Estimation of Heterogeneous Restricted Mean Survival Time Using Random Forest

Mingyang Liu and Hongzhe Li*

Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Estimation and prediction of heterogeneous restricted mean survival time (hRMST) is of great clinical importance, which can provide an easily interpretable and clinically meaningful summary of the survival function in the presence of censoring and individual covariates. The existing methods for the modeling of hRMST rely on proportional hazards or other parametric assumptions on the survival distribution. In this paper, we propose a random forest based estimation of hRMST for right-censored survival data with covariates and prove a central limit theorem for the resulting estimator. In addition, we present a computationally efficient construction for the confidence interval of hRMST. Our simulations show that the resulting confidence intervals have the correct coverage probability of the hRMST, and the random forest based estimate of hRMST has smaller prediction errors than the parametric models when the models are mis-specified. We apply the method to the ovarian cancer data set from The Cancer Genome Atlas (TCGA) project to predict hRMST and show an improved prediction performance over the existing methods. A software implementation, srf using R and C++, is available at https://github.com/lmy1019/SRF.

Keywords: estimating equation, high dimensional data, non-parametric survival estimation, regression forest, inference

1. INTRODUCTION

In epidemiological and biomedical studies, time to an event or survival time $T$ is often the primary outcome of interest. Important quantities related to survival time include hazard rate (HR), $t$-year survival probability, and the mean survival time. Among these, HR is one of the most commonly used quantity due to its strong connection to the proportional hazards regression model or Cox model. Cox model is a very popular regression model for censored survival data due to its computational feasibility and theoretical properties (Cox, 1972, 1975; Andersen and Gill, 1982; Gill and Gill, 1984; Huang et al., 2013; Fang et al., 2017). However, when there is a departure from the proportional hazards assumption, the connection between HR and survival function is lost and it is difficult to interpret HR (Wang and Schaubel, 2018). The $t$-year survival probability is the probability of survival time greater than a pre-specified time $t$. It is not suitable for summarizing the global profile of $T$ over the duration of a study (Tian et al., 2014). In contrast, mean survival time is an alternative quantity since it takes the whole distribution of $T$ into account. However, the mean of $T$ may not always be estimable in the presence of censoring. For example, let $C$ denotes the
There are two main approaches for hRMST regression. One approach is to estimate hRMST indirectly through hazard regression (Zucker, 1998; Chen and Tsiatis, 2001; Zhang and Schaubel, 2011). This approach starts by estimating the regression parameters and the baseline hazard from a Cox model, calculating the cumulative baseline hazard, transforming it to obtain the survival function and, finally, obtaining the hRMST through Equation (1). Such an indirect hRMST estimation is inconvenient and computationally cumbersome for obtaining a point estimate and its corresponding asymptotic standard error. An alternative approach is to model hRMST with the baseline covariates X directly via some parametric assumptions, eg. $g(\mu^L(X_i)) = \beta_0^L X_i$, where g is a strictly monotone link function with a continuous derivative within an open neighborhood (Tian et al., 2014; Wang and Schaubel, 2018). A major weakness of this approach, however, is their inability to choose a proper link function, which may lead to the model misspecification. As an example, we simulate $x_1, \ldots, x_n$ independently from the uniform distribution on [0, 1] with a survival time model

$$T = \exp(2X_1 + 5) + 1 + \epsilon, \quad \epsilon \sim N(0, 10^2),$$

where we assume that the censoring time $C$ and the restricted time $L$ satisfy $P(C \leq T \wedge L) = 33\%$ and $P(L \leq T \wedge C) = 11\%$. Our goal is to estimate $\mu^L(x)$. Figure 1 shows a set of predictions on an artificially generated data set from Equation (2). Compared with other methods, the random forest is able to estimate the target function closely, especially when $\mu^L(x)$ approaches $L$.

For the continuous outcomes without censoring, random forest (Breiman, 2001, 2004) is a popular method of non-parametric regression that has shown effectiveness in many applications (Svetnik et al., 2003; Díaz-Uriarte and Alvarez de Andrés, 2006; Cutler et al., 2007). It is invariant under scaling and various other transformations of feature values, robust to inclusion of irrelevant features (Hastie et al., 2001), and versatile enough to be applied to large-scale problems (Bias and Scornet, 2016). Besides strong empirical results, theoretical results such as consistency (Meinshausen, 2006; Biau et al., 2008; Biau, 2012; Denil et al., 2014) and asymptotic normality (Wager and Athey, 2015; Mentch and Hooker, 2016; Athey et al., 2018; Friedberg et al., 2018) have also been obtained for regression models without censoring. Extending random forest to censored survival data has been proposed in several recent papers (Ishwaran et al., 2008; Steingrimsson et al., 2019), focusing on implementations and algorithms. However, there has been little theoretical work in statistical inference of such random survival forest. Ishwaran and Kogalur (2011) proved the consistency of the random survival forest by showing that the forest ensemble survival function converges uniformly to the true population survival function.

