formula, data = tmp, etype = etype_val)
cox_formula <- formula(paste0("Surv(fgstart, fgstop, fgstatus) \sim ", pred))
sub_cox <- coxph(cox_formula, data = sub_fg, weight = fgwt)

# Reference weights
tmp2 <- tmp
tmp2$dummy <- factor(rep(1, nrow(tmp)))
if(is.null(ref)){
    ref_props <- as.data.frame(table(tmp2[, 4:ncol(tmp2)]) / nrow(tmp2[, 4:ncol(tmp2)]))
dummy <- do.call("rbind",
    replicate(nlevels(tmp2$strata),
        ref_props[, -ncol(ref_props)],
        simplify=F))
dummy <- cbind(strata = factor(rep(levels(tmp2$strata), each = nrow(dummy) / nlevels(tmp2$strata)))))
} else {
    ref_props <- as.data.frame(table(tmp2[tmp2$strata == paste0(strata, "=", ref), 4:ncol(tmp2)]) / nrow(tmp2[tmp2$strata == paste0(strata, "=", ref), 4:ncol(tmp2)]))
dummy <- do.call("rbind",
    replicate(nlevels(tmp2$strata),
        ref_props[, -ncol(ref_props)],
        simplify=F))
dummy <- cbind(strata = factor(rep(levels(tmp2$strata), each = nrow(dummy) / nlevels(tmp2$strata)))))
}
dummy <- dummy[, -ncol(dummy)]
sf_cox <- survfit(sub_cox, newdata = dummy, se.fit = F)

# Prediction times
if(is.null(times)){
    xtime <- sort(unique(c(0, tmp$time)))
} else {
    xtime <- times
}
sub_pred <- summary(sf_cox, times = xtime, extend = T)

# Extract curves per substrata and average
curves_i <- vector("list", nlevels(tmp$strata))
for(j in 1:nlevels(tmp$strata)){
curves_i[[j]] <- numeric(length(xtime))
for(k in 1:nrow(ref_props)){
    index <- k + (j - 1) * nrow(ref_props)
    sf_pred <- 1 - sub_pred$surv[, index]
    curves_i[[j]] <- curves_i[[j]] + ref_props[k, ncol(ref_props)] * sf_pred
}
}
surv_curve <- do.call("c", curves_i)

# Combine states and construct output
output <- data.frame(time = rep(xtime, times = nlevels(tmp$strata)),
    strata = factor(rep(levels(tmp$strata), each = length(xtime))),
    etype = rep(etype, length(surv_curve)))
A.5 True curve computation

true_comp(): hypothetical true curve computation for cause-specific hazards

This function was developed to calculate the hypothetical true curve for a series of time points, based on the cause-specific hazard model. The hypothetical true curve can be controlled by a list of baseline hazard, and treatment covariate effect coefficients per event.

time a vector of time points.
k the number of events.
base_haz a vector of baseline hazards for each event.
beta_coef a list of vectors, each vector containing the covariate coefficients for each event. The first element affects selection of the discrete covariate, and the second element affects the selection of the continuous covariate.

```r
true_comp <- function(time, k, base_haz, beta_coef){
  # Create combinations discrete covariates
  X_beta <- lapply(beta_coef, 
                  function(b) as.matrix(expand.grid(c(0, 1), c(0, 1, 2), 0)) %*% b)
  # Construct curve per discrete combination per event k
  curves_Ik <- vector("list", k)
  for(i in 1:k){
    curves_ij <- matrix(NA, nrow = length(time), ncol = 6)
    for(j in 1:6){
      lambda_k <- function(x) {
        base_haz[i] * exp(X_beta[[i]][j] + x * beta_coef[[i]][3])
      }
      sum_k <- paste(paste0("base_haz[", 1:k, "], exp(X_beta[[", 1:k, "]][j], x, "beta_coef[[", 1:k, "][3]]"), collapse = "+")
      lambda_K <- function(x) {
        eval(parse(text = sum_k))
      }
      for(l in 1:length(time)){
        surv_fun <- function(x) {
          lambda_k(x) / lambda_K(x) * (1 - exp(-time[l] * lambda_K(x))) * dnorm(x)
        }
        curves_ij[l, j] <- integrate(surv_fun, -10, 10)$value
      }
      curves_Ik[[i]] <- c(rowMeans(curves_ij[, c(1, 3, 5)]), 
                          rowMeans(curves_ij[, c(2, 4, 6)]))
    }
  }
  # Construct output
  pstate <- do.call("cbind", curves_Ik)
  pstate <- cbind(1 - rowSums(pstate), pstate)
  colnames(pstate) <- c("(s0)", 1:k)
}
```

73
true_sh(): hypothetical true curve computation for subdistribution hazards

This function was developed to calculate the hypothetical true curve for a series of time points, based on the subdistribution hazard model. The hypothetical true curve can be controlled by a list of baseline hazard, and treatment covariate effect coefficients per event.

time a vector of time points.

beta_coef a list of length two where each listed element is a vector of covariate coefficients for each event. The first element in the vector is the effect size of the cohort, the second element the effect of the discrete covariate, and the third element the effect of the continuous covariate.

p the baseline probability of occurrence for the first event.

true_sh <- function(time, beta_coef, p){
    # Create combinations discrete covariates
    X_beta <- as.matrix(expand.grid(c(0, 1), c(0, 1, 2), 0)) %*% beta_coef

    # Construct curve per discrete combination per event k
    curves_ij <- matrix(NA, nrow = length(time), ncol = 6)
    for(j in 1:6){
        lambda_k <- function(x) {exp(X_beta[j] + x * beta_coef[3])}
        for(l in 1:length(time)){
            surv_fun <- function(x) {
                (1 - (1 - p * (1 - exp(-time[l])))^(lambda_k(x))) * dnorm(x)
            }
            curves_ij[l, j] <- integrate(surv_fun, -10, 10)$value
        }
    }
    cuminc <- c(rowMeans(curves_ij[, c(1, 3, 5)]), rowMeans(curves_ij[, c(2, 4, 6)]))

    # Construct output
    output <- cbind(data.frame(time = rep(time, times = 2),
                               cohort = rep(c("control", "treat"), each = length(time))),
                               cuminc)
    return(output)
}
A.6 Validation of adjustment methods

To evaluate the accuracy and reproducibility of the adjustment methods, each method was applied on a biased sample in an attempt to approximate hypothetical the true curve. Therefore, a total of $B = 1000$ replications of population simulation, biased sampling, and covariate adjustment were performed, and the deviation from the hypothetical true curve was calculated and stored. This was done separately for the cause-specific hazard and subdistribution hazard model, as shown below.

Algorithm to evaluate adjustment methods for cause-specific hazards

The algorithm first computes the hypothetical true cumulative incidence curve for the cause-specific hazard model, then proceeds to generate $N = 10000$ observations under the cause-specific hazard model. A biased sample of $n = 1000$ observation per cohort is produced independently by each of the sampling schemes discussed in the section 2.2, and each adjustment method described in section 2.3 is applied on the biased sample. Lastly, the true curve is subtracted from each covariate-adjusted sample, and the results are restored in a list. Lastly, all replication are combined into a single output per biased sampling scheme, and saved as an Excel .xlsx file.

```r
# Start simulation
set.seed(1546732)
# Generate true curve for user-specified timeframe
times <- seq(0, 0.6, 0.02)
true_I <- true_comp(time = times, k = 2, base_haz = c(1, 1),
                    beta_coef = list(c(-0.5, 2, 0.5), c(0.3, 1, -0.4)))

# Start loop
B = 1000
deltas <- vector("list", B)
deltas2 <- vector("list", B)
for(i in 1:B){
  for(j in 1:2) {
    # Simulate data and sample (j = 1 treatment bias, j = 2 opposing bias)
    df <- sim_comp(N = 10000, k = 2, base_haz = c(1, 1),
                   beta_coef = list(c(-0.5, 2, 0.5), c(0.3, 1, -0.4)))
    if(j == 1) {
      df_sample <- sample_cov(df, n = 1000, beta_coef = c(2, 2), bias = "treat"
                              ref_input = "control"
    } else {
      df_sample <- sample_cov(df, n = 1000, beta_coef = c(1, 1), bias = "both"
                              ref_input = NULL
    }
    # Survfit objects and weights
    df_sf <- survfit(Surv(obs_time, status) ~ cohort,
                     data = df)
    sample_sf <- survfit(Surv(obs_time, status) ~ cohort,
                          data = df_sample)
    df_sample_ipw <- est_ipw(cohort ~ age * mutation,
                             data = df_sample, ref = ref_input)
    sample_sf_ipw <- survfit(Surv(obs_time, status) ~ cohort,
                             data = df_sample, ref = ref_input)
  }
}
```

75
data = df_sample, weights = ipw)

df_sample$estwt <- est_wt(df_sample[, 3:5], df[, 3:5])
sample_sf_estwt <- survfit(Surv(obs_time, status) ~ cohort, 
data = df_sample, weights = estwt)

# Construct cumulative incidence curves
df_I <- summary(df_sf, times = times, extend = T)$pstate
sample_I <- summary(sample_sf, times = times, extend = T)$pstate
ipw_I <- summary(sample_sf_ipw, times = times, extend = T)$pstate
estwt_I <- summary(sample_sf_estwt, times = times, extend = T)$pstate
rr_I <- rr_wt(Surv(obs_time, status) ~ cohort + mutation + age, 
strata = "cohort", ref = ref_input, 
data = df_sample, times = times)[, 3:5]
coxstrat_I <- coxstrat_wt(Surv(obs_time, status) ~ cohort + mutation + age, 
strata = "cohort", ref = ref_input, 
data = df_sample, times = times)[, 3:5]
strata_I <- strata_wt(Surv(obs_time, status) ~ cohort + mutation + age, 
strata = "cohort", ref = ref_input, 
data = df_sample, times = times)[, 3:5]

# Calculate average difference w/ true curve
tmp <- unname(cbind(rep(i, 14 * length(times)), 
rep(times, times = 14), 
rep(c("population", "sample", "sample_ipw", 
"sample_estwt", "sample_rr", 
"sample_coxstrat", "sample_sfstrat"), 
each = 2 * length(times)), 
rep(c("control", "treat"), 
each = length(times), times = 7), 
rbind(df_I - true_I[, 3:5], 
sample_I - true_I[, 3:5], 
ipw_I - true_I[, 3:5], 
estwt_I - true_I[, 3:5], 
rr_I - true_I[, 3:5], 
coxstrat_I - true_I[, 3:5], 
strata_I - true_I[, 3:5])))

names(tmp) <- c("iteration", "time", "method", "cohort", 
"event_free", "event_1", "event_2")
if(j == 1) deltas[[i]] <- tmp else deltas2[[i]] <- tmp
)

# Save the average differences
timebias_output <- do.call("rbind", deltas)
timebias_output2 <- do.call("rbind", deltas2)
csh_timebias <- list("treat_bias" = gather(timebias_output, event, delta, 5:7), 
"opp_bias" = gather(timebias_output2, event, delta, 5:7))
write.xlsx(csh_timebias, "csh_t00_t06.xlsx")
Algorithm to evaluate adjustment methods for subdistribution hazards

