Published in final edited form as:

Stat Biosci. 2017 December ; 9(2): 406–430. doi:10.1007/s12561-016-9162-z.

Censoring-robust estimation in observational survival studies: Assessing the relative effectiveness of vascular access type on patency among end-stage renal disease patients

Vinh Q. Nguyen¹, Daniel L. Gillen¹

Vinh Q. Nguyen: nguyenvq714@gmail.com

¹Department of Statistics, University of California, Irvine 92697-1250, USA

Abstract

The proportional hazards model is commonly used in observational studies to estimate and test a predefined measure of association between a variable of interest and the time to some event T . For example, it has been used to investigate the effect of vascular access type in patency among end-stage renal disease patients (Gibson et al., *J Vasc Surg* 34:694–700, 2001). The measure of association comes in the form of an adjusted hazard ratio as additional covariates are often included in the model to adjust for potential confounding. Despite its flexibility, the model comes with a rather strong assumption that is often not met in practice: a time-invariant effect of the covariates on the hazard function for T . When the proportional hazards assumption is violated, it is well known in the literature that the maximum partial likelihood estimator is consistent for a parameter that is dependent on the observed censoring distribution, leading to a quantity that is difficult to interpret and replicate as censoring is usually not of scientific concern and generally varies from study to study. Solutions have been proposed to remove the censoring dependence in the two-sample setting, but none has addressed the setting of multiple, possibly continuous, covariates. We propose a survival tree approach that identifies group-specific censoring based on adjustment covariates in the primary survival model that fits naturally into the theory developed for the two-sample case. With this methodology, we propose to draw inference on a predefined marginal adjusted hazard ratio that is valid and independent of censoring regardless of whether model assumptions hold.

Keywords

Censoring; Estimating equations; Model misspecification; Survival; Trees

[✉]Daniel L. Gillen, dgillen@uci.edu.

Electronic supplementary material The online version of this article (doi:10.1007/s12561-016-9162-z) contains supplementary material, which is available to authorized users.

Supplementary Materials

The Web Appendix referenced in Sects. 2.2 and 3 are available with this paper at the Statistics in Biosciences website.

1 Introduction

The proportional hazards model [1] is popular in the health sciences for the analysis of failure time data since it offers investigators a means to address scientific questions without assuming a full probability distribution on the event time and the ability to incorporate right-censored observations with ease. Although fully parametric assumptions are avoided, it nonetheless carries a particularly strong assumption: the presence of a time-invariant covariate effect on the hazard ratio scale.

The assumption is seldom met in the real world for the two-sample case and is even more susceptible to failure in the observational setting where multiple covariates are included in the model. To illustrate, consider the United States Renal Data System Dialysis Morbidity and Mortality Study Wave 2 as described in [2], hereafter referred to as the Vascular Access (VA) study. The VA study was an observational prospective study involving 4065 incident hemodialysis and peritoneal dialysis patients initiating dialysis in 1996 or early 1997 at 799 dialysis facilities across the United States. Of primary scientific interest in the VA study was an a priori test of the comparative effectiveness of three access types for chronic hemodialysis on the time to access revision among patients with end-stage renal disease. Besides this predictor of interest, it was also necessary to adjust for additional covariates in the model as the study was observational in nature.

To better understand the context of the study, end-stage renal disease (ESRD) is a condition where the filtration performed by the kidneys has been reduced to a point where life can no longer adequately be sustained. It is estimated that more than 300,000 persons in the United States have ESRD, and this number has been steadily rising over the past few decades. The standard of care for patients suffering from ESRD is renal replacement therapy in the form of dialysis or kidney transplantation. Hemodialysis is a technique for removing blood from the patient, cleansing toxins from the blood outside of the body, and then replacing the blood back into the patient. This process typically takes 3–4 h to complete and persons with ESRD typically undergo hemodialysis treatment three to four days a week. Given the frequency and duration of dialysis, it is infeasible to repeatedly insert a new access (needle) into the patient's vein at each dialysis visit. Such frequency would quickly result in irreparable damage to the vein eliminating a route to remove blood from the patient. As such, a “permanent” access is placed in the patient which remains there until either the access becomes clogged and inoperable or the patient stops dialysis (typically due to transplantation or death).

Despite improvements in permanent access technology, access failure and repair remains a major problem in the care of dialysis patients. Repeated interventions to maintain a working access exact an economic toll on the Medicare system (all the US citizens undergoing dialysis for treatment of ESRD are covered by Medicare), and the physical and emotional tolls on the patient are equally burdensome. Today, two main types of dialysis access are in use. The first, which has been in use the longest and is the cheapest to manufacture and easiest to insert in a patient, is the prosthetic graft. The other is the autogenous arteriovenous fistula (AVF) which can be placed in the patient in two different ways: standard attachment to a vein (SA fistula) or as a venous transposition (VT fistula). Venous transposition

placements require greater skill to place and longer time to mature, and are needed when veins are small and hard to find (such scenarios often occur in diabetics, smokers, and obese individuals).

To address this question, one might prespecify a proportional hazards model for the data and draw inference on the adjusted hazard ratios comparing groups with different access types; this is a common approach taken by many investigators with similar goals as reflected by the ubiquity of the Cox model in applied research [3]. However, as the crossing of the survival curves in Fig. 2 suggests, the proportional hazards assumption might not be reasonable when comparing different access types. One might attempt to incorporate a time-dependent covariate or a shift point for the hazard into the model to address the non-proportionality after observing such curves, but doing so would necessitate data-driven modeling that would inherently alter the preconceived question of interest expressed in the original model. In this setting, Type I error corresponding to testing the null hypothesis of no association between the covariate of interest and survival is likely to be inflated and the generalizability of the results may be questionable as the new model, and hence the scientific question, is chosen on the basis of the observed sample. On the other hand, basing inference on the misspecified model can lead to flawed results as [4] have shown that the regression coefficients in this model are consistent with parameters that depend upon the observed censoring distribution. As a result, point estimates from a misspecified proportional hazards model are influenced by the censoring pattern, a parameter which is usually not of scientific interest and may lead to results that can be difficult to replicate (censoring patterns change from study to study) and an interpretation that is not scientifically meaningful.

To address this deficiency [5] proposed an estimator based on the weighted score that removes the censoring dependence when censoring is covariate independent. These methods allow a marginal hazard ratio to be prespecified such that the resulting estimator from a modified estimating equation is robust to model misspecification to the extent that the resulting estimand is scientifically meaningful (weighted hazard ratio) and censoring independent; the resulting estimand is identical to that if a proportional hazards model was used for inference in the absence of censoring. That is, under non-proportional hazards, the target of inference is a weighted average of the time-varying coefficient, where the weights depend on the underlying distribution of T . To further elucidate the interpretation of the estimand considered by [5,6] derived an algebraic connection between the average regression effect estimated via their approach and a single parameter characterizing k -sample differences in a semiparametric model with random error belonging to the $\mathcal{GP}$ family of weighted log-rank statistics [7, Sect. 7.5]. Boyd et al. [8] extended the approach for covariate-dependent censoring in the two-sample setting. Their method consists of estimating the censoring distribution conditional on the binary group indicator, henceforth denoted as $S_C(\cdot|Z)$, and reweighing observations by the inverse of these estimates at different locations in the score equation. As with [5], the result is a censoring-robust estimator that is the solution to the weighted score equation. Inference can be based on the asymptotic distribution of the censoring-robust estimator where the weights are treated as fixed. When Z is a group indicator, estimation of $S_C(\cdot|Z)$ is straightforward by applying the left-continuous Kaplan–Meier estimator to the censoring times of each group. However, when multiple covariates (possibly taking on continuous values) are included in the model,

estimation of S_C is non-trivial unless a parametric or semiparametric model is assumed for the censoring time. In either case, these models are accompanied by assumptions that can fail and lead to consequences that may not be well understood, making them unattractive for our use.

This manuscript is concerned with censoring-robust estimation and inference for misspecified proportional hazards models involving multiple, possibly continuous, time-independent covariates as typically utilized in observational studies. The goal is to draw valid and censoring-robust inference on estimands that capture the scientific question of interest (marginal hazard ratios) regardless of whether the model assumptions hold. This is useful in a comparative effectiveness setting like the VA study as the hypothesis conveyed through a probability model is a priori specified. Post hoc modification of the model to conform with the observed data is not desired in order to control Type I error and preserve the reproducibility of the results. Many other studies have utilized the proportional hazards model to investigate access revision among hemodialysis patients [9–12]. In order to properly compare results across studies, one must, at the very least, obtain estimates that are not dependent on the censoring mechanism. In Sect. 2, we describe previously proposed censoring-robust methods in more detail and consider their asymptotic distributions for the two-sample case. We then outline a survival tree approach based on the work of [13] to identify group-specific censoring and illustrate how censoring-robust estimation and inference can be achieved in cases with multiple adjustment covariates. We evaluate our proposed methodology using simulation in Sect. 3 and apply the proposed methodology to the Vascular Access data in Sect. 4. We end with some concluding remarks in Sect. 5.