Instead of focusing on predicting the survival function or the survival probability as the algorithms implemented by Ishwaran et al. (2008) and Steingrimsson et al. (2019), we develop in this paper a random forest framework to model the hRMST directly given the baseline covariates in the presence of possibly covariate-dependent censoring. This approach provides a non-parametric estimation of hRMST adjusting for covariates. Due to the complex relationship between the survival time and the
covariates, it is desirable to have more flexible methods to estimate the hRMST than the approaches that a certain link function has to be assumed. Our construction of random forest is based on the estimated IPCW. We show that the resulting survival random forest estimates of hRMST has the asymptotic normality property that can be used to obtain the point-wise confidence interval with theoretical guarantees. To the best of our knowledge, it is the first asymptotic normality result for the predictions in the context of censored survival data using random forest.

The remainder of the paper is organized as follows. In section 2, we describe the proposed random forest estimator. Asymptotic properties are given in section 3. In section 4, we conduct simulation studies to evaluate the accuracy of the proposed method in the finite sample settings. In section 5, we apply our method to an ovarian cancer data set of The Cancer Genome Atlas (TCGA) project (http://cancergenome.nih.gov/abouttcga) to evaluate the predictions of the hRMST for ovarian cancer patients using their acylcarnitine measurements and clinical variables. We conclude this chapter with a brief discussion in section 6.

2. RANDOM FOREST FOR ESTIMATING THE hRMST

We begin with some notation. Let $X_i$ be the baseline covariates for subject $i$ from a cohort of sample size $n$ and $T_i$ be the survival time for subject $i$. Let $C_i$ be the censoring time, which is independent of $T_i$ conditional on the baseline covariates $X_i$. The observation time for subject $i$ is $Z_i = T_i \wedge C_i$, where $a \wedge b = \min\{a, b\}$. The indicator for censoring is denoted by $\delta_i = 1_{\{T_i \leq C_i\}}$. Our observed i.i.d. data are given as $\{(X_i, Z_i, \delta_i) : i = 1, \ldots, n\}$.

Let $L$ be a pre-specified time point of interest, before the maximum follow-up time $\tau = \max\{Z_i : i = 1, \ldots, n\}$. As in Wang and Schaubel (2018), $L$ is normally chosen as a time point of clinical relevance or, at least, of particular interest to the investigators, respecting the bound at the maximum follow-up time. Denote the restricted observation time as $Z_i^L = Z_i \wedge L$ and its corresponding indicator $\delta_i^L = 1_{\{T_i \wedge L \leq C_i\}}$. Our goal is to estimate covariate-adjusted RMST or hRMST $\mu^L(x) = E[Z^L|X = x]$ and to construct its confidence interval.

2.1. Forest-Based Local Estimating Equation for hRMST

Given the observed data $\{(X_i, \delta_i, Z_i)\}_{i=1}^n$, and a restriction threshold $L$, we first present a random forest method to estimate $\mu^L(x)$. The idea of the approach is to solve a weighted estimating equation for $\mu^L(x)$, where the estimating equation functions of the observations whose covariates to $x$ will have larger weights. Specifically, let $w_i = \delta_i^L/(1 - G(Z_i^L|X_i))$ be the IPCW of the $i$th data point under the true censoring distribution $G(\cdot|X_i)$. The (infeasible) estimating equation function $w_i(Z_i^L - \mu^L(x))$ of $X_i = x$ satisfies $E[w_i(Z_i^L - \mu^L(x))|X_i = x] = E[T_i \wedge L|X_i = x] - \mu^L(x) = 0$. If the local weights $\{\alpha_i(x)\}_{i=1}^n$ are also known, the solution to the empirical estimating equation for $\mu^L(x)$

$$
\sum_{i=1}^n \alpha_i(x)w_i(Z_i^L - \mu) = 0
$$

is given as

$$
\sum_{i=1}^n \alpha_i(x)w_iZ_i^L
$$

which provides a good candidate of estimator for $\mu^L(x)$. However we do not know the censoring distribution $G$ and the local weights $\{\alpha_i(x)\}_{i=1}^n$, which need to be estimated from the data. We assume censoring distribution $G$ follows a Cox model, a natural choice for modeling censoring times in the context of IPCW. Let

$$
\hat{w}_i = \frac{\delta_i^L}{1 - G(Z_i^L|X_i)}
$$

be the estimated IPCW for $i$th observation with $\hat{G}(\cdot|X_i)$ derived from the data through Cox model. We define the estimating equation function for $i$th observation with its corresponding estimated IPCW as

$$
\psi_{\mu^L(x)}(X_i, Z_i^L, \delta_i^L) = \hat{w}_i(Z_i^L - \mu^L(x)).
$$

Our approach to derive the local weights $\{\alpha_i(x)\}_{i=1}^n$ is through the random forest, which is an ensemble of survival trees constructed by Algorithm 1.