Similar to the procedure for the cause-specific hazards algorithm, the subdistribution hazards algorithm first computes the hypothetical true cumulative incidence curve under the subdistribution hazard model, then proceeds to generate $N = 10000$ observations under the subdistribution hazard model. A biased sample of $n = 1000$ observation per cohort is produced independently by each of the sampling schemes discussed in the section 2.2, and each adjustment method, described in section 2.3 is applied on the biased sample. Lastly, the true curve is subtracted from each covariate-adjusted sample, and the results are restored in a list. Lastly, all replication are combined into a single output per biased sampling scheme, and saved as an Excel .xlsx file. Unfortunately, the subdistribution hazard model was prone to a rare convergence error which could terminate the loop. To prevent this, alterations were made to proceed the loop upon encountering an error.

```
# Start simulation
set.seed(1546732)

# Generate true curve
times <- seq(0, 1.4, 0.02)
true_I <- true_sh(time = times, beta_coef = c(-0.5, 2, 0.5), p = 0.7)

# Start loop
B = 1000
deltas <- vector("list", B)
deltas2 <- vector("list", B)
for(i in 1:B){
  for(j in 1:2) {
    # Repeat iteration if erroneous fit
    error_message <- try(silent = TRUE, expr = {
      # Simulate data and sample (j = 1 treatment bias, j = 2 opposing bias)
      df <- sim_sh(N = 10000, rates = c(1, 1), shapes = c(1, 1),
                   beta_coef = list(c(-0.5, 2, 0.5), c(0.3, 1, -0.4)), p = 0.7)
      if(j == 1) {
        df_sample <- sample_cov(df, n = 1000, beta_coef = c(2, 2), bias = "treat")
        ref_input <- "control"
      } else {
        df_sample <- sample_cov(df, n = 1000, beta_coef = c(1, 1), bias = "both")
        ref_input = NULL
      }
    })
    # Finegray data.frames
    df_fg <- finegray(Surv(time, status) ~ cohort,
                      data = df, etype = "1")
    sample_fg <- finegray(Surv(time, status) ~ cohort,
                           data = df_sample, etype = "1")
    sample_fgipw <- ipw_finegray(Surv(time, status) ~ cohort + mutation + age,
                                  data = df_sample, strata = "cohort",
                                  ref = ref_input, etype = "1")
  }
  # Survfit objects and corrected output
  df_sf <- survfit(Surv(fgstart, fgstop, fgstatus) ~ cohort,
                   data = df_fg, weights = fgwt)
```
sample_sf <- survfit(Surv(fgstart, fgstop, fgstatus) ~ cohort, 
               data = sample_fg, weights = fgwt)
sample_sfipw <- survfit(Surv(fgstart, fgstop, fgstatus) ~ strata, 
                data = sample_fgipw, weights = fgwt)
sample_subwt <- sub_wt(Surv(time, status) ~ cohort + age + mutation, 
                data = df_sample, strata = "cohort", 
                ref = ref_input, etype = "1", times = times)
sample_substrat <- substrat_wt(Surv(time, status) ~ cohort + age + mutation, 
                data = df_sample, strata = "cohort", 
                ref = ref_input, etype = "1", times = times)

# Extract incidence curves
df_I <- 1 - summary(df_sf, times = times, extend = T)$surv
sample_I <- 1 - summary(sample_sf, times = times, extend = T)$surv
ipw_I <- 1 - summary(sample_sfipw, times = times, extend = T)$surv
subwt_I <- sample_subwt$cuminc
substrat_I <- sample_substrat$cuminc

# Calculate average difference w/ true curve
tmp <- data.frame(iteration = rep(i, 10 * length(times)),
       time = rep(times, times = 10),
       method = rep(c("population", "sample", "sample_ipw", 
                     "sample_subwt", "sample_substrat"), 
                     each = 2 * length(times)),
       cohort = rep(c("control", "treat"), 
                     each = length(times), times = 5),
       delta = c(df_I - true_I$cuminc, 
                sample_I - true_I$cuminc, 
                ipw_I - true_I$cuminc, 
                subwt_I - true_I$cuminc, 
                substrat_I - true_I$cuminc))
if(j == 1) deltas[[i]] <- tmp else deltas2[[i]] <- tmp
}
if(!is(error_message, "try-error")) break
}

# Denote iteration
print(paste("iteration", i, "of", B))

# Save the average differences
sdh_timebias <- list("treat_bias" = do.call("rbind", deltas),
                     "opp_bias" = do.call("rbind", deltas2))
write.xlsx(sdh_timebias, "sdh_t00_t14.xlsx")
Adjustment accuracy for cause-specific hazard methods over a time series

A plot was produced for each biased sampling scheme to indicate the adjustment accuracy and reproducibility of the adjustment methods for cause-specific hazards over a series of time points [0, 0.6].

```r
# 1) True curve deviation after adjustment of opposing bias between cohorts
# Loading the sheet
timebias_gather <- read.xlsx("csh_t00_t06.xlsx", sheet = 2)

# Estimate of mean and standard deviation
csh_timebias <- aggregate(delta ~ time + method + cohort + event, data = timebias_gather, FUN = mean)
csh_timebias$sd <- aggregate(delta ~ time + method + cohort + event, data = timebias_gather, FUN = sd)$delta

# Adjust group names
csh_timebias$method <- factor(csh_timebias$method, levels = c("population", "sample", "sample_ipw", "sample_estwt", "sample_rr", "sample_coxstrat", "sample_sfstrat"))
levels(csh_timebias$method) <- c("Target

population", "Sample

opposing bias", "Weights

inverse_prob", "Weights

external_ref", "Average_CI

riskReg_pred", "Average_CI

coxph_strata", "Average_CI

survfit_strata")
csh_timebias$event <- factor(csh_timebias$event, levels = c("event_1", "event_2", "event_free"))
levels(csh_timebias$event) <- c("Event

1", "Event

2", "Event

free")
csh_timebias$cohort <- ifelse(csh_timebias$cohort == "control", "Control", "Treatment")

# Produce plot
p1 <- ggplot(csh_timebias, aes(x = time, y = delta, col = cohort)) + facet_grid(event ~ method) + ylim(-0.2, 0.2) +
  scale_x_continuous(breaks = c(0, 0.2, 0.4, 0.6), limits = c(-0.05, 0.65)) +
  geom_hline(yintercept = 0, lwd = 1, col = 1, alpha = 0.3) +
  geom_line(lwd = 1) + geom_ribbon(aes(ymin = delta - sd, ymax = delta + sd), alpha = 0.3) +
  theme_bw() + theme(text = element_text(size = 16, face = "bold"),
    axis.title.x = element_text(size = 16),
    axis.title.y = element_text(size = 16),
    legend.position = "top") +
  ylab("Deviation from true CI") + xlab("Time") + labs(col = "Cohort")
```

# 2) True curve deviation after adjustment of treatment bias between cohorts
# Loading the sheet
timebias_gather2 <- read.xlsx("csh_t00_t06.xlsx", sheet = 1)

# Estimate of mean and standard deviation
csh_timebias2 <- aggregate(delta ~ time + method + cohort + event,)
data = timebias_gather2, FUN = mean)
csh_timebias2$sd <- aggregate(delta ~ time + method + cohort + event,
data = timebias_gather2,
FUN = sd)$delta

# Adjust group names

csh_timebias2$method <- factor(csh_timebias2$method,
levels = c("population", "sample", "sample_ipw",
"sample_estwt", "sample_rr",
"sample_coxstrat", "sample_sfstrat"))
levels(csh_timebias2$method) <- c("Target\_n\_population",
"Sample\_n\_treatment\_bias",
"Weights\_n\_inverse\_prob",
"Weights\_n\_external\_ref",
"Average\_CI\_n\_riskReg\_pred",
"Average\_CI\_n\_coxph\_strata",
"Average\_CI\_n\_survfit\_strata")

csh_timebias2$event <- factor(csh_timebias2$event,
levels = c("event_1",
"event_2",
"event_free"))
levels(csh_timebias2$event) <- c("Event\_1", "Event\_2", "Event\_free")
csh_timebias2$cohort <- ifelse(csh_timebias2$cohort == "control",
"Control",
"Treatment")

p2 <- ggplot(csh_timebias2, aes(x = time, y = delta, col = cohort)) +
  facet_grid(event ~ method) + ylim(-0.2, 0.2) +
  scale_x_continuous(breaks = c(0, 0.2, 0.4, 0.6),
  limits = c(-0.05, 0.65)) +
  geom_hline(yintercept = 0, lwd = 1, col = 1, alpha = 0.3) +
  geom_line(lwd = 1) + geom_ribbon(aes(ymin = delta - sd,
  ymax = delta + sd), alpha = 0.3) +
  theme_bw() + theme(text = element_text(size = 16, face = "bold"),
  axis.title.x = element_text(size = 16),
  axis.title.y = element_text(size = 16),
  legend.position = "top") +
  ylab("Deviation\_from\_true\_CI\_n") + xlab("Time") + labs(col = "Cohort:")
Adjustment accuracy for subdistribution hazard methods over a time series

A plot was produced for each biased sampling scheme to indicate the adjustment accuracy and reproducibility of the adjustment methods for subdistribution hazards over a series of time points \([0, 1.4]\).

### 1) True curve deviation after adjustment of opposing bias between cohorts

#### Loading the sheet
```r
timebias_gather <- read.xlsx("sdh_t00_t14.xlsx", sheet = 2)
```

#### Estimate mean and standard deviation
```r
sdh_timebias <- aggregate(delta ~ time + method + cohort, data = timebias_gather, FUN = mean)
sdh_timebias$sd <- aggregate(delta ~ time + method + cohort, data = timebias_gather, FUN = sd)$delta
```

#### Adjust group names
```r
sdh_timebias$method <- factor(sdh_timebias$method, levels = c("population", "sample", "sample_ipw", "sample_subst", "sample substrat"))
levels(sdh_timebias$method) <- c("Target population", "Sample opposing bias", "Weights inverse prob", "Average CI Coxph pred", "Average CI Coxph strata")
```

#### Produce plot
```r
sdh_timebias$cohort <- ifelse(sdh_timebias$cohort == "control", "Control", "Treatment")
sdh_timebias$event <- rep("Event1", nrow(sdh_timebias))
```

#### Produce plot
```r
p1 <- ggplot(sdh_timebias2, aes(x = time, y = delta, col = cohort)) +
  facet_grid(event ~ method) + ylim(-0.25, 0.25) +
  scale_x_continuous(breaks = c(0, 0.5, 1, 1.5),
                     limits = c(-0.05, 1.55)) +
  geom_hline(yintercept = 0, lwd = 1, col = 1, alpha = 0.3) +
  geom_line(lwd = 1) + geom_ribbon(aes(ymin = delta - sd, ymax = delta + sd), alpha = 0.3) +
  theme_bw() + theme(text = element_text(size = 16, face = "bold"),
                    axis.title.x = element_text(size = 16),
                    axis.title.y = element_text(size = 16),
                    legend.position = "top") +
  ylab("Deviation from true CI") + xlab("Time") + labs(col = "Cohort:")
```

### 2) True curve deviation after adjustment of treatment bias between cohorts

#### Loading the sheet
```r
timebias_gather2 <- read.xlsx("sdh_t00_t14.xlsx", sheet = 1)
```

#### Estimate mean and standard deviation
```r
sdh_timebias2 <- aggregate(delta ~ time + method + cohort, data = timebias_gather2, FUN = mean)
sdh_timebias2$sd <- aggregate(delta ~ time + method + cohort, data = timebias_gather2, FUN = sd)$delta
```

#### Adjust group names
```r
sdh_timebias2$method <- factor(sdh_timebias2$method,
                                 levels = c("population", "sample", "sample_ipw", "sample_subst", "sample substrat"))
```
levels = c("population", "sample", "sample.ipw", "sample.subwt", "sample.substrat")

levels(sdh_timebias2$method) <- c("Target \n \n population", "Sample \n \n treatment.bias", "Weights \n \n inverse.prob", "Average \n \n CI \n \n coxph.predict", "Average \n \n CI \n \n coxph.strata")

sdh_timebias2$cohort <- ifelse(sdh_timebias2$cohort == "control", "Control", "Treatment")

sdh_timebias2$event <- rep("Event \n \n 1", nrow(sdh_timebias2))

# Produce plot
p2 <- ggplot(sdh_timebias2, aes(x = time, y = delta, col = cohort)) + facet_grid(event ~ method) + ylim(-0.25, 0.25) + scale_x_continuous(breaks = c(0, 0.5, 1, 1.5), limits = c(-0.05, 1.55)) + geom_hline(yintercept = 0, lwd = 1, col = 1, alpha = 0.3) + geom_line(lwd = 1) + geom_ribbon(aes(ymin = delta - sd, ymax = delta + sd), alpha = 0.3) + theme_bw() + theme(text = element_text(size = 16, face = "bold"), axis.title.x = element_text(size = 16), axis.title.y = element_text(size = 16), legend.position = "top") + ylab("Deviation \n \n from \n \n true \n \n CI \n \n \n \n") + xlab("Time") + labs(col = "Cohort:")

82
Adjustment accuracy of cause-specific hazard methods at $t = 0.3$

The adjustment accuracy for the adjustment methods on the cause-specific hazard models was more closely observed at a center point in the range of event times ($t = 0.3$). Furthermore, tables were produced with the bias and RMSE values for each adjustment method, as described in section 2.4.