2 Methods

2.1 Two-Sample Censoring-Robust Estimation for the Proportional Hazards Model

Let $T > 0$ be a continuous-time random variable with hazard function

$$\lambda(t|Z) = \lambda_0(t) \exp(Z^T \beta),$$

where Z is a $p \times 1$ vector of covariates, β is a $p \times 1$ vector of regression coefficients, and λ_0 is a non-specified baseline hazard corresponding to $Z = 0$. Let $C > 0$ be a random censoring variable with distribution $F_C(\cdot|Z)$ where T is independent of C given Z , and let τ be the maximum follow-up time. For a sample of size n , let $\hat{\beta}$ denote the maximum partial likelihood estimator corresponding to the root of the estimating equation (derivative of the log partial likelihood)

$$U(\beta) = \sum_{i=1}^n \int_0^\tau \left[Z_i - \frac{\sum_{j=1}^n Y_j(t) Z_j \exp(Z_j^T \beta)}{\sum_{j=1}^n Y_j(t) \exp(Z_j^T \beta)} \right] dN_i(t),$$

where $Y_i(t) = I\{C_i \geq t, T_i \geq t\}$ and $N_i(t) = I\{T_i \leq t, T_i < C_i\}$.

When λ is misspecified [4] showed that, under mild regularity conditions, $\hat{\beta}$ is consistent to the root of

$$g(\beta) = \int_0^\tau E_Z \left[f_{T(t|Z)} S_C(t|Z) \left[Z - \frac{E_Z \{ Z \exp(Z^T \beta) S_T(t|Z) S_C(t|Z) \}}{E_Z \{ \exp(Z^T \beta) S_T(t|Z) S_C(t|Z) \}} \right] \right] dt,$$

where f_X and S_X , respectively, correspond to the true density and survival function of X ($X = C$ or T). The estimator $\hat{\beta}$ is consistent for a quantity that depends on the distribution of Z , the true distribution of the failure times, the maximal follow-up time τ , and the censoring distribution. To remove the censoring dependence [5] proposed weighting the summand of U by $W(t) = \hat{S}_T(t) / \sum_{i=1}^n Y_i(t)$ to estimate β (valid when censoring is covariate independent), whereas [8] proposed a censoring-robust estimator $\hat{\beta}_{CR}$ based on the root of the weighted estimating equation

$$U_{W(\beta)} = \sum_{i=1}^n \int_0^\tau \hat{W}(t|Z_i) \left[Z_i - \frac{S_W^{(1)}(\beta, t)}{S_W^{(0)}(\beta, t)} \right] dN_i(t),$$

where, for the two-sample case, $\hat{W}(t|Z_i) = 1/\hat{S}_C(t|Z_i)$, $\hat{S}_C(\cdot|Z_i)$ is the left-continuous Kaplan–Meier estimator of the censoring time for group Z_i , and where for $r = 0, 1, 2$,

$$S_W^{(r)}(\beta, t) = \frac{1}{n} \sum_{j=1}^n \hat{W}(t|Z_j) Y_j(t) Z_j^{\otimes r} \exp \left(Z_j^T \beta \right),$$

such that for a vector z , $z^{\otimes 0} = 1$, $z^{\otimes 1} = z$, and $z^{\otimes 2} = zz^T$. Provided that $\hat{S}_C(\cdot|Z_i)$ converges to $S_C(\cdot|Z_i)$ in probability, they showed that $\hat{\beta}_{CR}$ is consistent for β^* , the root of

$$g_W(\beta) = \int_0^\tau E_Z \left[f_{T(t|Z)} \left[Z - \frac{E_Z \{ Z \exp(Z^T \beta) S_T(t|Z) \}}{E_Z \{ \exp(Z^T \beta) S_T(t|Z) \}} \right] \right] dt,$$

and that $n^{1/2}(\hat{\beta}_{CR} - \beta^*)$ converges to a zero-mean Gaussian distribution with variance that can be estimated by $\hat{A}^{-1} \hat{B} \hat{A}^{-1}$, where

$$\hat{A} = -\frac{1}{n} \sum_{i=1}^n \int_0^\tau \hat{W}(t|Z_i) \left[\frac{S_W^{(2)}(\hat{\beta}_{CR}, t)}{S_W^{(0)}(\hat{\beta}_{CR}, t)} - \frac{S_W^{(1)}(\hat{\beta}_{CR}, t)^{\otimes 2}}{S_W^{(0)}(\hat{\beta}_{CR}, t)^2} \right] dN_i(t),$$

$$\hat{B} = \frac{1}{n} \sum_{i=1}^n \left\{ \int_0^\tau \hat{W}(t|Z_i) \left[Z_i(t) - \frac{S_W^{(1)}(\hat{\beta}_{CR}, t)}{S_W^{(0)}(\hat{\beta}_{CR}, t)} \right] dN_i(t) - \int_0^\tau \frac{\hat{W}(t|Z_i) Y_i(t) \exp\{Z_i(t)^T \hat{\beta}_{CR}\}}{S_W^{(0)}(\hat{\beta}_{CR}, t)} \left[Z_i(t) - \frac{S_W^{(1)}(\hat{\beta}_{CR}, t)}{S_W^{(0)}(\hat{\beta}_{CR}, t)} \right] d\hat{F}_W(t) \right\},$$

and $\hat{F}_W(t) = \hat{W}(t|Z_i) N_i(t)/n$. The vector β^* does not depend on the nuisance parameter S_C ; it corresponds to the regression coefficient vector from a Cox model if the proportional hazards relationship was correctly specified, and corresponds to the estimand of the maximum partial likelihood estimator for a misspecified Cox model in the absence of censoring.

2.2 Identifying Groups via Survival Trees

Censoring-robust estimation and inference in the proportional hazards model rely on consistent estimation of $S_C(\cdot|Z)$ and incorporating covariate-specific inverse probability of censoring weights into the score equation. When Z consists of group indicators, non-parametric estimation of S_C can be easily obtained by separating the observations into groups and applying the Kaplan–Meier estimator to each group. When Z consists of continuous variables, estimation of S_C is not straightforward. One could assume covariate-independent censoring, $S_C(t|Z) = S_C(t)$, as in [5] and estimate a single censoring distribution for the observed sample to be used as weights (equivalent to the weights used in [5]). However [8] showed that this does not remove the censoring dependence when censoring is dependent on Z . Thus, the key to extending previously developed censoring-robust methods is to obtain a consistent estimator for $S_C(\cdot|Z)$.

For our purpose, estimation of S_C represents a prediction problem that is not of direct scientific interest but is necessary to remove the dependence of the usual Cox estimator on the censoring distribution under a misspecified model. Our goal for this prediction problem is to devise a reasonable set of procedures that can be easily implemented in order to flexibly estimate S_C so as to be useful for the original inferential problem (scientific interest).

Based on our experiences, a parametric relationship of C on Z is unrealistic in most situations as censoring times often differ by groups, with groups usually defined based on low or high values of a certain variable or a certain covariate combination. For example, in the Vascular Access study, older subjects have higher mortality rates, and older patients are more susceptible to censoring due to their dependence on a caregiver, leading to shorter censoring times. The relationship between time to death and age may be modified by a patient's diabetic status. Thus, we prefer to not impose any strict assumptions on the relationship between C and Z , opting for a highly flexible relationship between adjustment variables and the probability of censoring.

To that end, we propose to discretize the relationship between Z and C by identifying clusters of observations based upon their covariate values that share a “similar” censoring distribution. That is, we wish to arrive at a function M_C that maps Z to a censoring group,

i.e., $M_C(Z) \rightarrow \{1, \dots, m\}$, where m is the number of censoring-specific groups. To identify clusters in a non-parametric fashion, we consider the survival tree approach of [13]. Although tree approaches identify relationships by dichotomizing covariates or taking subsets of values, they are flexible enough to capture parametric and non-parametric relationships given sufficient sample size at the cost of efficiency (more nodes). Another advantage of the tree-based approach is naturally incorporating interactions among covariates. While this may lead to less precise estimates of S_C at different time points if parametric relationships do exist, the tradeoff with flexibility is worthwhile in our view as it allows for the investigation of all possible relationships within a population (even though censoring is probably strongly influenced by only a small subset of variables).