### Algorithm 1: Survival tree

```
SurvivalTree (set of observations I, domain X)
IPCW ← CoxModel(I);
Root $P_0 ←$ CreateNode(I, X);
Queue Q → InitializeQueue($P_0$);
while Q is NotNull do
    node $P ←$ Pop(Q);
    Solve $\hat{\mu}^L_P = \arg\min_{\mu} \sum_{X_i \in P} \psi_{\mu}(X_i, Z_i^L, \delta_i^L) |\mu|$
    Set $\rho_i = \hat{w}_i(Z_i^L - \hat{\mu}_P^L)$
    Split $P$ by maximizing
    $\Delta(C_1, C_2) = \sum_{j=1}^2 \frac{1}{|\{X_i \in C_j\}|} \left( \sum_{i : X_i \in C_j} \rho_i \right)^2$
    if split succeeds then
        AddQueue($C_1$);
        AddQueue($C_2$);
    end
end
```

It can be shown that $\rho_i$ is the influence function of the $i$th observation for $\hat{\mu}_P^L$. Let $F_n$ be the empirical distribution of the observations in node $P$, and let $F_{n,i} = (1 - \epsilon)F_n + \epsilon v_i$, with $v_i$ be the Dirac delta function at $i$th observation. Set

$$
\hat{\mu}_{P,i}^L = \hat{\mu}_P^L + \Delta_i, \text{ where } \hat{\mu}_{P,i}^L = \arg\min_{\mu} \int \psi_{\mu}(X, Z^L, \delta^L) dF_{n,i} |\mu|
$$

By Taylor expansion,

$$
0 = \int \psi_{\hat{\mu}_{P,i}^L}(X, Z^L, \delta^L) dF_{n,i}
$$
\[
\frac{\Delta_i}{\epsilon} = \frac{\psi_{\hat{\mu}^L}(X_i, Z_i^l, \delta_i^l)}{\int \psi_{\hat{\mu}^L}(X, Z^l, \delta^l)dF_{n,i}}
\]

and therefore the influence function of the \(i\)th observation for \(\hat{\mu}^L\) is

\[
\lim_{\epsilon \to 0} \frac{\Delta_i}{\epsilon} = \frac{\psi_{\hat{\mu}^L}(X_i, Z_i^l, \delta_i^l)}{\int \psi_{\hat{\mu}^L}(X, Z^l, \delta^l)dF_{n,i}}\frac{\hat{w}_i}{\sum_{i \in P} \hat{w}_i} = \rho_i.
\]

Athey et al. (2018) shows that maximizing the splitting criterion \(\tilde{\Delta}(C_1, C_2)\) is approximately equivalent to minimizing the weighted mean squared error \(\text{err}(C_1, C_2) = \sum_{i=1,2} \mathbb{P}(X \in C_i | X \in P) \mathbb{E}[(\hat{\mu}^{L, C_i} - \mu^L(X))^2 | X \in C_i]\).

In order to achieve consistency and asymptotic normality, we split the tree and make predictions in an honest way as introduced in Wager and Athey (2015). Specifically, each tree in an honest forest is grown using two non-overlapping subsamples of the training data. For the \(b\)th tree, given \(I_b\) and \(J_b\), we first choose the tree structure \(T_b\) using only the data in \(J_b\), and write \(x \leftrightarrow x'\) as the boolean indicator for whether the points \(x\) and \(x'\) fall into the same leaf of \(T_b\). In a second step, we define the set of neighbors of \(x\) as \(L_b(x) = \{i \in I_b : x \leftrightarrow x_i\}\). The weights of point \(x\) from a survival forest with \(B\) trees can be written as

\[
\alpha_i(x) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}_{\{X_i \in L_b(x)\}} \frac{1}{|L_b(x)|}.
\]

The empirical locally weighted estimating equation for \(\hat{\mu}^L(x)\) is then defined as

\[
\sum_{i=1}^{n} \alpha_i(x) \psi_{\mu}(X_i, Z_i^l, \delta_i^l) = 0,
\]
and the random forest estimator for the hRMST is the solution of Equation (4), which is

$$\hat{\mu}_L(x) = \sum_{i=1}^{n} \frac{\alpha_i(x) \hat{w}_i Z_{Li}}{\sum_{i=1}^{n} \alpha_i(x) \hat{w}_i}.$$  

We emphasize the difference between the IPCW used in building the survival trees and IPCW used to derive $\hat{\mu}_L(x)$. The IPCW used in building survival trees is estimated only by the data points from $J_b$ so that the resulting survival forest is honest. The IPCW used to derive $\hat{\mu}_L(x)$ is estimated from all data points.

3. ASYMPTOTIC DISTRIBUTION OF $\hat{\mu}_L(X)$

3.1. Asymptotic Normality

We derive a central limit theorem for survival forest estimate of hRMST. We first give three common assumptions that required for the most of the theoretical analysis of random forests.

**Assumption 1.** $\mu_L(x)$ is Lipschitz continuous w.r.t $x$.

**Assumption 2.** There exists a restricted time threshold $L$, such that $P(C > t \land L|X = x) \geq \epsilon_L > 0$ for any $x, t$.

**Assumption 3.** $\text{Var}(T \land L|X = x) > 0$ for any $x$.

As mentioned in the previous section, we model the conditional survival function of censoring distribution $G$ given baseline covariates. Because of its flexibility and popularity in practice, we adopt the proportional hazards model for hazard function of censoring distribution.