# 1) True curve deviation at $t = 0.3$ following adjustment of opposing bias

```r
# Loading the sheet
delta_gather <- read.xlsx("csh_t00_t06.xlsx", sheet = 2)
delta_gather <- delta_gather[delta_gather$time == 0.3, ]

# Adjust group names
delta_gather$method <- factor(delta_gather$method, levels = c("population", "sample", "sample_ipw", "sample_estwt", "sample_rr", "sample_coxstrat", "sample_sfstrat"))
levels(delta_gather$method) <- c("Target_population", "Sample_opposing_bias", "Weights_inverse_prob", "Weights_external_ref", "Average_CI_riskReg_pred", "Average_CI_coxph_strata", "Average_CI_survfit_strata")
delta_gather$event <- factor(delta_gather$event, levels = c("event_1", "event_2", "event_free"))
levels(delta_gather$event) <- c("Event_1", "Event_2", "Event_free")
delta_gather$cohort <- ifelse(delta_gather$cohort == "control", "Control", "Treatment")

# Produce plot
p1 <- ggplot(delta_gather, aes(y = delta, x = cohort, col = cohort)) + facet_grid(event ~ method) + ylim(-0.25, 0.25) + geom_hline(yintercept = 0, lwd = 1, col = 1, alpha = 0.3) + geom_violin(position = position_dodge(1)) + stat_summary(fun.data = mean_sdl, geom = "pointrange", position = position_dodge(1)) + theme_bw() + theme(text = element_text(size = 16, face = "bold"), axis.title.x = element_blank(), axis.text.x = element_blank(), axis.ticks.x = element_blank(), axis.title.y = element_text(size = 16), legend.position = "top") + ylab("Deviation from true CI (t = 0.3)") + labs(col = "Cohort: ")
```

# 2) True curve deviation at $t = 0.3$ following adjustment of treatment bias

```r
# Loading the sheet
delta_gather2 <- read.xlsx("csh_t00_t06.xlsx", sheet = 1)
delta_gather2 <- delta_gather2[delta_gather2$time == 0.3, ]

# Adjust group names
delta_gather2$method <- factor(delta_gather2$method, levels = c("population", "sample", "sample_ipw", "sample_estwt", "sample_rr", "sample_coxstrat", "sample_sfstrat"))
levels(delta_gather2$method) <- c("Target_population", "Sample_opposing_bias", "Weights_inverse_prob", "Weights_external_ref", "Average_CI_riskReg_pred", "Average_CI_coxph_strata", "Average_CI_survfit_strata")
delta_gather2$event <- factor(delta_gather2$event, levels = c("event_1", "event_2", "event_free"))
levels(delta_gather2$event) <- c("Event_1", "Event_2", "Event_free")
delta_gather2$cohort <- ifelse(delta_gather2$cohort == "control", "Control", "Treatment")

# Produce plot
p2 <- ggplot(delta_gather2, aes(y = delta, x = cohort, col = cohort)) + facet_grid(event ~ method) + ylim(-0.25, 0.25) + geom_hline(yintercept = 0, lwd = 1, col = 1, alpha = 0.3) + geom_violin(position = position_dodge(1)) + stat_summary(fun.data = mean_sdl, geom = "pointrange", position = position_dodge(1)) + theme_bw() + theme(text = element_text(size = 16, face = "bold"), axis.title.x = element_blank(), axis.text.x = element_blank(), axis.ticks.x = element_blank(), axis.title.y = element_text(size = 16), legend.position = "top") + ylab("Deviation from true CI (t = 0.3)") + labs(col = "Cohort: ")
```
levels = c("population", "sample", "sample.ipw", "sample.estwt", "sample.rr", "sample.coxstrat", "sample.sfstrat")
levels(delta_gather2$method) <- c("Target\n\n\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntarget\n\ntar...
# 4) Bias and RMSE at t = 0.3 following adjustment of treatment bias

## Loading the sheet

delta_gather2 <- read.xlsx("csh_t00_t06.xlsx", sheet = 1)
delta_gather2 <- delta_gather2[delta_gather2$\time == 0.3,]

## Estimate bias and RMSE

delta_agg2 <- aggregate(delta ~ cohort + event + method, data = delta_gather2, FUN = mean)
delta_agg2$\text{mse} <- sqrt(aggregate(delta ~ cohort + event + method, data = delta_gather2, FUN = function(x) mean(x^2)))

## Order adjustment methods

delta_agg2[, 1:2] <- lapply(delta_agg2[, 1:2], factor)
delta_agg2$\text{method} <- factor(delta_agg2$\text{method}, levels = c("population", "sample", "sample\_ipw", "sample\_estwt", "sample\_rr", "sample\_coxstrat", "sample\_sfstrat"))

## Combine RMSE and bias columns

delta_comb <- rbind(delta_agg2[, 1:3], delta_agg2[, 1:3])
delta_comb$\text{measure} <- rep(c("Bias", "MSE"), each = nrow(delta_comb))
delta_comb$\text{value} <- round(c(delta_agg2$\delta, delta_agg2$\text{mse}), digits = 5)

## Produce table

tab2 <- tabular(Heading() * RowFactor(method, spacing = 1, levelnames = c("Target\_population", "Sample\_treatment\_bias", "Wts\_inverse\_prob", "Wts\_external\_ref", "Avg\_CI\_riskReg\_obs", "Avg\_CI\_coxph\_strata", "Avg\_CI\_survfit\_strata")) *

Heading() * RowFactor(cohort, levelnames = c("Control", "Treatment")) -

Heading() * Factor(event, levelnames = c("Event\_1", "Event\_2", "Event\_free")) *

Heading() * Factor(measure, levelnames = c("Bias", "MSE")) *

Heading() *(value)* Heading() *(mean), data = delta_comb)

toLatex(tab2)
Adjustment accuracy of subdistribution hazard methods at $t = 0.7$

The adjustment accuracy for the subdistribution hazards adjustment methods was more closely observed at a center point in the range of event times ($t = 0.7$). Furthermore, tables were produced with the bias and RMSE values for each adjustment method, as described in section 2.4.

```r
# 1) True curve deviation at $t = 0.3$ following adjustment of opposing bias
delta_gather <- read.xlsx("sdh_t00_t14.xlsx", sheet = 2)
delta_gather <- delta_gather[delta_gather$'time' == 0.7, ]

delta_gather$method <- factor(delta_gather$'method',
levels = c("population", "sample", "sample_ipw",
"sample_subwt", "sample_substrat"))
levels(delta_gather$'method') <- c("Target \ population",
"Sample \ opposing \ bias",
"Weights \ inverse \ prob",
"Average \ CI \ coxph \ pred",
"Average \ CI \ coxph \ strata")
delta_gather$cohort <- ifelse(delta_gather$'cohort' == "control",
"Control",
"Treatment")
delta_gather$event <- rep("Event 1", nrow(delta_gather))

delta_gather$cohort <- factor(delta_gather$cohort,
levels = c("control", "Treatment"))

delta_gather$event <- ifelse(delta_gather$event == "Event 2",
"Lower \ bound",
"Upper \ bound")

# Produce plot
p1 <- ggplot(delta_gather, aes(y = delta, x = cohort, col = cohort)) +
facet_grid(event ~ method) + ylim(-0.25, 0.25) +
geom_hline(yintercept = 0, lwd = 1, col = 1, alpha = 0.3) +
geom_violin(position = position_dodge(1)) +
stat_summary(fun.data = mean_sdl, geom = "pointrange",
position = position_dodge(1)) +
theme_bw() + theme(text = element_text(size = 16, face = "bold"),
axis.title.x=element_blank(),
axis.text.x = element_blank(),
axis.ticks.x = element_blank(),
axis.title.y = element_text(size = 16),
legend.position = "top") +
ylab("Deviation from true CI (t = 0.7)") +
labs(col = "Cohort: ")

# 2) True curve deviation at $t = 0.7$ following adjustment of treatment bias

# Loading the sheet
delta_gather2 <- read.xlsx("sdh_t00_t14.xlsx", sheet = 1)
delta_gather2 <- delta_gather2[delta_gather2$'time' == 0.7, ]

delta_gather2$method <- factor(delta_gather2$'method',
levels = c("population", "sample", "sample_ipw",
"sample_subwt", "sample_substrat"))
levels(delta_gather2$'method') <- c("Target \ population",
"Sample \ treatment \ bias",
"Weights \ inverse \ prob",
"Average \ CI \ coxph \ pred",
"Average \ CI \ coxph \ strata")

delta_gather2$cohort <- ifelse(delta_gather2$'cohort' == "control",
"Control",
"Treatment")

delta_gather2$event <- rep("Event 1", nrow(delta_gather2))

delta_gather2$cohort <- factor(delta_gather2$'cohort',
levels = c("control", "Treatment"))

delta_gather2$event <- ifelse(delta_gather2$'event' == "Event 2",
"Lower \ bound",
"Upper \ bound")

# Produce plot
p2 <- ggplot(delta_gather2, aes(y = delta, x = cohort, col = cohort)) +
facet_grid(event ~ method) + ylim(-0.25, 0.25) +
geom_hline(yintercept = 0, lwd = 1, col = 1, alpha = 0.3) +
geom_violin(position = position_dodge(1)) +
stat_summary(fun.data = mean_sdl, geom = "pointrange",
position = position_dodge(1)) +
theme_bw() + theme(text = element_text(size = 16, face = "bold"),
axis.title.x=element_blank(),
axis.text.x = element_blank(),
axis.ticks.x = element_blank(),
axis.title.y = element_text(size = 16),
legend.position = "top") +
ylab("Deviation from true CI (t = 0.7)") +
```
delta_gather2$cohort <- ifelse(delta_gather2$cohort == "control", "Control", "Treatment")
delta_gather2$event <- rep("Event\_1", nrow(delta_gather2))

# Produce plot
p2 <- ggplot(delta_gather2, aes(y = delta, x = cohort, col = cohort)) +
  facet_grid(event ~ method) + ylim(-0.25, 0.25) +
  geom_hline(yintercept = 0, lwd = 1, col = 1, alpha = 0.3) +
  geom_violin(position = position_dodge(1)) +
  stat_summary(fun.data = mean_sdl, geom = "pointrange",
               position = position_dodge(1)) +
  theme_bw() + theme(text = element_text(size = 16, face = "bold"),
                     axis.title.x = element_blank(),
                     axis.text.x = element_blank(),
                     axis.ticks.x = element_blank(),
                     axis.title.y = element_text(size = 16),
                     legend.position = "top") +
  ylab("Deviation\_from\_true\_CI\_\(t_\theta=0.7\)\_\(\mu\)\") + labs(col = "Cohort:\_\(\mu\)")

# 3) Bias and RMSE at t = 0.7 following adjustment of opposing bias

# Loading the sheet
delta_gather <- read.xlsx("sdh\_t00\_t14.xlsx", sheet = 2)
delta_gather <- delta_gather[delta_gather$"time == 0.7", ]

# Estimate bias and RMSE
delta_agg <- aggregate(delta ~ cohort + method,
                      data = delta_gather, FUN = mean)
delta_agg$mse <- sqrt(aggregate(delta ~ cohort + method,
                                 data = delta_gather,
                                 FUN = function(x) mean(x^2))$delta)

# Order adjustment methods
delta_agg[, 1:2] <- lapply(delta_agg[, 1:2], factor)
delta_agg$method <- factor(delta_agg$method,
                         levels = c("population", "sample", "sample\_ipw",
                                    "sample\_subwt", "sample\_substrat"))

# Combine MSE and delta columns
delta_comb2 <- rbind(delta_agg[, 1:3], delta_agg[, 1:3])
delta_comb2$measure <- rep(c("Bias", "MSE"), each = nrow(delta_agg))
delta_comb2$value <- round(c(delta_agg$delta, delta_agg$mse), digits = 5)
delta_comb2$event <- rep("Event\_1", nrow(delta_comb2))

# Produce table
tab1 <- tabular(Heading() * RowFactor(method, spacing = 1,
                                      levelnames = c("Target\_population",
                                                    "Sample\_opposing\_bias",
                                                    "Wts\_inverse\_prob",
                                                    "Avg\_CI\_coxph\_obs",
                                                    "Avg\_CI\_coxph\_strata")) *
                   Heading() * RowFactor(cohort,
                                      levelnames = c("Control", "Treatment")) *
                   Heading() * Factor(event,
                                      levelnames = c("Event\_1")) *
                   Heading() * Factor(measure,
                                      levelnames = c("Bias", "MSE")) *
                   Heading() * (value) * Heading() * (mean),
                      data = delta_comb2)
# 4) Bias and RMSE at $t = 0.7$ following adjustment of treatment bias

## Loading the sheet
```r
delta_gather2 <- read.xlsx("sdh_t00_t14.xlsx", sheet = 1)
delta_gather2 <- delta_gather2[delta_gather2$time == 0.7,]
delta_agg2 <- aggregate(delta ~ cohort + method, data = delta_gather2, FUN = mean)
delta_agg2$mse <- sqrt(aggregate(delta ~ cohort + method, data = delta_gather2, FUN = function(x) mean(x^2)))
delta_agg2[, 1:2] <- lapply(delta_agg2[, 1:2], factor)
delta_agg2$method <- factor(delta_agg2$method, levels = c("population", "sample", "sample_ipw", "sample_subwt", "sample_substr"))
delta_comb <- rbind(delta_agg2[, 1:3], delta_agg2[, 1:3])
delta_comb$measure <- rep(c("Bias", "MSE"), each = nrow(delta_agg2))
delta_comb$value <- round(c(delta_agg2$delta, delta_agg2$mse), digits = 5)
delta_comb$event <- rep("Event_1", nrow(delta_comb))
```