Once censoring-specific groups are identified based on Z , the Kaplan–Meier estimator can be used to provide a non-parametric estimate of the censoring distribution for each group. After the weights are obtained, they can be used for robust estimation and inference as described in Sect. 2.1, the only difference being the estimation of m censoring distributions instead of two. The inferential algorithm is outlined as follows for the proportional hazards model.

1. Specify the *a priori* scientific model:

$$\lambda(t|Z) = \lambda_0(t) \exp(Z^T \beta).$$

2. Identify censoring-specific groups using survival trees:

$$M_C(Z) \rightarrow \{1, \dots, m\}.$$

3. Estimate $S_C(\cdot|M_C(Z))$ using the left-continuous Kaplan–Meier estimator for each group $(1, \dots, m)$:

$$\hat{S}_C(\cdot|M_C(Z)).$$

4. Plug in the inverse probability of censoring in the estimating equation to form a weighted estimating equation:

$$U_W(\beta) = \sum_{i=1}^n \int_0^\tau \hat{W}(t|Z_i) \left[Z_i - \frac{\sum_{j=1}^n \hat{W}(t|Z_j) Y_j(t) Z_j \exp(Z_j^T \beta)}{\sum_{j=1}^n \hat{W}(t|Z_j) Y_j(t) \exp(Z_j^T \beta)} \right] dN_i(t),$$

$$\text{where } \hat{W}(t|Z_i) = 1/\hat{S}_C(t|M_C(Z_i)).$$

5. Solve $U_W(\beta) = 0$ to obtain the censoring-robust estimator, $\hat{\beta}_{CR}$.
6. Inference follows from the asymptotic distribution of $\hat{\beta}_{CR}$ as stated in Sect. 2.1.

The target of inference is β^* , the root of g_W , which also corresponds to the estimand of the maximum partial likelihood estimator from a Cox model under zero censorship. The weighted estimating equation is required to estimate β^* since [4,14] have shown that the usual (naive) estimator from a Cox model is consistent for a quantity that depends on the

censoring distribution. Thus, the current research focuses on the nonparametric estimation of S_C in order to facilitate inference about β^* .

For step 3 of the inferential algorithm, we consider the survival tree approach of [13]. They proposed growing a tree by considering splits based on the partitioning of the covariate space where the response is the time to some event; in our specific setting, the response is the censoring time. A standardized two-sample statistic that measures “difference” between groups is computed for each potential split, where the largest statistic is chosen for the next split. Once a maximal-sized tree is grown, pruning takes place based on their goodness-of-split statistic. See the Web Appendix of the manuscript for a summary of their algorithm as implemented in the current context (step 2 of the previously described inferential algorithm).

At each iteration of growing the tree, we are faced with selecting the split that separates the data into the two most heterogeneous groups with respect to censoring time. However, heterogeneity is difficult to summarize and is even more difficult to compare using a single statistic. The role of the standardized splitting statistic is to rank potential splits based on its measure of difference, but each two-sample statistic captures heterogeneity in a different way. For example [13] illustrated their algorithm with the log-rank statistic since it is commonly used for testing the equality of two distributions and is well understood. The log-rank statistic is most powerful when the difference between the two groups can be characterized using a proportional hazards relationship and is generally powerful when the hazards are stochastically ordered. However, when hazards are stochastically ordered but are non-proportional, it is not clear how splits should be ranked as the heterogeneity is not captured in a way that is comparable or even transitive. It is even more problematic when hazards cross as the two groups are clearly different but may not be reflected using the log-rank statistic.

Thus, it is crucial to define what heterogeneity precisely means and to select a statistic that reflects this difference. We require a statistic that can capture heterogeneity in a variety of alternatives, not just under proportional hazards. Recall that the motivation behind a censoring-robust estimator is to be able to *a priori* specify the question of interest and infer validly once data are available even when the proportional hazards assumption fails. To this end, it would seem contradictory to implicitly assume or prefer proportional hazards when splitting censoring groups using the log-rank statistic. As described previously, the log-rank statistic is probably not an optimal candidate due to its drawbacks under non-proportional hazards. We thus require a splitting statistic that is able to detect deviations from the strong null of equal censoring over a wide range of alternatives. Moreover, the splitting statistic should capture differences across the support of observable censoring times with no preference to any particular time point or region.

Some previously proposed versatile testing procedures that may be considered as possible splitting statistics include the maximum or linear combinations of several weighted log-rank statistics proposed and investigated in [7, Sect. 7.5], [15,16]; Kolmogorov–Smirnov type or Renyi-type statistics explored in [17–19]; Cramer–von Mises type statistics explored in [20,21]; weighted Kaplan–Meier statistics proposed and explored in [22,23]; and statistics

that capture overall survival differences based on squared differences in the hazard or absolute differences in the survival curves explored in [24,25], respectively.

For the purpose of identifying group-specific censoring in this manuscript, we consider the KG^p statistic of [17]. For a time-to-event random variable T differentiated by two groups, the statistic is defined to be

$$KG^p = \frac{\sup_{t \geq 0} \int_0^t \hat{S}^{\rho} + 1/2 \left(\frac{Y_1 Y_2}{Y_1 + Y_2} \right)^{1/2} \left(\frac{dN_1}{Y_1} - \frac{dN_2}{Y_2} \right)}{\left[\int_0^\infty \hat{S}^{2\rho} + 1 \left(1 - \frac{\Delta N_1 + \Delta N_2 - 1}{Y_1 + Y_2 - 1} \right) \times I\{Y_1 Y_2 > 0\} \frac{d(N_1 + N_2)}{Y_1 + Y_2} \right]^{1/2}},$$

where the time arguments are dropped in the integrals for each function for simplicity, the subscripts indicate group, $\hat{S}$ is the pooled Kaplan–Meier estimator for the time of interest, and $\rho \geq 0$ is a parameter that affects power at different alternatives. This particular statistic utilizes the maximum difference in the hazard function between the two groups, is easy to compute, and does not depend on weights involving the “censoring” time (in our case, the failure time). Thus, it can detect heterogeneity when the difference between groups satisfies proportional hazards, when hazard functions are non-proportional but ordered, and when hazard functions cross. Based on the optimality results discussed in [17] and the simulation results of [15] that include a similar statistic, we believe that the KG^p will be quite versatile for our use. The current use of the statistic is not for formal inference, so the choice ρ is not critical; we used $\rho = 0$ in our subsequent simulation study.

To reiterate, once censoring groups are identified based on the algorithm of [13], as outlined in the Web Appendix of the manuscript, using the KG^p statistic as the splitting criteria, the censoring distribution could be estimated by each group using the Kaplan–Meier estimator. The weights can be inserted by analogy with the two-sample case, and estimation and inference follow directly.

3 Numerical Studies

3.1 Simulation Setup

In this section, we compare the performance of (1) the naive estimator $\hat{\beta}$ where censoring is not accounted for, (2) the censoring-robust estimator $\hat{\beta}_{CRC}$ where S_C is estimated from a Cox proportional hazards model with all covariates included, and (3) the censoring-robust estimator $\hat{\beta}_{CR}$ where S_C is non-parametrically estimated using the survival tree approach (our proposed methodology) using simulation for the continuous-time proportional hazards model as described in Sect. 2.2 at $n = 400, 800$, and 2000 . We evaluate the estimators based on bias, efficiency, mean squared error (MSE), and coverage probability of 95 % confidence intervals under different data-generating mechanisms and as the censoring distribution varies.

Suppose $Z_1 \in \{0, 1\}$ is the variable of interest. Then $\exp(\beta_1)$ will be the corresponding parameter and focus of concern in our evaluations. Due to a lack of an analytic expression for β_1^* from g_W , we take the “true” value of β_1^* to be the Monte Carlo average of $\hat{\beta}_1$ (naive

estimator) under censoring case 1 (no random censorship with administrative censoring/truncation at $\tau = 4$) for $n = 2000$. That is, we use the average of the maximum partial likelihood estimator for large samples in the absence of intermittent censoring (administrative censoring is required to keep the support the same across comparisons) when evaluating the three estimators.