**Assumption 4.** The hazard function of censoring distribution follows $\lambda_L^C(t) = \lambda_0^C(t) \exp(X_i^C \beta_C)$

We make additional regularity assumptions that are widely used in analysis of estimates from the proportional hazards models. These assumptions are needed in order to quantify the difference between the estimated IPCW and true IPCW.

**Assumption 5.** $||X||_{\infty} < M_X < \infty$

**Assumption 6.** $\lambda_0^C(t) \leq \lambda_C^C < \infty$ for all $t$.

![FIGURE 3](Image) Simulation results of coverage probability for Model 2 with three different link functions, sample size of $n = 1,000, 2,000, 10,000$, and $p = 2, 4, 6, 8$. For each case, prediction coverage probability is calculated over the samples in the testing data set.
Assumption 7. \( \Omega_{C}(\beta) = E \left[ \int_0^\infty \frac{r_i^{(k)}(t, \beta)}{r_0^{(k)}(t, \beta)} dx_i^G(t) \right] \)

is positive definite, where \( R_i(t) = 1(Z_i \geq t) \), \( r^{(k)}(t, \beta) = E[\exp(\beta'X_i)R_i(t)X_i^{(k)}] \), \( \bar{x}(t, \beta) = E[\exp(\beta'X_i)R_i(t)X_i^{(k)}] \), \( N_i^C(t) = 1_{Z_i \leq t, k=0} \).

Assumption 8. \( P(R_i(t) = 1|X_i = x) \geq r > 0 \) for some positive constant and for any \( t, x \). This assumption implies that \( r^{(0)}(t, \beta) = E[\exp(\beta'X_i)R_i(t)] = E[\exp(\beta'X_i)E[R_i(t)|X_i]] \geq r > 0 \).

Following Wager and Athey (2015) and Athey et al. (2018), we assume that all trees are symmetric, in that their output is invariant to permuting the indices of Estimation-Part in training examples (see Corollary 6 of Wager and Athey (2015) for more details about this symmetry). They also require balanced splits in the sense that every split puts at least a fraction \( \omega \) of the
observations in the parent node into each child, for some \( \omega > 0 \).
Finally, the trees are randomized in such a way that, at every split, the probability that the tree splits on the \( j \)th feature is bounded from below by \( \pi > 0 \).
The forest is honest and built via subsampling with subsample size \( s \) satisfying \( s/n \to 0 \) and \( s \to \infty \).

Under the assumptions listed above, we have the following asymptotic distribution result for the random forest-based estimate of the hR MST.

**Theorem 1.** Under Assumptions 1, 2, 3, 4, 5, 6, 7, 8, for each fixed test point \( x \), there is a sequence \( \sigma_0(x) = \sqrt{\hat{\mu}^2(x)} / \sigma_n(x) \to 0 \),

\[
\frac{\hat{\mu}_n(x) - \mu(x)}{\sigma_n(x)} \to_d N(0, 1)
\]

if subsampling size

\[
\beta_{\min} = 1 - \left( 1 + \frac{\pi^{-1} \log(\omega^{-1})}{\log((1 - \omega^{-1}))} \right)^{-1}
\]

where \( \omega > 0 \) is the low-bound fraction for observations in the parent node into each child, and \( \pi > 0 \) is the lower-bound of the probability that the tree splits on any features.

We give a consistent estimate of \( \sigma_0^2(x) \) based on half-sampling (Efron, 1980) and the method of Sexton and Laake (2009).

### 3.2. Estimation of the Variance

Following Athey et al. (2018), we use the random forest delta method to develop a variance estimate of the survival forest prediction \( \hat{\mu}(x) \). Athey et al. (2018) provides a consistent estimate of \( \sigma_0^2(x) \) using \( \hat{s}_n^2(x) \), where \( \hat{s}_n^2(x) = (V(x)^{-1}) H_n(x) (V(x)^{-1})' \) with

\[
H_n(x) = \text{Var} \left[ \sum_{i=1}^n \alpha_i(x) \psi_{\mu(x)}(X_i, \delta^L, \delta^L) \right]
\]

\[
V(x) = \frac{\delta(x)}{\delta(\mu(x))} \sum_{i=1}^n \alpha_i(x) \psi_{\mu(x)}(X_i, \delta^L, \delta^L) (X_i \mid \mu = \mu(x))
\]

In our context, \( V(x) = -1 \), then simply we have \( \hat{s}_n^2(x) = H_n(x) \).

A consistent estimator for \( H_n(x) \) can be obtained using half-sampling estimator (Efron, 1980; Athey et al., 2018). Let \( \Psi_{H(i)} \) be the average of the empirical estimating equation functions averaged over the trees that only use the data from the half-sample \( H(i) \), denoted by \( S_{H(i)} \).

\[
\Psi_{H(i)}(x) = \frac{1}{|S_{H(i)}|} \sum_{b \in S_{H(i)}} \sum_{i \in L_b(x)} \psi_{\mu(x)}(X_{i}, \delta^L, \delta^L)
\]

where \( L_b(x) \) contains neighbors of \( x \) in the \( b \)th tree. An ideal half-sampling estimator is then defined as

\[
\hat{H}_{HS}^n(x) = \left( \frac{n}{n/2} \right)^{-1} \sum_{H(i) \in S_{H(i)}} (E_0[\Psi_{H(i)}(x)] - E_0 \tilde{\Psi}(x))^2
\]