## Convert factor and order groups
```r
delta_comb$measure <- rep(c("Bias", "MSE"), each = nrow(delta_agg2))
delta_comb$value <- round(c(delta_agg2$delta, delta_agg2$mse), digits = 5)
delta_comb$event <- rep("Event_1", nrow(delta_comb))
```

## Combine MSE and delta columns
```r
delta_comb$measure <- rep(c("Bias", "MSE"), each = nrow(delta_agg2))
delta_comb$value <- round(c(delta_agg2$delta, delta_agg2$mse), digits = 5)
delta_comb$event <- rep("Event_1", nrow(delta_comb))
```

## Produce table
```r
tab1 <- tabular(Heading()*RowFactor(method, spacing = 1, levelnames = c("Target_population", "Sample_treatment_bias", "Wts_inverse_prob", "Avg_CI_coxph_obs", "Avg_CI_coxph_strata"))*
              Heading()*RowFactor(cohort, levelnames = c("Control", "Treatment")) -
              Heading()*Factor(event, levelnames = c("Event_1")))*
              Heading()*Factor(measure, levelnames = c("Bias", "MSE"))*
              Heading()*(value)*Heading()*(mean), data = delta_comb)
toLatex(tab1)
```
Computation time evaluation

The following code was used to evaluate the computation times of the conditional adjustment methods for the cause-specific hazard model, which was discussed in section 3.4. The computation time was evaluated over $B = 100$, and only $n = 500$ observations were processed to reduce the overall computational burden.

```r
# 1) Evaluate computation time over $B = 100$ replication
set.seed(1546732)
comp_time <- microbenchmark(
  "coxwt" = cox_wt(Surv(obs_time, status) ~ cohort + age + mutation, 
  strata = "cohort", ref = "control", 
  data = sim_comp(N = 500, 
  k = 2, 
  base_haz = c(1, 1), 
  beta_coef = list(c(-0.5, 2, 0.5), 
  c(0.3, 1, -0.4)))),
  "rrwt" = rr_wt(Surv(obs_time, status) ~ cohort + age + mutation, 
  strata = "cohort", ref = "control", 
  data = sim_comp(N = 500, 
  k = 2, 
  base_haz = c(1, 1), 
  beta_coef = list(c(-0.5, 2, 0.5), 
  c(0.3, 1, -0.4)))),
  "coxstratwt" = coxstrat_wt(Surv(obs_time, status) ~ cohort + age + mutation, 
  strata = "cohort", ref = "control", n_qntl = 5, 
  data = sim_comp(N = 500, 
  k = 2, 
  base_haz = c(1, 1), 
  beta_coef = list(c(-0.5, 2, 0.5), 
  c(0.3, 1, -0.4)))),
  "stratawt" = strata_wt(Surv(obs_time, status) ~ cohort + age + mutation, 
  strata = "cohort", ref = "control", n_qntl = 5, 
  data = sim_comp(N = 500, 
  k = 2, 
  base_haz = c(1, 1), 
  beta_coef = list(c(-0.5, 2, 0.5), 
  c(0.3, 1, -0.4)))),
  times = 100)
write.xlsx(comp_time, "comp_time.xlsx")
```

# 2) Visually evaluate computation time of each adjustment method

```r
# Load the data
comp_time <- read.xlsx("comp_time.xlsx")
# Adjust group names and rescale nanoseconds to milliseconds
comp_time$expr <- factor(comp_time$expr, 
  levels = c("coxwt", 
  "coxstratwt", 
  "stratawt", 
  "rrwt"))
comp_time$time <- comp_time$time / 10^6 # convert nansec to millisecc
# Calculate mean and standard deviation
comp_time_summarized <- aggregate(time ~ expr, comp_time, FUN = mean)
comp_time_summarized$sd <- aggregate(time ~ expr, comp_time, FUN = sd)$time
```
# Produce plot

```r
ggplot(comp_time_summarized, aes(x = expr, y = time, fill = expr)) +
  geom_bar(stat = "identity", color = "black", width = 0.80, lwd = 1) +
  geom_errorbar(aes(ymin = time, ymax = time + sd), width = 0.5, lwd = 1) +
  coord_flip() +
  scale_y_log10(limits = c(1, 100000),
    breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x))) +
  scale_x_discrete(labels = c(expression(atop(italic("survival::coxph"), individual)),
    expression(atop(italic("survival::coxph"), stratified)),
    expression(atop(italic("survival::survfit"), stratified)),
    expression(atop(italic("riskRegression"), individual)))) +
  theme_classic() + theme(text = element_text(size = 18),
    legend.position = "none",
    axis.title.y = element_blank(),
    axis.line = element_line(size = 1),
    axis.ticks = element_line(size = 1)) +
  ylab("Time (ms)")
```
A.7 Adjustment of competing risks in breast cancer data

Load data and construct covariate matrix

First, the data had to be reconfigured in order to be able to work with it. Most notably, missing values were inconsistently labeled across covariates, characteristics that were represented by multiple covariates had to be reduced to one covariate, and some covariates consisted of many complex subgroups which required simplification. Importantly, in this study we are only interested in the subpopulation of patients that either underwent mastectomy or breast-conserving surgery with adjuvant radiotherapy. Furthermore, observations were removed if metastasis had already occurred or if they had undergone neoadjuvant therapies, as described by van Maaren et al. (2019).

# 1) Load data
df <- read.csv2("K21092.csv")

# 2) Reconstitute missing values
df[df == "X" ] <- NA
df[df == "" ] <- NA
miss_9 <- c("diffgr", "multifocaal", "tweede_later", 
"oestrrec_stat", "progrsec_stat", "reclater1",
"reclater2", "reclater3", "reclater4",
"reclater5", "reclater6", "reclater7",
"reclater8", "reclater9", "reclater10")
df[, miss_9][df[, miss_9] == "9"] <- NA
miss_0_3 <- c("oestrrec_uitslag", "progrsec_uitslag")
df[, miss_0_3][df[, miss_0_3] %in% c("0", "3") ] <- NA
df[, "her2_uitslag"][df[, "her2_uitslag"] %in% c("4", "7", "9") ] <- NA
df[, "lypos"] [df[, "lypos"] == "98"] <- NA
df[, "tumorgrootte"] [df[, "tumorgrootte"] == "999"] <- NA
rm(miss_0_3, miss_9)

# 3) Remove neoadjuvant observations and combine surgery and radiotherapy
df <- df[(df$chirurgie == "Mastectomie" & df$indrt == 0) |
(df$chirurgie == "Borstsparend" & df$indrt == 2), ]
df <- df[df$indchemo %in% c(0, 2), ]
df <- df[df$indhorm %in% c(0, 2), ]
df <- df[df$indtarget %in% c(0, 2), ]
df <- df[df$pm == 0 | is.na(df$pm), ]

# 4) Construct covariate matrix
covs <- df[, c(3:21, 23:24, 26:31)]
cov_fac <- c("ct", "cn", "cm", "pt", "pm", "pm",
"topog", "later", "morf", "diffgr", "multifocaal",
"oestrrec_stat", "oestrrec_uitslag", "progrsec_stat",
"progrsec_uitslag", "her2_uitslag", "indchemo",
"indhorm", "indtarget", "trastuzumab", "ses",
"okd_jn", "typezkh", "ikcentrumcode")
cov_num <- c("leeft", "tumorgrootte", "lypos")
covs[, cov_fac] <- lapply(covs[, cov_fac], factor)
covs[, cov_num] <- lapply(covs[, cov_num], as.numeric)
covs$cohort <- factor(paste(df$chirurgie, df$indrt))
rm(cov_num, cov_fac)

# 5) Remove duplicate covariates
covs <- covs[, which(!names(covs) %in% c("oestrrec_uitslag", 
"progrsec_uitslag", "reclater1", "reclater2", 
"reclater3", "reclater4", "reclater5", "reclater6", 
"reclater7", "reclater8", "reclater9", "reclater10")])
"indtarget"))]
covs <- covs[, which(!names(covs) %in% c("ct", "cn", "cm", "pm" ))]

# 6) Combine covariates and subcategories
covs$pt <- factor(substring(covs$pt, 1, 1))
covs$pn <- factor(substring(covs$pn, 1, 1))
covs$diffgr <- factor(ifelse(as.character(covs$diffgr) == "4", "3", as.character(covs$diffgr)))
covs$her2_uitslag <- factor(ifelse(as.character(covs$her2_uitslag) == "0", "1", as.character(covs$her2_uitslag)))
covs$systherapy <- factor(as.numeric(as.numeric(covs$indhorm) > 1) + ifelse(as.numeric(as.numeric(covs$indhorm) > 1) == 1, 2, 0))
covs$topog <- substring(covs$topog, 4, 4)
covs$subloc <- factor(ifelse(covs$topog %in% c(4, 5, 6), "Outer quadrants", ifelse(covs$topog %in% c(2, 3), "Inner quadrants", ifelse(covs$topog %in% c(0, 1), "Central parts", ifelse(covs$topog == 8, "Overlapping lesions", NA)))))
covs$sess <- as.numeric(covs$sess)
covs$sess <- factor(ifelse(covs$sess %in% 1:3, "Low", ifelse(covs$sess %in% 4:7, "Medium", "High")))
covs$sess <- relevel(covs$sess, ref = "Low")

# Remove old covariate structures
covs <- covs[, which(!names(covs) %in% c("indhorm", "indchemo", "morf", "topog" ))]

# 7) Final overview of covariate matrix
summary(covs)

Construct outcome variable and combine with covariate matrix

Each outcome was represented by a single column, which indicated whether the endpoint had occurred (0 = no, 1 = yes). Furthermore, outcomes within the same class (mortality, recurrence, or second primary tumor) shared the same observation time. Hence, the outcomes were first summarized within each class, and then a final endpoint was constructed based on the class with the smallest observation time. If there was overlap in observation time between endpoints, then the most severe endpoint was chosen as the observed endpoint, which was the case for a small fraction of observations (< 100).

# 1) Summarize second primary tumor response
resp_tweede <- df[, c(32:34)]
resp_tweede$andermal <- ifelse(rowSums(resp_tweede) > 1, 0, resp_tweede$andermal)
tweede_stat <- numeric(nrow(resp_tweede))
for(i in 1:nrow(resp_tweede)){

92
tweede_stat[i] <- ifelse(sum(resp_tweede[i, ]) > 0, 
  names(resp_tweede[i, ])[which(resp_tweede[i, ] == 1)], 
  NA)
}
tweede_end <- data.frame(tweede_stat = tweede_stat, 
  tweede_time = df$tweede_tijd)
tweede_end$tweede_stat[tweede_end$tweede_stat == "andermal"] <- "tweede_misc"

# Trim second primary tumor response to 10 year follow-up (=3653 days)
tweede_end$tweede_stat <- ifelse(tweede_end$tweede_time > 3653, 
  0, tweede_end$tweede_stat)
tweede_end$tweede_time <- ifelse(tweede_end$tweede_time > 3653, 
  3653, tweede_end$tweede_time)

# 2) Summarize recurrence response
resp_rec <- df[, c(79, 81, 83)]
resp_rec[is.na(resp_rec)] <- 0

# Simplify recurrence endpoint
rec_stat <- numeric(nrow(resp_rec))
for(i in 1:nrow(resp_rec)) {
  rec_stat[i] <- ifelse(resp_rec[i, 2] == 1, "rr_metastasis", 
    ifelse(resp_rec[i, 3] == 1, "rr_regional", 
      ifelse(resp_rec[i, 1] == 1, "rr_local", NA)))
}
rec_end <- data.frame(rec_stat = rec_stat, 
  rec_time = df$fu10y_fuptijd)
rec_end$rec_time <- ifelse(is.na(rec_end$rec_stat), NA, rec_end$rec_time)

# 3) Summarize mortality response
surv_end <- df[, c(87, 88)]
surv_end$vitstat <- ifelse(surv_end$vitstat == 1, "mortality", 0)

# Trim mortality to 10 year follow-up (=3653 days)
surv_end$vitfup <- ifelse(surv_end$vitfup > 3653, 
  3653, surv_end$vitfup)

# 4) Combine response matrices
obs_times <- data.frame(tweede_time = tweede_end$tweede_time, 
  rec_time = rec_end$rec_time, 
  surv_time = surv_end$vitfup)
osb_states <- data.frame(tweede_stat = tweede_end$tweede_stat, 
  rec_stat = rec_end$rec_stat, 
  surv_stat = surv_end$vitstat)

# 5) Select most severe endpoint as the observed endpoint
obs_time <- unname(unlist(apply(obs_times, 1, 
  function(x) min(x, na.rm = T))))
obs_ind <- unlist(lapply(apply(obs_times, 1, 
  function(x) which(x == min(x, na.rm = T))), 
  max))
obs_status <- numeric(nrow(obs_states))
for(i in 1:nrow(obs_states)) {
  obs_status[i] <- obs_states[i, obs_ind[i]]
}
resp <- data.frame(obs_time, 
  obs_status = factor(obs_status))

# 6) Remove incomplete cases and combine covariate matrix and response
df_comp <- cbind(resp, covs)

Generate frequency tables

An overview of the composition of the breast cancer data was constructed, and presented in Table 5. The cohort-specific group frequencies and proportions were obtained separately for each discrete covariate. For continuous covariates, the cohort-specific mean and interquartile range (IQR) were estimated instead.