We generate data for the null case according to

$$\lambda(t|Z) = \exp\{0 \times z_1 - 0.4 \times z_2 - 0.4 \times z_3\},$$

the non-null proportional hazards case according to

$$\lambda(t|Z) = \exp\{-0.4 \times z_1 - 0.4 \times z_2 - 0.4 \times z_3\},$$

and the non-proportional hazards case (late diverging hazard) according to

$$\lambda(t|Z) = 1 + 0.3 \times I\{t > 0.5, z_1 = 1\}.$$

For each data-generating scenario and for each sample size n , we study the properties of the three estimators as censoring varies. See Table 1 for a description of the censoring scenarios, which consists of a variety of cases: administrative censoring by truncation (case 1), covariate-independent censoring (case 2), censoring by grouping (cases 3–4), censoring by parametric relationships (cases 5–6), and crossing hazard censoring (case 7). We require administrative censoring at time $\tau = 4$ in all cases to keep the support of observable time constant. In most cases, censoring times are generated according to a power-function distribution with parameters (b, r) , where $0 < C < b$ and $r > 0$. When $r = 1$, C is the uniform distribution over $(0, b)$. When $r \rightarrow 0$ more probability is concentrated toward 0, and when $r \rightarrow \infty$ more probability is concentrated toward b . As typical of observational studies, we generate covariates that are correlated by design: $Z_1 \sim \text{Bernoulli}(0.5)$, $Z_2 \sim \text{power-function}(1, 1 + 2 \times I\{Z_1 = 1\})$, and $Z_3 \sim 2 \times Z_2 + \text{Normal}(0, 1)$. We replicate each data-generating scenario and censoring case 1000 times.

To find censoring-specific groups, we apply the algorithm described in the Web Appendix of the manuscript to the censoring times using the KG^0 statistic. The parameter ρ affects power at different alternatives; since we are not using the statistic in a formal testing framework, we choose $\rho = 0$ for simplicity. In our tree-building algorithm, we restrict each node to have at least 20 events (censored observations) to ensure enough information on the censoring times for each group; we are not comfortable saying two groups are different with respect to censoring time when each group has, e.g., 5 observations each. We use 5-fold cross-validation to select the optimally sized tree. The results of the simulation study are presented in Tables 2, 3, and 4 for the $n = 400, 800$, and 2000 scenarios, respectively.

3.2 Simulation Results

In the subsequent paragraphs, we focus primarily on the comparison of the naive estimator ($\hat{\beta}$) with the censoring-robust estimator based on survival trees ($\hat{\beta}_{\text{CR}}$). We defer the

comparison of the censoring-robust estimator based on the Cox proportional hazards model for censoring ($\hat{\beta}_{\text{CRC}}$) towards the end.

Under the null and PH (correct model specification) scenarios, we find that the naive estimator performs as expected: it is approximately unbiased and coverage is attained using the 95 % confidence intervals regardless of the censoring mechanism. In contrast, $\hat{\beta}_{\text{CRC}}$ and $\hat{\beta}_{\text{CR}}$ are nearly unbiased and coverage is near 95 % for all censoring mechanisms except for a few cases (4–7) where coverage is 90–94 % and 91–93 %, respectively. However, as the sample size grows, coverage improves to 93–95 % for $\hat{\beta}_{\text{CR}}$. As expected when the correct model is specified, we observe that the naive estimator is a more efficient estimator.

In the case that the model is misspecified (non-PH scenario), we observe higher variability for $\hat{\beta}_{\text{CRC}}$ and $\hat{\beta}_{\text{CR}}$ as compared to $\hat{\beta}$. However, the comparison of variability with $\hat{\beta}$ is inappropriate since the estimators are estimating different quantities; this is akin to comparing apples and oranges. In a misspecified setting, the estimand for the naive estimator is a quantity that depends on censoring and the estimand for the censoring-robust estimator is a marginal adjusted hazard ratio that is independent of censoring, a quantity that can be meaningful to an investigator in the presence of a time-varying effect: a weighted average of hazard ratios. As can be seen, the naive estimator is severely biased and coverage can decrease to 0, both of which illustrate that what $\hat{\beta}$ is consistent for is a function of censoring. On the other hand, $\hat{\beta}_{\text{CR}}$ is nearly unbiased and coverage is attained or nearly attained in all scenarios.

The improvement in unbiasedness and coverage probability of $\hat{\beta}_{\text{CR}}$ might seem like a direct result of a bias–variance tradeoff induced by the weighted estimating equation. If we consider MSE, then $\hat{\beta}$ dominates for the null and proportional hazards settings under all censoring scenarios; this is expected as the model was correctly specified. However, for the non-proportional hazards setting, we find that the MSE of $\hat{\beta}_{\text{CR}}$ dominates the MSE of $\hat{\beta}$ in all censoring scenarios. We note, however, that MSE is not an ideal summary measure for our comparisons as the estimators are consistent for different quantities. If we let $n \rightarrow \infty$, then the variance component of MSE goes to 0 for all estimators, and what remains is the bias (consistency). In our case, the bias reflects the varying estimand for varying censoring mechanisms.

If we consider coverage probability in more detail, we can explain the lack of nominal coverage attainment by either of two reasons: either the estimator was not estimating the estimand of interest (e.g., $\hat{\beta}$ in the non-proportional hazards setting) or that the variability of the estimator was not adequately estimated (e.g., $\hat{\beta}_{\text{CR}}$ as reflected by the average ratio of analytic standard error to empirical standard error, SE/ESE). As inference of β^* is based on the asymptotic distribution of $\hat{\beta}_{\text{CR}}$, coverage probability improves as $n \rightarrow \infty$. A second reason might be due to the fact that the variability was induced by the survival tree algorithm and the estimation of S_C was not incorporated in the variance formula. That is, the weights in U_W were treated as fixed. To incorporate this source of variability, one can easily incorporate the bootstrap to obtain a more correct estimate of the standard error. However, as Fig. 1 portrays, not much is lost when this source of variability is ignored. In most cases, the average ratio of analytic standard error to bootstrap standard error is about 0.95–1.00, with

the worst-case scenario being 0.914 for the proportional hazards setting with censoring scenario 6 at $n = 400$. With respect to coverage probability, the use of bootstrap standard errors make negligible difference, with the biggest difference being a 1 % improvement in the coverage probability.

When considering the average number of censoring groups obtained from the survival tree algorithm, we find that more groups are identified when there is a parametric relationship between the covariates and the censoring time (cases 5–6). We also observe that as the sample size is increased, the algorithm yields more groups. Both observations are expected for the survival tree algorithm. We also note that the KG^0 statistic works reasonably well when hazards cross (case 7).

If we compare the results of $\hat{\beta}_{\text{CRC}}$ with those of $\hat{\beta}_{\text{CR}}$ from Tables 2, 3, and 4, we find that in nearly all settings, the survival tree approach ($\hat{\beta}_{\text{CR}}$) performs better with respect to bias, efficiency, MSE, and coverage probability. It is noteworthy that censoring scenario 4 leads to extremely large values for MSE. This is probably due to the fact that the Cox model for the censoring time is estimating extremely small probabilities caused by large or small covariate values combined with the linear relationship on the hazard ratio scale. Small probabilities of not-yet-censored would lead to large weights and numerical instability with the estimates. This probably was not a problem for the survival tree approach since we restricted at least 20 censoring events for each node, reducing the possibility for small probability of not-yet-censored.

Finally, we were not able to manually build good predictive models for $C|Z$ using the proportional hazards model to obtain $\hat{\beta}_{\text{CRC}}$ for each dataset generated in the simulation study. Since only three covariates were part of the scientific model in our simulation study, we included all three covariates for the distribution of $C|Z$. As we have described in our inferential algorithm in Sect. 2.2, any prediction method could be used to estimate $S_C(\cdot|Z)$. If one were to use a (semi)parametric model, then transformations or discretizations of the covariates and interactions should be considered to obtain better models. The possibilities are limitless. However, exploring all functional forms and interactions is nearly equivalent to the survival tree approach, where covariates are repeatedly discretized and interactions are inherently incorporated due the recursive nature of the tree approach. Thus, a survival tree approach may be preferred.

4 Application

Recall the Vascular Access study introduced in Sect. 1. In this section, we focus on the subgroup of patients ($n = 1542$) receiving hemodialysis; see Table 5 for a description of the sample. A scientific goal of the study was to compare the effect of different access types on the time to access revision (863 events). As noted before, using a proportional hazards model can lead to estimates that depend upon the censoring distribution since the data indicate a failure of the proportional hazards assumption (see Fig. 2), potentially leading to results that are difficult to replicate across multiple studies. This is of particular relevance in the area of comparative effectiveness where one would seek to determine if the relative performance of access type is consistent across multiple studies and patient populations. We

thus apply the censoring-robust methods outlined in Sect. 2.2 to obtain censoring-specific groups and apply the results outlined in Sect. 2.1 to obtain estimates and draw inference.