### Table 1: Comparison of Mean-Absolute-Error (MAE) and Rooted-Mean-Squared-Error (RMSE) for Model 1 with different link functions.

| \( p \) | SRF | Naive.Cox | Naive.km | Lu.id | Lu.exp | Wang.id | Wang.exp |
|------|-----|----------|----------|-------|--------|---------|----------|
| 5    | 0.1359 | 0.1371 | 0.2067 | 0.1341 | 0.1346 | 0.1341 | 0.1346 |
| 10   | 0.1699 | 0.1695 | 0.2466 | 0.1697 | 0.1691 | 0.1686 | 0.1691 |
| 20   | 0.1721 | 0.1710 | 0.2497 | 0.1710 | 0.1715 | 0.1709 | 0.1714 |
| 30   | 0.1703 | 0.1693 | 0.2464 | 0.1686 | 0.1691 | 0.1685 | 0.1690 |
| 5    | 0.1347 | 0.1359 | 0.2048 | 0.1330 | 0.1335 | 0.1330 | 0.1335 |
| 10   | 0.1684 | 0.1680 | 0.2441 | 0.1673 | 0.1677 | 0.1672 | 0.1677 |
| 20   | 0.1706 | 0.1695 | 0.2472 | 0.1695 | 0.1701 | 0.1695 | 0.1699 |
| 30   | 0.1761 | 0.1690 | 0.2444 | 0.1631 | 0.1637 | 0.1630 | 0.1636 |
| 5    | 0.1689 | 0.1679 | 0.2439 | 0.1672 | 0.1678 | 0.1671 | 0.1676 |
| 10   | 0.2472 | 0.2538 | 33.888 | 24.496 | 24.723 | 24.436 | 24.709 |
| 20   | 30.827 | 30.860 | 39.296 | 30.608 | 30.773 | 30.577 | 30.749 |
| 30   | 31.058 | 31.052 | 39.621 | 30.889 | 31.076 | 30.850 | 31.048 |
| 20   | 30.744 | 30.695 | 39.181 | 30.479 | 30.689 | 30.438 | 30.646 |

The number of covariates \( p = 5, 10, 20 \), for each \( p \), the first row is MAE, the second row is RMSE. SRF, proposed random forest-based estimator; Naive.km, estimate based on Kaplan–Meier estimator without adjusting for the covariates; Naive.Cox, Cox regression based estimator; Lu.id, method of Tian et al. (2014) with identity link; Lu.exp, method of Tian et al. (2014) with exponential link; Wang.id, method of Wang and Schaubel (2018) with identity link; Wang.exp, method of Wang and Schaubel (2018) with exponential link.
\[
\Psi(x) = \left( \frac{n}{n/2} \right)^{-1} \sum_{H : |H| = n/2} \Psi_H(x)
\]

where \( \Theta \) is the randomness in building honest tree, including splitting data into random halves and randomness in selecting variables to split. \( \hat{H}^{HIS}_n(x) \) is similar to classic bootstrap estimator for the standard error, except that the sampling distribution for \( \hat{H}^{HIS}_n(x) \) is the half sampling distribution instead of the bootstrap sampling. Denote \( \hat{E}_n \) and \( \hat{V}ar_n \) as the expectation and variance under the half sampling distribution, then \( \hat{H}^{HIS}_n(x) = \hat{V}ar_n[\Psi(x)] \).

Since carrying out the full half-sampling computation and expectation with respect to \( \Theta \) are impractical, Sexton and Laake (2009) pointed out that \( \hat{H}^{HIS}_n(x) \) can be efficiently approximated by the following law of total variance:

\[
\hat{H}^{HIS}_n(x) = \hat{V}ar_n \left[ \hat{E}_n \left[ \frac{1}{M} \sum_{m=1}^{M} \Psi_{H,\phi_m}(x) \right] \right]
\]

\[
= \hat{V}ar_n \left[ \frac{1}{M} \sum_{m=1}^{M} \Psi_{H,\phi_m}(x) \right] - \hat{E}_n \left[ \hat{V}ar_n \left[ \frac{1}{M} \sum_{m=1}^{M} \Psi_{H,\phi_m}(x) \right] \right]
\]

which leads to a Monte Carlo approximation of \( \hat{H}^{HIS}_n(x) \) by

\[
\hat{\sigma}^2_n(x) = \hat{V}ar_n \left[ \frac{1}{M} \sum_{m=1}^{M} \Psi_{H,\phi_m}(x) \right] - \hat{E}_n \left[ \hat{V}ar_n \left[ \frac{1}{M} \sum_{m=1}^{M} \Psi_{H,\phi_m}(x) \right] \right].
\]

In order to approximate random forest randomness quantity \( \hat{V}ar_n \) and sampling randomness quantities \( \hat{V}ar_n, \hat{E}_n \), we split \( B \) trees in \( G \) groups and each group has \( l \) trees, and the trees in the same group have the same half sample. The final consistent estimator \( \hat{\sigma}^2_n(x) \) can be written as

\[
\hat{\sigma}^2_n(x) = \frac{1}{G-1} \sum_{g=1}^{G} \left( \hat{\Psi}_g(x) - \hat{\Psi}(x) \right)^2 - \frac{1}{(l-1)} \frac{1}{B} \sum_{g=1}^{G} \sum_{i=1}^{l} \left( \hat{\Psi}_g(x) - \hat{\Psi}_i(x) \right)^2
\]

where \( \hat{\Psi}_g(x) = \frac{1}{l} \sum_{i=1}^{l} \hat{\Psi}_g(x) \), and \( \hat{\Psi}(x) = \frac{1}{G} \sum_{g=1}^{G} \hat{\Psi}_g(x) \).