# 1) Extract discrete covariates and recode age

cov_tbl <- covs[, c(-4, -7)]
cov_tbl$leeft <- cut(as.numeric(cov_tbl$leeft),
  breaks = c(0, 40, 49, 59, 69, 79),
  max(as.numeric(cov_tbl$leeft))),
  labels = c("<40", "40-49", "50-59",
  "60-69", "70-79", "80-80")

# Replace NAs by a unique value

cov_tbl <- apply(cov_tbl, 2, function(x) ifelse(is.na(x), "Unknown", x))
cov_tbl <- as.data.frame(cov_tbl)
cov_tbl[, 1:ncol(cov_tbl)] <- lapply(cov_tbl[, 1:ncol(cov_tbl)], factor)

# 2) Frequency and proportion tables for discrete covariates

cov_freq <- apply(cov_tbl, MARGIN = 2,
  FUN = function(x) table(x, cov_tbl$cohort))
cov_prop <- lapply(cov_freq,
  FUN = function(x) round(proportions(x, 2) * 100, 1))

# 3) Mean and IQR for continuous covariates (tumor size and # lymph nodes)
cov_tumorsize <- aggregate(tumorgrootte ~ cohort, data = covs,
  FUN = function (x) c(mean = mean(x),
  low_IQR = quantile(x, 0.25),
  high_IQR = quantile(x, 0.75)))
cov_nlymph <- aggregate(lypos ~ cohort, data = covs,
  FUN = function (x) c(mean = mean(x),
  low_IQR = quantile(x, 0.25),
  high_IQR = quantile(x, 0.75)))

---

94
Produce inverse probability weighted cumulative incidence curves

Inverse probability weights were constructed as described in section 2.3, and both inverse probability weighted and unweighted cumulative incidence curves were constructed and compared between each other. Covariate effects on treatment allocation were examined by using an ANOVA table.

```r
# 1) Compute inverse probability weights
comp_ipw <- df_comp
comp_ipw$ipw <- est_ipw(cohort ~ ., data = df_comp[, 3:ncol(df_comp)])

# 2) Plot distribution of propensity scores
score_plot <- ggplot(comp_ipw, aes(x = 1 / ipw, fill = cohort)) +
stat_density(alpha = 0.4, col = "black",
position = "identity", outline.type = "full") +
theme_bw() +
theme(text = element_text(size = 14, face = "bold"),
axis.title.x = element_text(size = 12),
axis.title.y = element_text(size = 12),
legend.position = "top",
legend.title = element_blank(),
legend.box = "vertical") +
scale_y_continuous(breaks = c(0, 1, 2, 3, 4),
limits = c(0, 3.5)) +
scale_x_continuous(limits=c(0, 1),) +
scale_fill_discrete(labels = c("Breast-conserving\therapy",
"Mastectomy")) +
xlab("Propensity␣score") + ylab("Density␣\n")

# 3) Examine summary output and Chi-square test
ipw_model <- glm(cohort ~ ., data = df_comp[, 3:ncol(df_comp)], family = "binomial")
anova(ipw_model, test = "Chisq")

# 4) Construct cumulative incidence curves

# Construct cumulative incidence curves
sf_unadj <- long_sfc(survfit(Surv(obs_time, obs_status) ~ cohort,
data = comp_ipw)) # Unadjusted
ipw_adj <- long_sfc(survfit(Surv(obs_time, obs_status) ~ cohort,
data = comp_ipw, weights = ipw)) # IPW

# Combine data sets to produce plot
sf_both <- rbind(ipw_adj, sf_unadj)
sf_both$type <- c(rep("IPW-adjusted", nrow(ipw_adj)),
rep("Unadjusted", nrow(sf_unadj)))

# Add class and refine group names and time points
sf_both$class <- ifelse(sf_both$event %in% c("rr_local",
"rr_regional",
"rr_metastasis"), "recurrence",
ifelse(sf_both$event %in% c("tweede_DCIS",
"tweede_invasief",
"tweede_misc"), "2nd\primary\tumor",
"mortality"))
sf_both$time_year <- sf_both$time / 365.25
sf_both$event <- factor(sf_both$event,
levels = c("mortality", "rr_metastasis", "rr_regional",
"rr_local", "tweede_invasief",
"tweede_DCIS", "tweede_misc"))
```

95
levels(sf_both$event) <- c("Mortality", "Metastasis", "Regional", 
"Local", "Invasive", 
"DCIS", "Miscellaneous")

# 5) Produce multi-panel plot
p1 <- ggplot(sf_both[sf_both$class == "mortality", ], 
aes(x = time_year, y = prob, col = strata, lty = type)) +
geom_step() + facet_wrap("event", ncol = 3) +
ylim(0, 1) + scale_x_continuous(breaks = seq(0, 10, 2)) +
theme_bw() + theme(text = element_text(size = 12, face = "bold"),
axis.title.x = element_text(size = 12),
axis.title.y = element_text(size = 12),
legend.position = "right") +
scale_color_discrete(labels = c("Breast-conserving therapy", 
"Mastectomy")) +
labs(col = "Cohort", lty = "Method") +
ylab("Cumulative incidence\n\n")

p2 <- ggplot(sf_both[sf_both$class == "recurrence", ],
aes(x = time_year, y = prob, col = strata, lty = type)) +
geom_step() + facet_wrap("event", ncol = 3) +
ylim(0, 1) + scale_x_continuous(breaks = seq(0, 10, 2)) +
theme_bw() + theme(text = element_text(size = 12, face = "bold"),
axis.title.x = element_text(size = 12),
axis.title.y = element_text(size = 12),
legend.position = "none") +
xlab("Time from initial diagnosis\n\nyears") +
ggtitle("Tumor recurrence")

p3 <- ggplot(sf_both[sf_both$class == "2nd primary tumor", ],
aes(x = time_year, y = prob, col = strata, lty = type)) +
geom_step() + facet_wrap("event", ncol = 3) +
ylim(0, 1) + scale_x_continuous(breaks = seq(0, 10, 2)) +
theme_bw() + theme(text = element_text(size = 12, face = "bold"),
axis.title.x = element_text(size = 12),
axis.title.y = element_text(size = 12),
legend.position = "none") +
xlab("Time from initial diagnosis\n\nyears") +
ylab("Cumulative incidence\n\n") +
ggtitle("Second primary tumour")

ipw_plot <- ggarrange(p1, p2, p3, nrow = 3, ncol = 1,
labels = c("A", "B", "C"))

# 6) Extract coefficient table
xtable(ipw_model, digits = c(0, 3, 2, 3))
Produce conditional adjusted cumulative incidence curves

Lastly, unadjusted cumulative incidence curves were also compared to the conditional
adjustment method, as described in section 2.3. This method was carried out by us-
using the computationally efficient cause-specific Cox regression model provided by the
riskRegression package. Furthermore, each covariate was tested for violation of the
proportional hazards assumption prior to the construction of the covariate-adjusted cu-
mulative incidence curve.

# 1) Construct and examine cause-specific hazard model

# Construct formula
pred <- which(!names(df_comp) %in% c("obs_time", "obs_status", "cohort"))
pred_formula <- paste(names(df_comp[, pred]), collapse = "+")
bc_formula <- as.formula(paste("Surv(obs_time, obs_status) ~", pred_formula))

# Cause-specific Cox regression model
bc_coxph <- coxph(bc_formula, data = df_comp, id = 1:nrow(df_comp))
bc_zph <- cox.zph(bc_coxph) # verify proportional hazards

# 2) Construct cumulative incidence curves

sf_unadj <- long_sfc(survfit(Surv(obs_time, obs_status) ~ cohort, data = df_comp))
rr_adj <- rr_wt(Surv(obs_time, obs_status) ~ ., strata = "cohort", data = df_comp)
rr_adj <- gather(rr_adj[, c(1:2, 4:10)], event, prob, 3:9) # long format

# Combine data sets to produce plot
sf_comb <- rbind(rr_adj, sf_unadj)
sf_comb$type <- c(rep("Cox-adjusted", nrow(rr_adj)),
                rep("Unadjusted", nrow(sf_unadj)))

# Add class and refine group names and time points
sf_comb$class <- ifelse(sf_comb$event %in% c("rr_local",
                         "rr_regional", "rr_metastasis"), "recurrence",
    ifelse(sf_comb$event %in% c("tweede_DCIS",
                         "tweede_invasief", "tweede_misc"),
    "2nd primary tumor", "mortality"))

sf_comb$time_year <- sf_comb$time / 365.25
sf_comb$event <- factor(sf_comb$event, levels = c("Mortality", "Metastasis", "Regional",
                         "Local", "Invasive", "DCIS", "Miscellaneous")

# 3) Produce multi-panel plot

pl <- ggplot(sf_comb[, sf_comb$class == "mortality", ],
aes(x = time_year, y = prob, col = strata, lty = type)) +
geom_step() + facet_wrap("event", ncol = 3) +
ylim(0, 1) + scale_x_continuous(breaks = seq(0, 10, 2)) +
theme_bw() + theme(text = element_text(size = 12, face = "bold"),
axis.title.x = element_text(size = 12),
axis.title.y = element_text(size = 12),
axis.title.y = element_text(size = 12),
p2 <- ggplot(sf_comb[sf_comb$class == "recurrence",], aes(x = time_year, y = prob, col = strata, lty = type)) + geom_step() + facet_wrap("event", ncol = 3) + ylim(0, 1) + scale_x_continuous(breaks = seq(0, 10, 2)) + theme_text(element_text(size = 12, face = "bold"), axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12), legend.position = "none") + xlab("Time from initial diagnosis (years)") + ylab("Cumulative incidence") + ggtitle("Tumor recurrence")

p3 <- ggplot(sf_comb[sf_comb$class == "2nd primary tumor",], aes(x = time_year, y = prob, col = strata, lty = type)) + geom_step() + facet_wrap("event", ncol = 3) + ylim(0, 1) + scale_x_continuous(breaks = seq(0, 10, 2)) + theme_text(element_text(size = 12, face = "bold"), axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12), legend.position = "none") + xlab("Time from initial diagnosis (years)") + ylab("Cumulative incidence") + ggtitle("Second primary tumor")

rr_plot <- ggarrange(p1, p2, p3, nrow = 3, ncol = 1, labels = c("A", "B", "C"))

# 4) Extract cause-specific hazard model coefficients
xtable(bc_coxph, digits = c(0, 3, 3, 3, 2, 3))

# 2) Mortality
# 3) Local recurrence
# 4) Metastasis
# 5) Regional recurrence
# 6) DCIS
# 7) Invasive second primary tumor
# 8) Miscellaneous second primary tumor
B Breast cancer model coefficients

B.1 Covariate effects on treatment allocation (IPW)

Table 6: Coefficient table from the IPW adjustment method. The coefficients were obtained my using a logistic regression model, which modeled treatment allocation (Breast-conserving therapy = 0, Mastectomy = 1) as a function of the tumour-, patient-, treatment-, and hospital-related characteristics. The estimates and their standard errors (SE) are reported on the log odds (logit) scale. Covariate effects on treatment allocation were considered to be statistically significant at $p < 0.05$.