After applying the survival tree approach using the same restrictions on the nodes and number of cross-validations as in the simulation study, we arrive at 24 censoring groups, with ethnicity being the first covariate that a split occurs on. That is, there is heterogeneity in censoring time across ethnicity. The censoring curves for each node of the pruned tree can be seen in Fig. 3. Estimates and 95 % confidence intervals for the scientific model are presented in Table 6; we include results from both the naive estimator and the censoring-robust estimator. Clearly, the point estimates differ by a noticeable amount. The marginal adjusted hazard ratio comparing VT fistula to SA fistula is estimated to be 2.08 (95 % CI 1.449–2.985). The marginal adjusted hazard ratio comparing prosthetic graft to SA fistula is 1.640 (95 % CI 1.234–2.179). Thus, VT fistula, the most complicated access type for insertion, is associated with longer time to access revision.

One might argue that the same conclusions would have been obtained using the naive estimator. However, the estimates are clearly different, and if the study were repeated, it would not be surprising if the naive estimator resulted in noticeably different estimates. That is, even if the survival times remained the same in a new study but that a different censoring mechanism is utilized, the results from the naive estimator would change, whereas the results based on the censoring-robust estimator would not.

When real effects are near a hazard ratio of 1, the impact might lead to a different conclusion, from statistical significance to non-significance (or vice versa). For example, consider the estimate for gender. The naive estimator yields a censoring-dependent marginal adjusted hazard ratio of 1.150 (95 % CI 0.991–1.335), and the censoring-robust estimator yields a marginal adjusted hazard ratio of 1.248 (95 % CI 1.012–1.506). The important fact that should not be forgotten is that censoring is a nuisance that should be removed in order for results to bear scientific meaning.

5 Discussion

In this manuscript, we outlined a set of procedures that allow for robust estimation and inference of a predefined marginal adjusted hazard ratio using a proportional hazards model involving multiple covariates that is commonly used for observational studies. The methodology is useful in scientific practice since the hypothesis (expressed using a measure of association from a probability model) is defined a priori using a prespecified probability model. Since the model is chosen before data analysis, the adequacy of model assumptions is often not met. Modifying the model to conform with assumptions alters the preconceived hypothesis, potentially leading to an inflation of Type I error and results that may be difficult to replicate. The proposed methodology facilitates the estimation of a marginal adjusted hazard ratio that is independent of censoring when the assumption of a time-invariant covariate effect is violated. In the case that the assumption does in fact hold, the methodology still yields a consistent estimate of the fixed adjusted hazard ratio. The tradeoff for robustness is a cost in the efficiency of the estimator when model assumptions do in fact hold. However, we believe that the cost for robustness is valuable in the scientific setting

where prespecification of the hypothesis is required and model assumptions are not known to hold at the design phase.

The extension from the two-sample case to the observational study case is not immediately obvious when S_C depends on multiple covariates as the relationship between C and Z is unknown. We treat the estimation of S_C as a data-driven prediction problem since it is not of direct scientific interest but its accuracy is crucial in removing the nuisance censoring dependence from the estimand of interest. As seen in Sect. 3, the algorithm of [13] combined with the KG^P statistic is sufficiently flexible to identify censoring-specific groups under a wide variety of censoring scenarios. Once the groups are identified, extension to the multiple groups case is trivial using developed theory for the two-sample case. The novelty of the method presented in this manuscript is the use of survival trees on the censoring time that facilitates the estimation goal of a scientific study. To draw proper inference, one could employ the bootstrap procedure to account for the variability from the tree-building algorithm and the estimation of S_C . However, as our simulation results have shown, little is lost if this source of variability is unaccounted for.

Our proposed procedures can be modified in a number of ways. Any method of prediction for S_C is a possible candidate. However, based on our experiences with censoring, we believe that censoring usually differs by the clustering of a small subset of covariates. Hence, the tree approach seems natural. Within the tree framework, a user could choose any other reasonable two-sample statistic that detects heterogeneity. We understand that there is often a time constraint on any given data analysis. Should computing be an issue, the use of readily available algorithms such as the tree algorithm based on an exponential model [26] is probably reasonable in most situations. However, the flexibility in detecting heterogeneity in cases where the hazard functions cross might be lost.

Users should consider placing restrictions on nodes to have more confidence in the splitting of nodes. Examples of restrictions include the number of events or the number of observations. The illustrations in Sects. 3 and 4 require 20 censored events in each node. This number should be dictated by the comfort level of the user. Also, instead of using cross-validation to select an optimal tree that takes into account the tradeoff of complexity and overfitting, the users can predefine the number of nodes that they are willing to accept. For example, four or five nodes are probably more than enough to account for most of the variability arising from differential censoring in most situations.

We have also illustrated the estimation of S_C with a Cox proportional hazards model in our simulation study and found that it performs worse than the survival tree approach. We do not discount this method immediately as our simulation study did not optimize the CZ model with respect to prediction. In practice where one spends more time on predicting S_C , results may be improved. However, such an approach would require manual human intervention and guidance in building the model to explore different transformations, combinations, and interactions. In this context, the survival tree approach is attractive due to its automatic nature and inherent design to explore interactions and functional forms (via discretization of the covariates).

We note that the proposed estimation procedure has particular relevance in health policy where a great deal of research and emphasis is being placed on the comparative effectiveness of existing interventions. Such analyses are often carried out in a meta-analysis framework where the relative performance of existing interventions are compared and contrasted across multiple observational and interventional studies. Along with the usual complications involved in performing meta-analyses (accounting for heterogeneity of data collection procedures, different adjustment covariates, different patient populations, publication bias, etc.), most studies will tend to have differential patient accrual and dropout patterns. The result could be perceived effect modification that is solely driven by differences in the censoring distribution across studies. The methods proposed here could alleviate this problem if each individual analysis reported a censoring-robust estimate of intervention effect. For example, other authors have also studied vascular access patency [9–12]. If we suppose that all inclusion and exclusion criteria and adjustment variables were identical in these studies, then the individual (naively) estimated hazard ratios cannot be compared as estimands are most likely different due to differential censoring. Only if censoring were constant could a fair comparison be made. For example, if a censoring-robust estimator were used, then an adjusted marginal hazards ratio under zero censorship could be used for comparison. Further utility of censoring-robust estimators in the context of meta-analyses remains an area of future research.

One limitation to our methodology is that it does not extend to the time-varying covariates setting as it is unclear how the sample could be partitioned when each “subject” experiences different covariate values at different times. In such a setting, building a parametric model selected by cross-validation or validated by a hold-out sample seems like a more pragmatic approach.

Finally, the methodology proposed in this manuscript can easily be extended for censoring-robust estimation in the discrete survival setting as described by [14]. Again, survival trees can be used to identify censoring-specific groups, and weights can be incorporated into the estimating equation as in the two-sample case.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Gibson KD, Gillen DL, Caps MT, Kohler TR, Sherrard DJ, Stehman-Breen CO. Vascular access survival and incidence of revisions: A comparison of prosthetic grafts, simple autogenous fistulas, and venous transposition fistulas from the United States Renal Data System Dialysis Morbidity and Mortality Study* 1. *J Vasc Surg.* 2001; 34(4):694–700. [PubMed: 11668326]
2. Cox DR. Regression models and life-tables. *J R Stat Soc Ser B (Methodol).* 1972; 34(2):187–220.
3. Stigler SM. Citation patterns in the journals of statistics and probability. *Stat Sci.* 1994; 9(1):94–108.
4. Struthers CA, Kalbfleisch JD. Misspecified proportional hazard models. *Biometrika.* 1986; 73(2):363–369.
5. Xu R, O’Quigley J. Estimating average regression effect under non-proportional hazards. *Biostatistics.* 2000; 1(4):423–439. [PubMed: 12933565]