The following diagram summarizes the procedure of estimating the variance \( \hat{\sigma}^2_n(x) \).

\[
\frac{\sigma^2_n(x)}{\text{Asym.equivalent}} \quad \frac{\hat{\sigma}^2_n(x)}{\text{Half-Sampling estimator}}
\]
\[ \hat{H}_{n}^{HS}(x) \xleftarrow{\text{Empirical estimator}} \hat{a}_{n}^{2}(x) \]

where from left to right, the first arrow is based on Theorem 5 of Athey et al. (2018), the second arrow is based on half-sampling of Efron (1980), and the third arrow is supported by Equations (5) and (6) and the method of Sexton and Laake (2009).

4. SIMULATION STUDIES

We present simulations to evaluate the performance of the proposed method in finite sample setting. Two different models for the survival time are considered

- Model 1: \( T = g^{-1}(\alpha_0 + \sum_{i=1}^{n} \alpha_i X_i) + \epsilon \)
- Model 2: \( T = g^{-1}(\alpha_0 + \sum_{i=1}^{n} \alpha_i X_i^2) + \epsilon \)

where \( X_i, \ldots, X_p \) are independently generated from \( \text{Unif}(-1, 1) \), \( \alpha_0 = 5 \), \( \alpha_1 = \alpha_2 = 0.25 \) and \( \alpha_i = 0 \) for \( i > 2 \), and \( \epsilon \sim N(0, \sigma^2) \). The variance \( \sigma^2 \) is chosen to have proper signal-noise ratio (SNR),

\[
\text{SNR} = \frac{\text{Var}(g^{-1}(\alpha_0 + \sum_{i=1}^{n} \alpha_i X_i))}{\text{Var}(\epsilon)}.
\]

We generate the independent censoring time \( C_i \) from a Cox model with the following hazard \( \lambda = \lambda_C \exp(X_i \log 2) \) and \( \lambda_C \) is chosen to have a proper un-censoring rate. The link function \( g \) can have the following form

- Identity link: \( g^{-1}(x) = x \);
- Exp link: \( g^{-1}(x) = \exp(x) \);
- Log-exp link: \( g^{-1}(x) = \log(\exp(x) + 1) \).

4.1. Evaluation of Coverage Probability of Predictions

To evaluate the asymptotic results in Theorem 1, we generate five training data sets and one testing data set with the same sample size. The coverage probability performance is evaluated on the testing data set with predictions and confidence intervals derived from 5 independent training data sets. More specifically, for each observation in the testing sample, we obtain the 95% confidence intervals and record how many times an hRMST observation in the testing sample is within five estimated 95% confidence intervals. The coverage probability of an observation is defined by the proportion of being covered, and the overall coverage probability of the testing sample is defined by the average of coverage probability of each of its observation. We present the coverage probability results with sample size \( n = 1,000, 2,000, 5,000 \) for Model 1, and \( n = 1,000, 2,000, 10,000 \) for Model 2. By choosing the proper \( \lambda_C \), we control the un-censoring rate around 60–70% for different link functions: \( \lambda_C \sim 0.08 \) for Identity link and Log-exp link, and \( \lambda_C \sim 0.003 \) for Exp link. The truncation time \( L \) is

| \( p \) | SRF | Naive.Cox | Naive.km | Lu.id | Lu.exp | Wang.id | Wang.exp |
|-----|-----|-----------|----------|-------|--------|---------|---------|
| 5   | 0.1361 | 0.1353 | 0.2051 | 0.1377 | 0.1344 | 0.1336 | 0.1342 |
| 0.1706 | 0.1681 | 0.2457 | 0.1687 | 0.1693 | 0.1685 | 0.1690 |
| 10  | 0.1444 | 0.1430 | 0.2160 | 0.1402 | 0.1408 | 0.1403 | 0.1408 |
| 0.1755 | 0.1732 | 0.2523 | 0.1726 | 0.1731 | 0.1725 | 0.1730 |
| 0.1392 | 0.1372 | 0.2078 | 0.1345 | 0.1351 | 0.1345 | 0.1351 |
| 0.1723 | 0.1699 | 0.2484 | 0.1694 | 0.1700 | 0.1692 | 0.1698 |

Model 1: identity link, \( n = 3,000, \text{SNR} = 0.3 \)

Model 1: log-exp link, \( n = 3,000, \text{SNR} = 0.3 \)

Model 1: exp link, \( n = 3,000, \text{SNR} = 0.3 \)

The number of covariates \( p = 5, 10, 20 \), for each \( p \), the first row is MAE, the second row is RMSE. SRF, proposed random forest-bases estimator; Naive.km, estimate based on Kaplan–Meier estimator without adjusting for the covariates; Naive.Cox, Cox regression based estimator; Lu.id, method of Tian et al. (2014) with identity link; Lu.exp, method of Tian et al. (2014) with exponential link; Wang.id, method of Wang and Schaubel (2018) with identity link; Wang.exp, method of Wang and Schaubel (2018) with exponential link.
chosen to make the truncation rate fall into 2% – 5%. Specifically, $L \sim 5.4$ for Identity link and Log-exp link, and $L \sim 220$ for Exp link.