| Characteristics                                      | Estimate | SE  | z value | Pr(|z|) |
|------------------------------------------------------|----------|-----|---------|--------|
| (Intercept)                                          | -4.139   | 0.258 | -16.04  | 0.000  |
| Age                                                  | 0.031    | 0.003 | 10.61   | 0.000  |
| Tumour size in mm                                    | 0.058    | 0.006 | 9.43    | 0.000  |
| T stage (T1 = 0)                                     |          |      |         |        |
| T2                                                   | 0.153    | 0.101 | 1.51    | 0.131  |
| N stage (N0 = 0)                                     |          |      |         |        |
| N1                                                   | -0.503   | 0.137 | -3.69   | 0.000  |
| Number of positive lymph nodes                       | -0.009   | 0.070 | -0.13   | 0.897  |
| Grade (1 = 0)                                        |          |      |         |        |
| 2                                                    | 0.037    | 0.078 | 0.48    | 0.634  |
| 3                                                    | 0.064    | 0.097 | 0.66    | 0.507  |
| Multifocality (No = 0)                               |          |      |         |        |
| Yes                                                  | 1.764    | 0.090 | 19.49   | 0.000  |
| Histological tumour type (Ductal = 0)                |          |      |         |        |
| Lobular                                              | 0.328    | 0.100 | 3.29    | 0.001  |
| Mixed                                                | 0.395    | 0.148 | 2.67    | 0.008  |
| Other                                                | -0.153   | 0.151 | -1.01   | 0.311  |
| Sublocalisation (Outer quadrants = 0)                |          |      |         |        |
| Inner quadrants                                       | 0.015    | 0.078 | 0.20    | 0.844  |
| Central parts                                        | 0.597    | 0.112 | 5.36    | 0.000  |
| Overlapping lesions                                  | 0.268    | 0.073 | 3.68    | 0.000  |
| Lateralisation (Left = 0)                            |          |      |         |        |
| Right                                                | 0.076    | 0.058 | 1.31    | 0.191  |
| Estrogen status (Negative = 0)                       |          |      |         |        |
| Positive                                             | -0.114   | 0.121 | -0.94   | 0.346  |
| Progesteron status (Negative = 0)                    |          |      |         |        |
| Positive                                             | -0.099   | 0.078 | -1.26   | 0.206  |
| HER2 status (Negative = 0)                           |          |      |         |        |
| Unclear                                              | -0.003   | 0.118 | -0.02   | 0.982  |
| Positive                                             | 0.269    | 0.106 | 2.52    | 0.012  |
| Characteristics                                           | Estimate | SE   | z value | Pr(>|z|) |
|-----------------------------------------------------------|----------|------|---------|---------|
| Targeted therapy (trastuzumab) (No = 0)                   | -0.158   | 0.160| -0.99   | 0.322   |
| Yes                                                       | -0.262   | 0.145| -1.81   | 0.071   |
| Adjuvant systematic therapy (None = 0)                    | -0.092   | 0.092| -1.00   | 0.319   |
| Chemotherapy                                             | 0.040    | 0.111| 0.36    | 0.717   |
| Endocrine therapy                                        | 1.253    | 0.083| 15.17   | 0.000   |
| Axillary lymph node dissection (No = 0)                   | 0.007    | 0.071| -0.10   | 0.922   |
| Social economic status (Low = 0)                          | -0.121   | 0.077| -1.58   | 0.115   |
| Medium                                                   | -0.165   | 0.066| -2.51   | 0.012   |
| High                                                     | -0.008   | 0.115| -0.07   | 0.947   |
| Hospital type (General = 0)                               | 0.312    | 0.127| 2.46    | 0.014   |
| Top-clinical                                             | -0.170   | 0.133| -1.28   | 0.201   |
| Academic                                                 | -0.162   | 0.108| -1.51   | 0.132   |
| Hospital region (A = 0)                                  | 0.208    | 0.119| 1.74    | 0.081   |
| B                                                        | 0.451    | 0.106| 4.25    | 0.000   |
| C                                                        | 0.266    | 0.127| 2.09    | 0.036   |
| D                                                        | 0.135    | 0.111| 1.22    | 0.223   |
| E                                                        | -0.005   | 0.109| -0.04   | 0.967   |
B.2 Covariate effects on competing outcomes

Table 7: Coefficient table from the cause-specific hazard model used to construct the predicted cumulative incidence curves for the conditional adjustment method. Cause-specific estimates, hazard ratios, and standard error (SE) are given for each tumour-, patient-, treatment-, and hospital-related characteristic. Covariate effects on the cause-specific hazard were considered to be statistically significant at $p < 0.05$.

| Outcome Characteristics | Estimate | HR  | SE  | z value | Pr($>|z|)$ |
|-------------------------|----------|-----|-----|---------|-----------|
| Mortality               |          |     |     |         |           |
| Age                     | 0.109    | 1.115 | 0.005 | 22.53   | 0.000     |
| Tumour size in mm       | 0.023    | 1.023 | 0.007 | 3.22    | 0.001     |
| T stage (T1 = 0)        |          |     |     |         |           |
| T2                      | 0.009    | 1.009 | 0.137 | 0.06    | 0.949     |
| N stage (N0 = 0)        |          |     |     |         |           |
| N1                      | -0.281   | 0.755 | 0.183 | -1.54   | 0.124     |
| Number of positive lymph nodes | 0.206 | 1.229 | 0.100 | 2.06    | 0.039     |
| Grade (1 = 0)           |          |     |     |         |           |
| 2                       | -0.032   | 0.969 | 0.103 | -0.31   | 0.758     |
| 3                       | -0.079   | 0.924 | 0.135 | -0.58   | 0.560     |
| Multifocality (No = 0)  |          |     |     |         |           |
| Yes                     | 0.140    | 1.150 | 0.123 | 1.14    | 0.255     |
| Histological tumour type (Ductal = 0) |          |     |     |         |           |
| Lobular                 | -0.322   | 0.725 | 0.141 | -2.29   | 0.022     |
| Mixed                   | 0.171    | 1.187 | 0.204 | 0.84    | 0.401     |
| Other                   | 0.149    | 1.160 | 0.174 | 0.85    | 0.393     |
| Sublocalisation (Outer quadrants = 0) |          |     |     |         |           |
| Inner quadrants         | -0.084   | 0.919 | 0.112 | -0.75   | 0.452     |
| Central parts           | -0.195   | 0.823 | 0.149 | -1.31   | 0.191     |
| Overlapping lesions     | 0.003    | 1.003 | 0.100 | 0.03    | 0.980     |
| Lateralisation (Left = 0) |          |     |     |         |           |
| Right                   | -0.036   | 0.964 | 0.080 | -0.45   | 0.652     |
| Estrogen status (Negative = 0) |          |     |     |         |           |
| Positive                | 0.004    | 1.004 | 0.169 | 0.02    | 0.981     |
| Progesteron status (Negative = 0) |          |     |     |         |           |
| Positive                | 0.007    | 1.007 | 0.107 | 0.06    | 0.949     |
| HER2 status (Negative = 0) |          |     |     |         |           |
| Unclear                 | 0.034    | 1.034 | 0.154 | 0.22    | 0.828     |
| Positive                | -0.183   | 0.833 | 0.147 | -1.25   | 0.212     |
| Targeted therapy (trastuzumab) (No = 0) |          |     |     |         |           |
| Yes                     | -1.132   | 0.322 | 0.533 | -2.12   | 0.034     |
| Outcome Characteristics                                      | Estimate | HR   | SE   | z value | Pr(>|z|) |
|---------------------------------------------------------------|----------|------|------|---------|---------|
| Adjuvant systematic therapy (None = 0)                        |          |      |      |         |         |
| Chemotherapy                                                 | 0.361    | 1.435| 0.223| 1.62    | 0.105   |
| Endocrine therapy                                            | -0.213   | 0.808| 0.112| -1.91   | 0.057   |
| Both                                                         | -0.315   | 0.730| 0.243| -1.30   | 0.195   |
| Axillary lymph node dissection (No = 0)                      |          |      |      |         |         |
| Yes                                                          | 0.424    | 1.528| 0.100| 4.23    | 0.000   |
| Social economic status (Low = 0)                             |          |      |      |         |         |
| Medium                                                        | -0.062   | 0.939| 0.095| -0.66   | 0.509   |
| High                                                         | -0.176   | 0.839| 0.104| -1.68   | 0.092   |
| Hospital type (General = 0)                                  |          |      |      |         |         |
| Top-clinical                                                 | -0.075   | 0.928| 0.090| -0.83   | 0.406   |
| Academic                                                     | 0.048    | 1.049| 0.179| 0.27    | 0.789   |
| Hospital region (A = 0)                                      |          |      |      |         |         |
| B                                                            | 0.211    | 1.235| 0.171| 1.24    | 0.217   |
| C                                                            | -0.201   | 0.818| 0.212| -0.95   | 0.343   |
| D                                                            | 0.049    | 1.050| 0.147| 0.33    | 0.738   |
| E                                                            | 0.271    | 1.312| 0.165| 1.65    | 0.099   |
| F                                                            | 0.270    | 1.310| 0.143| 1.88    | 0.059   |
| G                                                            | 0.062    | 1.064| 0.186| 0.33    | 0.739   |
| H                                                            | 0.171    | 1.187| 0.150| 1.14    | 0.254   |
| I                                                            | 0.014    | 1.014| 0.155| 0.09    | 0.928   |
| Local recurrence                                             |          |      |      |         |         |
| Age                                                          | -0.009   | 0.991| 0.007| -1.28   | 0.200   |
| Tumour size in mm                                             | 0.029    | 1.029| 0.016| 1.84    | 0.066   |
| T stage (T1 = 0)                                              |          |      |      |         |         |
| T2                                                           | 0.059    | 1.061| 0.285| 0.21    | 0.837   |
| N stage (N0 = 0)                                              |          |      |      |         |         |
| N1                                                           | 0.062    | 1.064| 0.392| 0.16    | 0.874   |
| Number of positive lymph nodes                                | 0.171    | 1.186| 0.206| 0.83    | 0.406   |
| Grade (1 = 0)                                                |          |      |      |         |         |
| 2                                                            | 0.468    | 1.597| 0.206| 2.27    | 0.023   |
| 3                                                            | 0.723    | 2.061| 0.269| 2.68    | 0.007   |
| Multifocality (No = 0)                                       |          |      |      |         |         |
| Yes                                                          | 0.706    | 2.026| 0.200| 3.53    | 0.000   |
| Histological tumour type (Ductal = 0)                        |          |      |      |         |         |
| Lobular                                                     | -0.215   | 0.806| 0.280| -0.77   | 0.442   |
| Mixed                                                        | 0.227    | 1.255| 0.338| 0.67    | 0.501   |
| Other                                                        | -0.019   | 0.981| 0.394| -0.05   | 0.962   |
| Outcome Characteristics | Estimate | HR  | SE  | z value | Pr(>|z|) |
|--------------------------|----------|-----|-----|---------|---------|
| Sublocalisation (Outer quadrants = 0) |          |     |     |         |         |
| Inner quadrants          | 0.272    | 1.313 | 0.200 | 1.36    | 0.173   |
| Central parts             | 0.392    | 1.480 | 0.286 | 1.37    | 0.170   |
| Overlapping lesions      | 0.060    | 1.062 | 0.196 | 0.31    | 0.759   |
| Lateralisation (Left = 0) |          |     |     |         |         |
| Right                    | -0.113   | 0.893 | 0.154 | -0.74   | 0.462   |
| Estrogen status (Negative = 0) |          |     |     |         |         |
| Positive                 | 0.136    | 1.146 | 0.311 | 0.44    | 0.661   |
| Progesteron status (Negative = 0) |          |     |     |         |         |
| Positive                 | 0.100    | 1.105 | 0.217 | 0.46    | 0.646   |
| HER2 status (Negative = 0) |          |     |     |         |         |
| Unclear                  | -0.206   | 0.813 | 0.317 | -0.65   | 0.515   |
| Positive                 | 0.132    | 1.141 | 0.260 | 0.51    | 0.611   |
| Targeted therapy (trastuzumab) (No = 0) |          |     |     |         |         |
| Yes                      | -0.238   | 0.789 | 0.503 | -0.47   | 0.637   |
| Adjuvant systematic therapy (None = 0) |          |     |     |         |         |
| Chemotherapy             | -1.229   | 0.293 | 0.394 | -3.12   | 0.002   |
| Endocrine therapy        | -1.051   | 0.349 | 0.263 | -4.00   | 0.000   |
| Both                     | -1.775   | 0.169 | 0.321 | -5.54   | 0.000   |
| Axillary lymph node dissection (No = 0) |          |     |     |         |         |
| Yes                      | 0.005    | 1.005 | 0.218 | 0.02    | 0.983   |
| Social economic status (Low = 0) |          |     |     |         |         |
| Medium                   | -0.006   | 0.994 | 0.195 | -0.03   | 0.975   |
| High                     | 0.246    | 1.279 | 0.199 | 1.23    | 0.217   |
| Hospital type (General = 0) |          |     |     |         |         |
| Top-clinical             | 0.115    | 1.122 | 0.176 | 0.65    | 0.514   |
| Academic                 | -0.215   | 0.807 | 0.326 | -0.66   | 0.511   |
| Hospital region (A = 0)  |          |     |     |         |         |
| B                        | 0.633    | 1.884 | 0.320 | 1.98    | 0.048   |
| C                        | 0.395    | 1.485 | 0.334 | 1.18    | 0.237   |
| D                        | 0.391    | 1.478 | 0.291 | 1.34    | 0.179   |
| E                        | 0.228    | 1.256 | 0.327 | 0.70    | 0.486   |
| F                        | 0.303    | 1.354 | 0.297 | 1.02    | 0.308   |
| G                        | 0.203    | 1.225 | 0.345 | 0.59    | 0.557   |
| H                        | -0.092   | 0.912 | 0.334 | -0.28   | 0.783   |
| I                        | 0.301    | 1.351 | 0.287 | 1.05    | 0.295   |
| Outcome Characteristics | Estimate | HR | SE  | z value | Pr(>|z|) |
|-------------------------|----------|----|-----|---------|---------|
| **Regional recurrence**  |          |    |     |         |         |
| Age                     | -0.043   | 0.958 | 0.012 | -3.42 | 0.001 |
| Tumour size in mm       | 0.079    | 1.082 | 0.022 | 3.62  | 0.000 |
| T stage (T1 = 0)        |          |    |     |         |         |
| T2                      | -0.905   | 0.405 | 0.435 | -2.08 | 0.037 |
| N stage (N0 = 0)        |          |    |     |         |         |
| N1                      | 0.951    | 2.589 | 0.559 | 1.70  | 0.089 |
| Number of positive lymph nodes | 0.061 | 1.063 | 0.281 | 0.22  | 0.828 |
| Grade (1 = 0)           |          |    |     |         |         |
| 2                       | 2.117    | 8.303 | 0.734 | 2.88  | 0.004 |
| 3                       | 2.784    | 16.178 | 0.764 | 3.64  | 0.000 |
| Multifocality (No = 0)  |          |    |     |         |         |
| Yes                     | 0.298    | 1.348 | 0.364 | 0.82  | 0.413 |
| Histological tumour type (Ductal = 0) |          |    |     |         |         |
| Lobular                 | -0.127   | 0.881 | 0.487 | -0.26 | 0.794 |
| Mixed                   | -0.101   | 0.904 | 0.732 | -0.14 | 0.891 |
| Other                   | -12.810  | 0.000 | 357.987 | -0.04 | 0.971 |
| Sublocalisation (Outer quadrants = 0) |          |    |     |         |         |
| Inner quadrants         | -0.223   | 0.800 | 0.335 | -0.67 | 0.505 |
| Central parts           | -0.262   | 0.770 | 0.531 | -0.49 | 0.622 |
| Overlapping lesions     | -0.388   | 0.678 | 0.330 | -1.17 | 0.240 |
| Lateralisation (Left = 0) |          |    |     |         |         |
| Right                   | 0.118    | 1.125 | 0.247 | 0.48  | 0.633 |
| Estrogen status (Negative = 0) |          |    |     |         |         |
| Positive                | -0.079   | 0.924 | 0.460 | -0.17 | 0.863 |
| Progesteron status (Negative = 0) |          |    |     |         |         |
| Positive                | -0.463   | 0.630 | 0.328 | -1.41 | 0.158 |
| HER2 status (Negative = 0) |          |    |     |         |         |
| Unclear                 | -1.645   | 0.193 | 1.013 | -1.62 | 0.104 |
| Positive                | 0.089    | 1.093 | 0.390 | 0.23  | 0.819 |
| Targeted therapy (trastuzumab) (No = 0) |          |    |     |         |         |
| Yes                     | -0.740   | 0.477 | 0.709 | -1.04 | 0.297 |
| Adjuvant systematic therapy (None = 0) |          |    |     |         |         |
| Chemotherapy            | -1.153   | 0.316 | 0.518 | -2.23 | 0.026 |
| Endocrine therapy       | -0.051   | 0.950 | 0.396 | -0.13 | 0.898 |
| Both                    | -1.404   | 0.246 | 0.465 | -3.02 | 0.003 |
| Axillary lymph node dissection (No = 0) |          |    |     |         |         |
| Yes                     | -0.674   | 0.510 | 0.385 | -1.75 | 0.080 |
### Outcome