6. Xu R, Harrington DP. A semiparametric estimate of treatment effects with censored data. *Biometrics*. 2001; 57(3):875–885. [PubMed: 11550940]
7. Fleming TR, Harrington DP. Counting processes and survival analysis. Applied Probability and Statistics Section, Wiley Series in Probability and Mathematical Statistics. 1991
8. Boyd AP, Kittelson JM, Gillen DL. Estimation of treatment effect under non-proportional hazards and conditionally independent censoring. *Stat Med*. 2012; 31(28):3504–3515. [PubMed: 22763957]
9. Woods JD, Turenne MN, Strawderman RL, Young EW, Hirth RA, Port FK, Held PJ. Vascular access survival among incident hemodialysis patients in the united states. *Am J Kidney Diseases*. 1997; 30(1):50–57. [PubMed: 9214401]
10. Dixon BS, Novak L, Fangman J, et al. Hemodialysis vascular access survival: upper-arm native arteriovenous fistula. *Am J Kidney Diseases*. 2002; 39(1):92. [PubMed: 11774107]
11. Sheth RD, Brandt ML, Brewer ED, Nuchtern JG, Kale AS, Goldstein SL. Permanent hemodialysis vascular access survival in children and adolescents with end-stage renal disease. *Kidney Int*. 2002; 62(5):1864–1869. [PubMed: 12371990]
12. Ramage IJ, Bailie A, Tyerman KS, McColl JH, Pollard SG, Fitzpatrick MM, et al. Vascular access survival in children and young adults receiving long-term hemodialysis. *Am J Kidney Diseases*. 2005; 45(4):708. [PubMed: 15806474]
13. LeBlanc M, Crowley J. Survival trees by goodness of split. *J Am Stat Assoc*. 1993; 88(422):457–467.
14. Nguyen VQ, Gillen DL. Robust inference in discrete hazard models for randomized clinical trials. *Lifetime Data Anal*. 2012; 18(4):446–469. [PubMed: 22810273]
15. Lee JW. Some versatile tests based on the simultaneous use of weighted log-rank statistics. *Biometrics*. 1996; 52(2):721–725.
16. Lang, Wu; Peter, B. Gilbert. Flexible weighted log-rank tests optimal for detecting early and/or late survival differences. *Biometrics*. 2002; 58(4):997–1004. [PubMed: 12495155]
17. Fleming TR, Harrington DP, O’Sullivan M. Supremum versions of the log-rank and generalized Wilcoxon statistics. *J Am Stat Assoc*. 1987; 82(397):312–320.
18. Fleming TR, O’Fallon JR, O’Brien PC, Harrington DP. Modified Kolmogorov-Smirnov test procedures with application to arbitrarily right-censored data. *Biometrics*. 1980; 36(4):607–625.
19. Fleming TR, Harrington DP. A class of hypothesis tests for one and two samples of censored survival data. *Commun Stat Theory Meth*. 1981; 10:763–794.
20. Schumacher M. Two-sample tests of Cramer–von Mises- and Kolmogorov–Smirnov-type for randomly censored data. *Int Stat Rev*. 1984; 52(3):263–281.
21. Koziol JA. A two sample Cramer-von Mises test for randomly censored data. *Biom J*. 1978; 20(6):603–608.
22. Pepe MS, Fleming TR. Weighted kaplan–Meier statistics: a class of distance tests for censored survival data. *Biometrics*. 1989; 45(2):497–507. [PubMed: 2765634]
23. Pepe MS, Fleming TR. Weighted Kaplan-Meier statistics: large sample and optimality considerations. *J R Stat Soc Ser B (Methodol)*. 1991; 53(2):341–352.
24. Lin X, Wang H. A new testing approach for comparing the overall homogeneity of survival curves. *Biom J*. 2004; 46(5):489–496.
25. Lin X, Xu Q. A new method for the comparison of survival distributions. *Pharm Stat*. 2009
26. LeBlanc M, Crowley J. Relative risk trees for censored survival data. *Biometrics*. 1992; 48(2):411–425. [PubMed: 1637970]

Fig. 1. Bootstrap standard error of $\hat{\beta}_{CR}$ compared with analytic standard error not accounting for the variability induced by the estimation of $\hat{S}_C$. For each scenario, the procedure was repeated 600 times (as opposed to 1000 replicates), where at each iteration, 200 bootstrap samples were drawn. The symbol *unfilled square* corresponds to analytic standard error, and *unfilled triangle* corresponds to the bootstrap standard error

SA fistula	401 (0)	234 (99)	35 (170)	0 (183)
VT fistula	111 (0)	44 (42)	2 (62)	0 (63)
Prosthetic graft	1030 (0)	611 (265)	44 (594)	1 (617)
Total	1542 (0)	889 (406)	81 (826)	1 (863)

Fig. 2. Estimated survival curves depicting the probability of access survival for three access types (SA fistula, VT fistula, and prosthetic graft). Statistics below the x axis depict the number of accesses at risk and the cumulative number of access failures

Censoring Curves

Fig. 3.
Estimated censoring distributions for the groups identified by the CART procedure

Table 1

Censoring scenarios

Cases	Censoring
1	$C = 4$
2	$C \sim \text{power-function}(4, 1)$
3	$C \sim \text{power-function}(4, 1 + 0.5 \times I\{z_1 = 1, z_2 = 0.5\})$
4	$C \sim \text{power-function}(4, 1 + 2.0 \times I\{z_1 = 1, z_2 = 0.5\})$
5	$C \sim \text{power-function}(4, \exp\{-z_2\})$
6	$C \sim \text{power-function}(4, \exp\{-0.5 \times z_1 - 0.5 \times z_2 - z_1 \times z_2\})$
7	$Z_1 = 0 : \lambda_C(t Z) = 1 \times I\{t = 0.5\} + 0.360 \times I\{0.5 < t \leq 1\} + 1 \times I\{1 < t \leq 4\}$ $Z_1 = 1 : \lambda_C(t Z) = 0.516 \times I\{t = 0.5\} + 1 \times I\{0.5 < t \leq 4\}$

Scenario 1 is that of censoring by truncation. Scenario 2 is that of a single censoring mechanism. Scenarios 3–4 are that of two censoring mechanisms dictated by the covariates. Scenarios 5–6 have censoring dependent on Z in a parametric relationship. Scenario 7 is that of crossing hazards

Table 2

Results of the simulation study with $n = 400$

Event	Naive estimator ($\hat{\beta}$)						Robust estimator with Cox PH Censoring ($\hat{\beta}_{CR}$)						Robust estimator with Survival Trees ($\hat{\beta}_{CR}$)						Avg
Scenario	Rate	Truth	Bias	ESE	MSE	SE/ESE	CP	Bias	ESE	MSE	SE/ESE	CP	Bias	ESE	MSE	SE/ESE	CP	Groups	
$n = 400$																			
Null																			
C: 1	0.819	0.000	0.005	0.120	1.449	1.042	0.965	-0.003	0.132	1.754	0.945	0.928	0.005	0.120	1.449	1.042	0.965	1.000	
C: 2	0.549	0.000	0.006	0.152	2.327	1.015	0.962	0.001	0.199	3.959	0.922	0.934	0.004	0.185	3.418	0.954	0.949	5.001	
C: 3	0.606	0.000	0.005	0.140	1.957	1.037	0.959	-0.014	0.363	13.166	0.586	0.915	0.002	0.164	2.681	0.984	0.938	4.319	
C: 4	0.678	0.000	0.007	0.133	1.759	1.021	0.956	-0.024	1.237	152.934	0.334	0.824	0.005	0.145	2.110	1.011	0.951	3.566	
C: 5	0.416	0.000	0.005	0.181	3.259	1.007	0.944	-0.006	0.190	3.611	0.995	0.948	0.002	0.221	4.859	0.982	0.943	6.449	
C: 6	0.374	0.000	0.016	0.205	4.206	1.014	0.952	0.002	0.226	5.104	0.935	0.937	0.026	0.277	7.734	0.907	0.930	8.321	
C: 7	0.380	0.000	0.002	0.185	3.435	1.017	0.952	-0.008	0.308	9.491	0.838	0.903	0.002	0.262	6.858	0.931	0.934	6.760	
PH																			
C: 1	0.766	-0.404	-0.004	0.135	1.828	0.977	0.944	-0.002	0.132	1.739	0.998	0.952	-0.004	0.135	1.828	0.977	0.944	1.000	
C: 2	0.493	-0.404	-0.003	0.168	2.827	0.985	0.948	0.004	0.231	5.324	0.877	0.939	-0.005	0.206	4.221	0.939	0.943	5.484	
C: 3	0.550	-0.404	-0.004	0.158	2.482	0.983	0.949	-0.002	0.300	9.002	0.738	0.921	-0.007	0.186	3.444	0.943	0.937	4.984	
C: 4	0.619	-0.404	-0.002	0.150	2.244	0.960	0.944	-0.047	1.166	136.049	0.335	0.839	-0.004	0.168	2.823	0.933	0.931	4.464	
C: 5	0.371	-0.404	-0.012	0.193	3.740	1.008	0.947	-0.001	0.212	4.500	0.968	0.936	-0.014	0.262	6.855	0.903	0.923	6.810	
C: 6	0.341	-0.404	-0.007	0.233	5.421	0.961	0.934	-0.002	0.232	5.395	0.987	0.944	-0.007	0.320	10.223	0.874	0.918	8.811	
C: 7	0.333	-0.404	0.000	0.204	4.145	0.994	0.952	-0.004	0.339	11.491	0.848	0.919	-0.002	0.299	8.947	0.890	0.916	7.223	
Non-PH																			
C: 1	0.885	-0.647	-0.003	0.124	1.536	1.002	0.944	0.009	0.122	1.504	1.013	0.955	-0.003	0.124	1.536	1.002	0.944	1.000	
C: 2	0.664	-0.647	0.133	0.144	3.850	0.994	0.842	0.001	0.177	3.118	0.886	0.908	0.011	0.158	2.506	0.977	0.944	3.617	
C: 3	0.721	-0.647	0.095	0.138	2.821	0.992	0.898	-0.032	0.204	4.253	0.795	0.924	0.003	0.146	2.144	0.985	0.939	3.029	
C: 4	0.784	-0.647	0.045	0.131	1.905	1.002	0.937	-0.110	1.089	119.679	0.210	0.908	-0.009	0.136	1.866	0.988	0.941	2.377	
C: 5	0.508	-0.647	0.202	0.163	6.719	1.001	0.763	0.106	0.173	4.121	0.971	0.898	0.023	0.190	3.641	0.971	0.942	5.642	
C: 6	0.485	-0.647	0.202	0.194	7.855	0.995	0.810	0.163	0.196	6.496	1.001	0.863	0.027	0.247	6.164	0.923	0.924	6.836	
C: 7	0.500	-0.647	0.263	0.165	9.638	0.983	0.615	-0.004	0.270	7.283	0.793	0.901	0.029	0.202	4.161	0.966	0.941	5.571	