Figures 2, 3 present the results for Model 1 and Model 2 under three different link functions. We see that the coverage probability approaches to nominal level 95% when the sample size gets larger. If $p$ is smaller, the coverage probability is closer to 95%. This corresponds to the result of Theorem 3 in Wager and Athey (2015), which states that the rate of convergence of the bias of random forest estimator is $O(n^{-p})$ for some constant $K$. When the sample size $n$ is fixed, bigger $p$ leads to larger bias in the estimates of hRMST, and under-coverage of the confidence interval. On the other hand, when $p$ is fixed, bigger $n$ results in a smaller bias and leads to a better coverage of the confidence interval.

4.2. Comparison of Prediction Performance With Existing Methods

We compare our proposed method with several existing methods for hRMST estimation, including

- **Naive.km**: using Kaplan–Meier estimator for survival function and computing hRMST by Equation (1). Covariates are not adjusted.
- **Naive.Cox**: using proportional hazards estimator for the survival function and computing hRMST by Equation (1). The censoring distribution is assumed to follow the proportional hazards assumption.
- **Lu.method**: using some parametric forms of hRMST and computing hRMST by solving a weighted estimating equation. The censoring distribution is assumed to be independent of the covariates (Tian et al., 2014). We consider Identity link and Exp link in the simulations.
- **Wang.method**: using some parametric forms of hRMST and computing hRMST by solving a weighted estimating equation. The censoring distribution is assumed to follow the proportional hazards assumption. We consider Identity link and Exp link in the simulations (Wang and Schaubel, 2018).

We compare all these methods under Model 1 and Model 2, and use the Mean-Absolute-Error (MAE) and Rooted-Mean-Squared-Error (RMSE), introduced in Davison and Hinkley (1997), Tian et al. (2007), and Wang and Schaubel (2018), to measure the performance of these methods.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i^L}{1 - \hat{G}(Z_i^L | X_i = x)} \left| Z_i^L - \hat{L}(X_i) \right|,$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i^L}{1 - \hat{G}(Z_i^L | X_i = x)} \left[ Z_i^L - \hat{L}(X_i) \right]^2}.$$

### TABLE 4 | Comparison of Mean-Absolute-Error (MAE) and Rooted-Mean-Squared-Error (RMSE) for Model 2 with different link functions and the censoring distribution is mis-specified with $\alpha = 0.5$.

| $p$ | SRF | Naive.Cox | Naive.km | Lu.id | Lu.exp | Wang.id | Wang.exp |
|-----|-----|-----------|----------|-------|--------|---------|---------|
| 5   | 0.1230 | 0.1378 | 0.1374 | 0.1385 | 0.1385 | 0.1377 | 0.1377 |
| 10  | 0.1514 | 0.1657 | 0.1653 | 0.1683 | 0.1663 | 0.1658 | 0.1658 |
| 10  | 0.1310 | 0.1450 | 0.1442 | 0.1457 | 0.1457 | 0.1447 | 0.1447 |
| 10  | 0.1562 | 0.1704 | 0.1695 | 0.1712 | 0.1712 | 0.1704 | 0.1704 |
| 20  | 0.1533 | 0.1894 | 0.1884 | 0.1843 | 0.1843 | 0.1892 | 0.1892 |

**Model 1: identity link, $n = 3,000$, SNR = 0.3**

| $p$ | SRF | Naive.Cox | Naive.km | Lu.id | Lu.exp | Wang.id | Wang.exp |
|-----|-----|-----------|----------|-------|--------|---------|---------|
| 5   | 0.1213 | 0.1359 | 0.1355 | 0.1385 | 0.1385 | 0.1358 | 0.1358 |
| 10  | 0.1494 | 0.1634 | 0.1630 | 0.1640 | 0.1640 | 0.1636 | 0.1636 |
| 10  | 0.1292 | 0.1430 | 0.1422 | 0.1437 | 0.1437 | 0.1427 | 0.1427 |
| 20  | 0.1244 | 0.1681 | 0.1673 | 0.1689 | 0.1689 | 0.1681 | 0.1681 |
| 20  | 0.1512 | 0.1845 | 0.1834 | 0.1838 | 0.1838 | 0.1832 | 0.1832 |