#### Characteristics

| Social economic status (Low = 0) | Estimate | HR  | SE  | z value | Pr(>|z|) |
|---------------------------------|----------|-----|-----|---------|---------|
| Medium                          | 0.087    | 1.091 | 0.328 | 0.27 | 0.790 |
| High                            | 0.337    | 1.400 | 0.325 | 1.04 | 0.300 |

| Hospital type (General = 0)     | Estimate | HR  | SE  | z value | Pr(>|z|) |
|---------------------------------|----------|-----|-----|---------|---------|
| Top-clinical                    | -0.453   | 0.636 | 0.287 | -1.58 | 0.115 |
| Academic                        | 0.527    | 1.693 | 0.400 | 1.32 | 0.188 |

| Hospital region (A = 0)         | Estimate | HR  | SE  | z value | Pr(>|z|) |
|---------------------------------|----------|-----|-----|---------|---------|
| B                               | -0.433   | 0.649 | 0.775 | -0.56 | 0.577 |
| C                               | 0.252    | 1.286 | 0.585 | 0.43  | 0.667 |
| D                               | 0.613    | 1.846 | 0.455 | 1.35  | 0.178 |
| E                               | 0.004    | 1.004 | 0.589 | 0.01  | 0.995 |
| F                               | 0.409    | 1.505 | 0.453 | 0.90  | 0.367 |
| G                               | 0.720    | 2.054 | 0.572 | 1.26  | 0.208 |
| H                               | 0.705    | 2.024 | 0.437 | 1.61  | 0.107 |
| I                               | 0.960    | 2.611 | 0.441 | 2.18  | 0.029 |