Each scenario is repeated 1000 times. Event rate refers to the average proportion of the sample with an observed event. Truth refers to the Monte Carlo average of $\hat{\beta}$ at $n = 2000$ and under censoring scenario 1. ESE refers to the empirical standard error as calculated through the replicated datasets. MSE refers to the mean squared error times 1000. SE/ESE refers to the average ratio of analytically calculated

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

sandwich standard error to the ESE. CP refers to the coverage probability of the 95 % confidence intervals. Avg Groups refer to the average number of censoring-specific groups derived from the chosen tree used to obtain $\hat{\beta}_{CR}$

Table 3

Results of the simulation study with $n = 800$

Event	Naive estimator ($\hat{\beta}$)							Robust estimator with Cox PH Censoring ($\hat{\beta}_{CR}$)							Robust estimator with Survival Trees ($\hat{\beta}_{CR}$)							Avg
Scenario	Rate	Truth	Bias	ESE	MSE	SE/ESE	CP	Bias	ESE	MSE	SE/ESE	CP	Bias	ESE	MSE	SE/ESE	CP	Groups				
$n = 800$																						
Null																						
C: 1	0.819	0.000	-0.003	0.090	0.812	0.983	0.937	0.000	0.092	0.851	0.958	0.941	-0.003	0.090	0.812	0.983	0.937	1.000				
C: 2	0.547	0.000	-0.001	0.109	1.189	1.001	0.952	0.002	0.152	2.310	0.885	0.933	-0.006	0.136	1.856	0.948	0.940	9.378				
C: 3	0.606	0.000	-0.002	0.101	1.026	1.010	0.943	-0.009	0.193	3.712	0.820	0.931	-0.002	0.118	1.393	0.997	0.949	7.888				
C: 4	0.677	0.000	-0.003	0.096	0.913	1.001	0.948	-0.060	1.353	183.228	0.307	0.842	-0.008	0.107	1.154	0.981	0.944	7.599				
C: 5	0.415	0.000	-0.001	0.132	1.733	0.973	0.945	-0.001	0.138	1.908	0.973	0.948	-0.003	0.171	2.919	0.930	0.923	13.725				
C: 6	0.374	0.000	-0.002	0.148	2.185	0.990	0.951	0.003	0.149	2.226	1.006	0.946	0.002	0.197	3.875	0.953	0.941	17.677				
C: 7	0.380	0.000	0.002	0.131	1.719	1.011	0.956	0.007	0.215	4.607	0.914	0.944	0.006	0.193	3.740	0.941	0.944	12.784				
PH																						
C: 1	0.767	-0.404	0.001	0.093	0.873	0.996	0.947	0.006	0.096	0.924	0.968	0.935	0.001	0.093	0.873	0.996	0.947	1.000				
C: 2	0.494	-0.404	-0.004	0.120	1.451	0.968	0.942	0.004	0.162	2.613	0.901	0.933	-0.007	0.152	2.322	0.922	0.932	10.281				
C: 3	0.548	-0.404	0.001	0.113	1.265	0.970	0.933	0.003	0.218	4.748	0.761	0.931	0.001	0.137	1.881	0.935	0.944	9.250				
C: 4	0.618	-0.404	0.001	0.103	1.064	0.984	0.944	-0.051	1.055	111.532	0.371	0.859	0.001	0.124	1.531	0.930	0.939	8.751				
C: 5	0.370	-0.404	-0.003	0.138	1.895	0.995	0.956	0.008	0.149	2.212	0.981	0.945	-0.008	0.180	3.241	0.967	0.945	14.329				
C: 6	0.342	-0.404	-0.004	0.156	2.435	1.008	0.954	0.002	0.167	2.770	0.975	0.938	-0.009	0.208	4.327	0.993	0.943	18.525				
C: 7	0.332	-0.404	0.005	0.144	2.081	0.989	0.942	-0.001	0.248	6.128	0.879	0.931	-0.001	0.216	4.670	0.927	0.930	14.440				
Non-PH																						
C: 1	0.885	-0.647	-0.002	0.086	0.740	1.017	0.952	-0.004	0.085	0.731	1.025	0.952	-0.002	0.086	0.740	1.017	0.952	1.000				
C: 2	0.664	-0.647	0.132	0.100	2.725	1.013	0.739	-0.003	0.122	1.495	0.925	0.932	0.003	0.112	1.260	0.985	0.941	6.989				
C: 3	0.723	-0.647	0.095	0.094	1.799	1.023	0.839	-0.034	0.136	1.969	0.858	0.936	0.001	0.102	1.043	1.011	0.956	5.962				
C: 4	0.784	-0.647	0.047	0.090	1.036	1.021	0.912	-0.035	0.344	11.954	0.472	0.906	-0.010	0.096	0.933	1.002	0.951	5.071				
C: 5	0.508	-0.647	0.204	0.115	5.487	1.003	0.568	0.095	0.127	2.509	0.942	0.858	0.022	0.135	1.870	0.988	0.945	11.260				
C: 6	0.486	-0.647	0.208	0.139	6.244	0.977	0.650	0.153	0.136	4.188	1.016	0.802	0.041	0.178	3.322	0.930	0.925	14.517				
C: 7	0.500	-0.647	0.261	0.111	8.049	1.024	0.363	-0.032	0.190	3.725	0.836	0.895	0.025	0.143	2.113	1.000	0.946	9.887				

Each scenario is repeated 1000 times. Event rate refers to the average proportion of the sample with an observed event. Truth refers to the Monte Carlo average of $\hat{\beta}$ at $n = 2000$ and under censoring scenario 1. ESE refers to the empirical standard error as calculated through the replicated datasets. MSE refers to the mean squared error times 1000. SE/ESE refers to the average ratio of analytically calculated

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

sandwich standard error to the ESE. CP refers to the coverage probability of the 95 % confidence intervals. Avg Groups refer to the average number of censoring-specific groups derived from the chosen tree used to obtain $\hat{\beta}_{CR}$