| Metastasis                      | Estimate | HR  | SE  | z value | Pr(>|z|) |
|---------------------------------|----------|-----|-----|---------|---------|
| Age                             | -0.008   | 0.992 | 0.004 | -1.90 | 0.058 |
| Tumour size in mm               | 0.051    | 1.052 | 0.007 | 6.77  | 0.000 |
| T stage (T1 = 0)                |          |      |      |         |         |
| T2                              | -0.046   | 0.955 | 0.144 | -0.32 | 0.752 |
| N stage (N0 = 0)                |          |      |      |         |         |
| N1                              | 0.349    | 1.418 | 0.190 | 1.84  | 0.066 |
| Number of positive lymph nodes  | 0.172    | 1.187 | 0.089 | 1.92  | 0.055 |
| Grade (1 = 0)                   |          |      |      |         |         |
| 2                               | 0.599    | 1.820 | 0.145 | 4.12  | 0.000 |
| 3                               | 0.989    | 2.688 | 0.163 | 6.06  | 0.000 |
| Multifocality (No = 0)          |          |      |      |         |         |
| Yes                             | 0.350    | 1.419 | 0.121 | 2.89  | 0.004 |
| Histological tumour type (Ductal = 0) |      |      |      |         |         |
| Lobular                         | -0.309   | 0.734 | 0.173 | -1.79 | 0.074 |
| Mixed                           | -0.150   | 0.861 | 0.251 | -0.60 | 0.551 |
| Other                           | -0.663   | 0.515 | 0.295 | -2.25 | 0.025 |
| Sublocalisation (Outer quadrants = 0) |      |      |      |         |         |
| Inner quadrants                 | 0.261    | 1.299 | 0.111 | 2.35  | 0.019 |
| Central parts                   | -0.015   | 0.985 | 0.178 | -0.08 | 0.934 |
| Overlapping lesions             | 0.033    | 1.034 | 0.109 | 0.31  | 0.758 |
| Lateralisation (Left = 0)       |          |      |      |         |         |
| Right                           | -0.006   | 0.994 | 0.086 | -0.08 | 0.940 |
| Outcome Characteristics | Estimate | HR  | SE  | z value | Pr(>|z|) |
|-------------------------|----------|-----|-----|---------|---------|
| Estrogen status (Negative = 0) |          |     |     |         |         |
| Positive                | -0.015   | 0.985 | 0.167 | -0.09 | 0.927   |
| Progesteron status (Negative = 0) |          |     |     |         |         |
| Positive                | -0.232   | 0.793 | 0.115 | -2.02 | 0.043   |
| HER2 status (Negative = 0) |          |     |     |         |         |
| Unclear                 | 0.066    | 1.068 | 0.170 | 0.39  | 0.698   |
| Positive                | 0.294    | 1.341 | 0.136 | 2.16  | 0.031   |
| Targeted therapy (trastuzumab) (No = 0) |          |     |     |         |         |
| Yes                     | -0.505   | 0.603 | 0.221 | -2.29 | 0.022   |
| Adjuvant systemic therapy (None = 0) |          |     |     |         |         |
| Chemotherapy            | -0.517   | 0.596 | 0.188 | -2.74 | 0.006   |
| Endocrine therapy       | -0.361   | 0.697 | 0.141 | -2.56 | 0.011   |
| Both                    | -0.752   | 0.471 | 0.163 | -4.62 | 0.000   |
| Axillary lymph node dissection (No = 0) |          |     |     |         |         |
| Yes                     | -0.117   | 0.890 | 0.129 | -0.91 | 0.364   |
| Social economic status (Low = 0) |          |     |     |         |         |
| Medium                  | 0.040    | 1.041 | 0.106 | 0.37  | 0.708   |
| High                    | 0.092    | 1.097 | 0.113 | 0.82  | 0.413   |
| Hospital type (General = 0) |          |     |     |         |         |
| Top-clinical            | -0.099   | 0.905 | 0.097 | -1.03 | 0.304   |
| Academic                | -0.411   | 0.663 | 0.197 | -2.09 | 0.036   |
| Hospital region (A = 0) |          |     |     |         |         |
| B                       | 0.087    | 1.091 | 0.196 | 0.44  | 0.659   |
| C                       | -0.049   | 0.953 | 0.203 | -0.24 | 0.811   |
| D                       | 0.443    | 1.558 | 0.145 | 3.05  | 0.002   |
| E                       | -0.130   | 0.878 | 0.187 | -0.70 | 0.487   |
| F                       | -0.192   | 0.825 | 0.170 | -1.13 | 0.259   |
| G                       | 0.100    | 1.105 | 0.189 | 0.53  | 0.597   |
| H                       | -0.110   | 0.896 | 0.173 | -0.64 | 0.525   |
| I                       | 0.217    | 1.242 | 0.154 | 1.40  | 0.160   |
| DCIS                    |          |     |     |         |         |
| Age                     | -0.008   | 0.992 | 0.013 | -0.63 | 0.528   |
| Tumour size in mm       | -0.045   | 0.956 | 0.030 | -1.48 | 0.139   |
| T stage (T1 = 0)        |          |     |     |         |         |
| T2                      | 0.750    | 2.117 | 0.496 | 1.51  | 0.130   |
| N stage (N0 = 0)        |          |     |     |         |         |
| N1                      | -0.006   | 0.994 | 0.692 | -0.01 | 0.993   |
| Number of positive lymph nodes | 0.015 | 1.015 | 0.373 | 0.04  | 0.968   |
| Outcome                                                                 | Estimate | HR    | SE     | z value | Pr(>|z|) |
|-------------------------------------------------------------------------|----------|-------|--------|---------|---------|
| Grade (1 = 0)                                                           |          |       |        |         |         |
| 2                                                                       | -0.212   | 0.809 | 0.318  | -0.67   | 0.505   |
| 3                                                                       | -0.342   | 0.710 | 0.456  | -0.75   | 0.454   |
| Multifocality (No = 0)                                                  |          |       |        |         |         |
| Yes                                                                     | 0.626    | 1.870 | 0.334  | 1.88    | 0.061   |
| Histological tumour type (Ductal = 0)                                   |          |       |        |         |         |
| Lobular                                                                 | 0.093    | 1.098 | 0.453  | 0.21    | 0.837   |
| Mixed                                                                   | 0.624    | 1.867 | 0.541  | 1.15    | 0.248   |
| Other                                                                   | 1.056    | 2.874 | 0.426  | 2.48    | 0.013   |
| Sublocalisation (Outer quadrants = 0)                                   |          |       |        |         |         |
| Inner quadrants                                                         | 0.399    | 1.490 | 0.349  | 1.14    | 0.253   |
| Central parts                                                           | 0.817    | 2.264 | 0.446  | 1.83    | 0.067   |
| Overlapping lesions                                                     | 0.303    | 1.354 | 0.342  | 0.89    | 0.375   |
| Laterisation (Left = 0)                                                 |          |       |        |         |         |
| Right                                                                   | 0.100    | 1.105 | 0.264  | 0.38    | 0.705   |
| Estrogen status (Negative = 0)                                          |          |       |        |         |         |
| Positive                                                                | 0.623    | 1.864 | 0.733  | 0.85    | 0.395   |
| Progesteron status (Negative = 0)                                       |          |       |        |         |         |
| Positive                                                                | 0.365    | 1.441 | 0.416  | 0.88    | 0.380   |
| HER2 status (Negative = 0)                                              |          |       |        |         |         |
| Unclear                                                                 | 0.027    | 1.027 | 0.533  | 0.05    | 0.960   |
| Positive                                                                | -0.091   | 0.913 | 0.572  | -0.16   | 0.873   |
| Targeted therapy (trastuzumab) (No = 0)                                 |          |       |        |         |         |
| Yes                                                                     | 0.435    | 1.544 | 0.768  | 0.57    | 0.572   |
| Adjuvant systematic therapy (None = 0)                                  |          |       |        |         |         |
| Chemotherapy                                                            | 0.263    | 1.301 | 0.819  | 0.32    | 0.748   |
| Endocrine therapy                                                       | 0.039    | 1.039 | 0.449  | 0.09    | 0.931   |
| Both                                                                    | -0.178   | 0.837 | 0.507  | -0.35   | 0.726   |
| Axillary lymph node dissection (No = 0)                                 |          |       |        |         |         |
| Yes                                                                     | -0.266   | 0.767 | 0.402  | -0.66   | 0.509   |
| Social economic status (Low = 0)                                        |          |       |        |         |         |
| Medium                                                                  | 0.200    | 1.221 | 0.308  | 0.65    | 0.516   |
| High                                                                    | -0.196   | 0.822 | 0.372  | -0.53   | 0.599   |
| Hospital type (General = 0)                                             |          |       |        |         |         |
| Top-clinical                                                            | 0.482    | 1.619 | 0.342  | 1.41    | 0.159   |
| Academic                                                                | 1.437    | 4.207 | 0.389  | 3.69    | 0.000   |
| Outcome Characteristics | Estimate | HR | SE | z value | Pr(>|z|) |
|--------------------------|----------|----|----|---------|---------|
| Hospital region (A = 0)  |          |    |    |         |         |
| B                        | 0.064    | 1.066 | 0.482 | 0.13    | 0.894   |
| C                        | -0.853   | 0.426 | 0.763 | -1.12   | 0.264   |
| D                        | 0.540    | 1.716 | 0.402 | 1.34    | 0.179   |
| E                        | -0.499   | 0.607 | 0.583 | -0.86   | 0.392   |
| F                        | -0.060   | 0.942 | 0.505 | -0.12   | 0.906   |
| G                        | 0.193    | 1.213 | 0.528 | 0.37    | 0.714   |
| H                        | -1.120   | 0.326 | 0.764 | -1.47   | 0.143   |
| I                        | -0.793   | 0.453 | 0.597 | -1.33   | 0.184   |
| Invasive second primary tumor |      |    |    |         |         |
| Age                      | 0.026    | 1.026 | 0.003 | 7.81    | 0.000   |
| Tumour size in mm        | 0.006    | 1.006 | 0.007 | 0.89    | 0.375   |
| T stage (T1 = 0)         |          |    |    |         |         |
| T2                       | -0.248   | 0.780 | 0.122 | -2.04   | 0.042   |
| N stage (N0 = 0)         |          |    |    |         |         |
| N1                       | -0.027   | 0.973 | 0.162 | -0.17   | 0.868   |
| Number of positive lymph nodes | 0.062 | 1.064 | 0.086 | 0.72    | 0.472   |
| Grade (1 = 0)            |          |    |    |         |         |
| 2                        | -0.006   | 0.994 | 0.083 | -0.08   | 0.940   |
| 3                        | -0.005   | 0.995 | 0.111 | -0.05   | 0.962   |
| Multifocality (No = 0)   |          |    |    |         |         |
| Yes                      | 0.004    | 1.004 | 0.103 | 0.04    | 0.966   |
| Histological tumour type (Ductal = 0) |         |    |    |         |         |
| Lobular                  | 0.203    | 1.225 | 0.105 | 1.94    | 0.053   |
| Mixed                    | -0.073   | 0.930 | 0.178 | -0.41   | 0.682   |
| Other                    | 0.031    | 1.031 | 0.162 | 0.19    | 0.848   |
| Sublocalisation (Outer quadrants = 0) |      |    |    |         |         |
| Inner quadrants          | 0.070    | 1.073 | 0.087 | 0.81    | 0.419   |
| Central parts            | -0.019   | 0.981 | 0.130 | -0.15   | 0.883   |
| Overlapping lesions      | 0.078    | 1.081 | 0.082 | 0.94    | 0.347   |
| Lateralisation (Left = 0) |          |    |    |         |         |
| Right                    | -0.104   | 0.901 | 0.066 | -1.59   | 0.112   |
| Estrogen status (Negative = 0) |         |    |    |         |         |
| Positive                 | -0.043   | 0.958 | 0.136 | -0.31   | 0.754   |
| Progesteron status (Negative = 0) |       |    |    |         |         |
| Positive                 | -0.084   | 0.919 | 0.086 | -0.98   | 0.327   |
| HER2 status (Negative = 0) |         |    |    |         |         |
| Unclear                  | -0.119   | 0.888 | 0.137 | -0.87   | 0.385   |
| Positive                 | -0.199   | 0.819 | 0.128 | -1.56   | 0.118   |
| Outcome Characteristics                                      | Estimate | HR    | SE    | z value | Pr(>|z|) |
|--------------------------------------------------------------|----------|-------|-------|---------|---------|
| Targeted therapy (trastuzumab) (No = 0)                      |          |       |       |         |         |
| Yes                                                          | 0.003    | 1.003 | 0.215 | 0.01    | 0.990   |
| Adjuvant systematic therapy (None = 0)                       |          |       |       |         |         |
| Chemotherapy                                                | -0.148   | 0.862 | 0.173 | -0.86   | 0.392   |
| Endocrine therapy                                           | -0.063   | 0.939 | 0.103 | -0.61   | 0.543   |
| Both                                                        | -0.150   | 0.861 | 0.136 | -1.11   | 0.268   |
| Axillary lymph node dissection (No = 0)                      |          |       |       |         |         |
| Yes                                                         | 0.023    | 1.023 | 0.093 | 0.25    | 0.806   |
| Social economic status (Low = 0)                             |          |       |       |         |         |
| Medium                                                      | -0.067   | 0.935 | 0.081 | -0.83   | 0.406   |
| High                                                        | 0.016    | 1.017 | 0.085 | 0.19    | 0.847   |
| Hospital type (General = 0)                                  |          |       |       |         |         |
| Top-clinical                                                | -0.016   | 0.985 | 0.075 | -0.21   | 0.837   |
| Academic                                                    | 0.323    | 1.382 | 0.117 | 2.75    | 0.006   |
| Hospital region (A = 0)                                     |          |       |       |         |         |
| B                                                           | -0.109   | 0.897 | 0.142 | -0.77   | 0.443   |
| C                                                           | -0.231   | 0.793 | 0.150 | -1.54   | 0.123   |
| D                                                           | -0.168   | 0.846 | 0.120 | -1.40   | 0.163   |
| E                                                           | 0.038    | 1.039 | 0.127 | 0.30    | 0.764   |
| F                                                           | -0.064   | 0.938 | 0.120 | -0.54   | 0.590   |
| G                                                           | -0.215   | 0.806 | 0.152 | -1.42   | 0.155   |
| H                                                           | -0.190   | 0.827 | 0.125 | -1.52   | 0.129   |
| I                                                           | -0.125   | 0.882 | 0.120 | -1.04   | 0.298   |
| Miscellaneous second primary tumor                           |          |       |       |         |         |
| Age                                                          | 0.062    | 1.064 | 0.012 | 5.14    | 0.000   |
| Tumour size in mm                                            | -0.060   | 0.942 | 0.025 | -2.38   | 0.017   |
| T stage (T1 = 0)                                             |          |       |       |         |         |
| T2                                                          | 0.864    | 2.372 | 0.398 | 2.17    | 0.030   |
| N stage (N0 = 0)                                             |          |       |       |         |         |
| N1                                                          | -0.101   | 0.904 | 0.513 | -0.20   | 0.844   |
| Number of positive lymph nodes                               | 0.210    | 1.233 | 0.256 | 0.82    | 0.412   |
| Grade (1 = 0)                                                |          |       |       |         |         |
| 2                                                           | -0.255   | 0.775 | 0.280 | -0.91   | 0.362   |
| 3                                                           | 0.229    | 1.257 | 0.357 | 0.64    | 0.521   |
| Multifocality (No = 0)                                       |          |       |       |         |         |
| Yes                                                         | -0.110   | 0.896 | 0.353 | -0.31   | 0.755   |
| Outcome Characteristics | Estimate | HR  | SE  | z value | Pr(>|z|) |
|-------------------------|----------|-----|-----|---------|---------|
| Histological tumour type (Ductal = 0) |          |     |     |         |         |
| Lobular | 0.472 | 1.604 | 0.347 | 1.36 | 0.173 |
| Mixed | 0.952 | 2.591 | 0.412 | 2.31 | 0.021 |
| Other | 0.606 | 1.832 | 0.442 | 1.37 | 0.171 |
| Sublocalisation (Outer quadrants = 0) |          |     |     |         |         |
| Inner quadrants | 0.175 | 1.192 | 0.278 | 0.63 | 0.528 |
| Central parts | -0.122 | 0.885 | 0.443 | -0.27 | 0.784 |
| Overlapping lesions | 0.034 | 1.035 | 0.280 | 0.12 | 0.903 |
| Lateralisation (Left = 0) |          |     |     |         |         |
| Right | 0.032 | 1.033 | 0.218 | 0.15 | 0.882 |
| Estrogen status (Negative = 0) |          |     |     |         |         |
| Positive | 1.195 | 3.302 | 0.544 | 2.20 | 0.028 |
| Progesteron status (Negative = 0) |          |     |     |         |         |
| Positive | -0.406 | 0.666 | 0.269 | -1.51 | 0.131 |
| HER2 status (Negative = 0) |          |     |     |         |         |
| Unclear | -0.822 | 0.440 | 0.597 | -1.38 | 0.169 |
| Positive | -0.743 | 0.476 | 0.531 | -1.40 | 0.162 |
| Targeted therapy (trastuzumab) (No = 0) |          |     |     |         |         |
| Yes | 0.719 | 2.052 | 0.681 | 1.06 | 0.291 |
| Adjuvant systematic therapy (None = 0) |          |     |     |         |         |
| Chemotherapy | 1.474 | 4.366 | 0.575 | 2.56 | 0.010 |
| Endocrine therapy | -0.175 | 0.840 | 0.347 | -0.50 | 0.614 |
| Both | 0.274 | 1.316 | 0.475 | 0.58 | 0.564 |
| Axillary lymph node dissection (No = 0) |          |     |     |         |         |
| Yes | 0.155 | 1.168 | 0.301 | 0.52 | 0.606 |
| Social economic status (Low = 0) |          |     |     |         |         |
| Medium | -0.561 | 0.570 | 0.260 | -2.16 | 0.031 |
| High | -0.479 | 0.619 | 0.275 | -1.74 | 0.081 |
| Hospital type (General = 0) |          |     |     |         |         |
| Top-clinical | -0.084 | 0.919 | 0.251 | -0.33 | 0.738 |
| Academic | -0.040 | 0.961 | 0.458 | -0.09 | 0.931 |
| Hospital region (A = 0) |          |     |     |         |         |
| B | 0.135 | 1.145 | 0.467 | 0.29 | 0.772 |
| C | 0.238 | 1.269 | 0.459 | 0.52 | 0.604 |
| D | -0.052 | 0.949 | 0.413 | -0.13 | 0.899 |
| E | -0.339 | 0.713 | 0.517 | -0.65 | 0.513 |
| F | -0.108 | 0.898 | 0.424 | -0.25 | 0.799 |
| G | 0.092 | 1.096 | 0.503 | 0.18 | 0.855 |
| H | 0.192 | 1.212 | 0.414 | 0.47 | 0.642 |
| I | 0.486 | 1.626 | 0.378 | 1.29 | 0.199 |