Table 4

Results of the simulation study with $n = 2000$

Event	Naive estimator ($\hat{\beta}$)							Robust estimator with Cox PH censoring ($\hat{\beta}_{CRC}$)							Robust estimator with Survival Trees ($\hat{\beta}_{CR}$)							Avg		
Scenario	Rate	Truth	Bias	ESE	MSE	SE/ESE	CP	Bias	ESE	MSE	SE/ESE	CP	Bias	ESE	MSE	SE/ESE	CP	Bias	ESE	MSE	SE/ESE	CP	Groups	
$n = 2000$																								
Null																								
C: 1	0.819	0.000	-0.001	0.055	0.298	1.024	0.954	-0.001	0.056	0.313	1.000	0.957	-0.001	0.055	0.298	1.024	0.954	-0.001	0.055	0.298	1.024	0.954	1.000	
C: 2	0.548	0.000	-0.002	0.069	0.472	1.003	0.957	-0.001	0.101	1.010	0.874	0.914	-0.003	0.086	0.733	0.979	0.943	-0.003	0.086	0.733	0.979	0.943	27.084	
C: 3	0.606	0.000	-0.001	0.062	0.389	1.035	0.963	0.010	0.156	2.441	0.754	0.926	-0.001	0.077	0.595	0.994	0.941	-0.001	0.077	0.595	0.994	0.941	22.101	
C: 4	0.678	0.000	-0.001	0.059	0.346	1.025	0.951	0.041	1.143	130.607	0.389	0.831	-0.003	0.067	0.451	1.008	0.945	-0.003	0.067	0.451	1.008	0.945	19.191	
C: 5	0.416	0.000	-0.000	0.081	0.654	1.000	0.953	-0.001	0.087	0.759	0.984	0.940	-0.001	0.106	1.118	0.976	0.947	-0.001	0.106	1.118	0.976	0.947	40.183	
C: 6	0.374	0.000	-0.001	0.090	0.806	1.026	0.952	-0.001	0.098	0.957	0.979	0.948	0.007	0.123	1.515	0.992	0.948	0.007	0.123	1.515	0.992	0.948	46.701	
C: 7	0.380	0.000	-0.002	0.082	0.674	1.019	0.958	-0.004	0.143	2.035	0.919	0.934	0.001	0.127	1.602	0.943	0.944	0.001	0.127	1.602	0.943	0.944	39.311	
PH																								
C: 1	0.766	-0.401	0.000	0.057	0.327	1.028	0.954	-0.000	0.060	0.355	0.986	0.948	-0.000	0.057	0.327	1.028	0.954	-0.000	0.057	0.327	1.028	0.954	1.000	
C: 2	0.494	-0.401	-0.001	0.073	0.539	1.002	0.953	0.001	0.102	1.039	0.950	0.949	-0.001	0.091	0.832	1.001	0.954	-0.001	0.091	0.832	1.001	0.954	30.884	
C: 3	0.549	-0.401	-0.000	0.066	0.434	1.044	0.963	0.002	0.162	2.631	0.761	0.946	-0.001	0.081	0.658	1.024	0.955	-0.001	0.081	0.658	1.024	0.955	26.256	
C: 4	0.618	-0.401	0.001	0.063	0.398	1.016	0.947	-0.029	1.207	145.633	0.337	0.825	-0.003	0.074	0.554	0.989	0.946	-0.003	0.074	0.554	0.989	0.946	22.798	
C: 5	0.370	-0.401	0.000	0.084	0.702	1.031	0.953	-0.003	0.094	0.892	0.986	0.943	0.000	0.113	1.266	1.005	0.956	0.000	0.113	1.266	1.005	0.956	43.059	
C: 6	0.342	-0.401	-0.000	0.096	0.926	1.030	0.953	-0.001	0.109	1.195	0.947	0.934	0.010	0.135	1.841	0.998	0.947	0.010	0.135	1.841	0.998	0.947	49.230	
C: 7	0.332	-0.401	-0.002	0.090	0.813	0.998	0.950	0.004	0.154	2.377	0.947	0.940	-0.008	0.140	1.977	0.935	0.935	-0.008	0.140	1.977	0.935	0.935	42.194	
Non-PH																								
C: 1	0.885	-0.646	0.000	0.058	0.335	0.954	0.945	-0.004	0.056	0.319	0.980	0.947	-0.000	0.058	0.335	0.954	0.945	-0.000	0.058	0.335	0.954	0.945	1.000	
C: 2	0.664	-0.646	0.136	0.065	2.265	0.974	0.420	-0.002	0.078	0.604	0.946	0.938	0.005	0.074	0.548	0.969	0.942	0.005	0.074	0.548	0.969	0.942	17.648	
C: 3	0.723	-0.646	0.098	0.063	1.367	0.963	0.622	-0.029	0.085	0.816	0.883	0.929	0.003	0.070	0.493	0.950	0.931	0.003	0.070	0.493	0.950	0.931	14.312	
C: 4	0.784	-0.646	0.048	0.060	0.596	0.963	0.861	-0.042	0.381	14.682	0.370	0.888	-0.011	0.065	0.429	0.960	0.931	-0.011	0.065	0.429	0.960	0.931	12.588	
C: 5	0.508	-0.646	0.202	0.073	4.636	0.989	0.210	0.096	0.080	1.574	0.947	0.737	0.016	0.086	0.759	1.008	0.945	0.016	0.086	0.759	1.008	0.945	33.533	
C: 6	0.486	-0.646	0.204	0.087	4.904	0.985	0.334	0.151	0.089	3.072	0.981	0.574	0.028	0.109	1.266	0.986	0.940	0.028	0.109	1.266	0.986	0.940	38.197	
C: 7	0.500	-0.646	0.263	0.073	7.475	0.983	0.051	-0.032	0.129	1.769	0.855	0.906	0.027	0.096	0.999	0.968	0.937	0.027	0.096	0.999	0.968	0.937	30.439	

Each scenario is repeated 1000 times. Event rate refers to the average proportion of the sample with an observed event. Truth refers to the Monte Carlo average of $\hat{\beta}$ at $n = 2000$ and under censoring scenario 1. ESE refers to the empirical standard error as calculated through the replicated datasets. MSE refers to the mean squared error times 1000. SE/ESE refers to the average ratio of analytically calculated

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

sandwich standard error to the ESE. CP refers to the coverage probability of the 95 % confidence intervals. Avg Groups refer to the average number of censoring-specific groups derived from the chosen tree used to obtain $\hat{\beta}_{CR}$

Table 5

Patient characteristics by access type

Variable	SA fistula (<i>n</i> = 401)	VT fistula (<i>n</i> = 111)	Prosthetic graft (<i>n</i> = 1030)
Age	58.42 ± 16.60	65.15 ± 14.94	63.29 ± 14.49
BMI	26.13 ± 14.82	25.10 ± 6.18	28.30 ± 21.27
Gender			
Male	281 (70 %)	68 (61 %)	475 (46 %)
Female	120 (30 %)	43 (39 %)	555 (54 %)
Race			
Caucasian	255 (64 %)	76 (68 %)	583 (57 %)
African–Americans	105 (26 %)	25 (23 %)	366 (36 %)
Others (mainly Asians)	38 (9 %)	9 (8 %)	74 (7 %)
NA	3 (1 %)	1 (1 %)	7 (1 %)
Smoking			
Non-smoker	193 (48 %)	58 (52 %)	536 (52 %)
Former smoker	110 (27 %)	32 (29 %)	317 (31 %)
Current smoker	69 (17 %)	13 (12 %)	116 (11 %)
NA	29 (7 %)	8 (7 %)	61 (6 %)
Diabetes			
No	205 (51 %)	56 (50 %)	420 (41 %)
Yes	188 (47 %)	52 (47 %)	582 (57 %)
NA	8 (2 %)	3 (3 %)	28 (3 %)
Serum albumin	3.58 ± 0.62	3.45 ± 0.54	3.45 ± 0.56

Table 6

Adjusted hazard ratio estimates and 95 % confidence intervals based on the naive estimator and the censoring-robust estimator

	Adj. hazard ratio (naive)	Adj. hazard ratio (censoring-robust)
Age	1.003 (0.997–1.008)	1.001 (0.994–1.008)
BMI	0.996 (0.991–1.002)	0.996 (0.990–1.002)
Female	1.150 (0.991–1.335)	1.248 (1.034–1.506)
Race: African–Americans	1.131 (0.967–1.323)	1.206 (1.012–1.437)
Race: others	0.808 (0.595–1.099)	0.537 (0.265–1.091)
Diabetes	1.060 (0.912–1.231)	1.077 (0.892–1.301)
Serum albumin	0.823 (0.725–0.935)	0.849 (0.738–0.977)
Access type: venous transposition fistula	1.857 (1.366–2.525)	2.080 (1.449–2.985)
Access type: prosthetic graft	1.426 (1.183–1.720)	1.640 (1.234–2.179)