**Model 1: log-exp link, $n = 3,000$, SNR = 0.3**

| $p$ | SRF | Naive.Cox | Naive.km | Lu.id | Lu.exp | Wang.id | Wang.exp |
|-----|-----|-----------|----------|-------|--------|---------|---------|
| 5   | 21.270 | 23.793 | 23.697 | 24.016 | 24.009 | 23.535 | 23.534 |
| 10  | 26.187 | 28.147 | 28.075 | 28.329 | 28.322 | 28.133 | 28.132 |
| 10  | 22.824 | 25.159 | 24.948 | 25.408 | 25.399 | 24.843 | 24.842 |
| 20  | 27.067 | 29.009 | 28.823 | 29.293 | 29.227 | 28.945 | 28.943 |
| 20  | 21.832 | 23.896 | 23.708 | 24.188 | 24.177 | 23.698 | 23.697 |
| 20  | 26.635 | 28.417 | 28.221 | 28.753 | 28.740 | 28.499 | 28.499 |
We set \( n = 3,000, \text{SNR} = 0.3 \). For Identity link and Log-exp link, \( \lambda_C = 0.08, L = 5.3 \). For Exp link \( \lambda_C = 0.0026, L = 190 \). We calculate the MAE and RMSE for our method and four existing methods (both Lu.method and Wang.method have two link functions) under Model 1 and Model 2 and \( p = 5,10,20 \). Among all the considered models, our method in general has a better performance. As an example, Figure 4 visualizes the observed hRMST generated from Log-exp link and predicted hRMST from our method and Wang.method, showing that the random forest can give better predictions.

Tables 1, 2 show the MAE and RMSE for Model 1 and Model 2, respectively. For Model 1, the parametric methods are correctly specified using the methods of Tian et al. (2014), Wang and Schaubel (2018), we expect that both methods perform well, and our method can have a comparable performance. For Model 2, our proposed method dominates all other methods. Increasing the number of non-predictive covariates does not have a big impact on the performance of our method.

When \( \alpha = 1 \), the gamma distribution degenerates to the exponential distribution we used for Tables 1, 2. Tables 3, 4 show the MAE and RMSE for Model 1 and Model 2 when \( \alpha = 0.5, \) and Tables 5, 6 show the MAE and RMSE for Model 1 and Model 2 when \( \alpha = 1.5 \). Results of \( \alpha \in (0.5, 1.5) \) are not very different from the results of \( \alpha = 1 \). Under Model 1, our method performs comparably well as methods of Tian et al. (2014), Wang and Schaubel (2018), and it dominates the others under Model 2. When feature dimension is low\((p = 5)\), the error metrics of our method when \( \alpha = 1 \) are in general lower than the error metrics when \( \alpha = 0.5, 1.5 \) for both Model 1 and Model 2. The additional errors can be regarded as the bias induced from the violation of PH assumption of the censoring distribution. When feature dimension is high\((p = 10, 20)\), bias from large \( p \) may dominate the bias from the violation of PH assumption of the censoring distribution.

### 5. APPLICATION TO THE TCGA OVARIAN CANCER DATA SET

We apply the proposed method to The Cancer Genome Atlas (TCGA) ovarian cancer functional proteomics data set (Akbani et al., 2015) that is publicly available (http://gdac.broadinstitute.org). The data sets include proteomic characterization of tumors using reverse-phase protein arrays (RPPA). Specifically, Akbani et al. (2015) reported an RPPA-based proteomic analysis using 195 high-quality antibodies that target total, cleaved, acetylated...
and phosphorylated forms of proteins in 412 high-grade serous ovarian cystadenocarcinoma (OVCA) samples. The function space covered by the antibodies used in the RPPA analysis encompasses major functional and signaling pathways of relevance to human cancer, including proliferation, DNA damage, polarity, vesicle function, EMT, invasiveness, hormone signaling, apoptosis, metabolism, immunological, and stromal function as well as transmembrane receptors, integrin, TGFβ, LKB1/AMPK, TSC/mTOR, PI3K/Akt, Ras/ MAPK, Hippo, Notch, and Wnt/beta-catenn signaling (Akbani et al., 2015).

After removing a few samples with missing data, the final data set includes 407 OVCA samples with a mean/median follow-up of 3.20/2.79 years and a total of 242 deaths and 40% censoring. To assess how different methods predict the hRMST, we performed the following cross-validation analysis. For a given $L$, we did 10-fold cross-validation on the data set. For each training data set in the cross-validation, we perform a univariate analysis to select top 5 most significant features based on univariate Cox regression analysis. We then estimate the hRMST on the test set using the training data sets with these 5 features as the predictors. We apply 7 different methods, including estimate based on the KM estimator, estimate based on the Cox model, the method of Tian et al. (2014) and the method of Wang and Schaubel (2018). We report the average

| $p$ | SRF | Naive.Cox | Naive.km | Lu.id | Lu.exp | Wang.id | Wang.exp |
|-----|-----|-----------|----------|-------|--------|---------|---------|
| 5   | 0.1227 | 0.1396 | 0.1395 | 0.1397 | 0.1397 | 0.1394 | 0.1394 |
| 10  | 0.1507 | 0.1666 | 0.1664 | 0.1668 | 0.1668 | 0.1666 | 0.1666 |
| 20  | 0.1514 | 0.167 | 0.1664 | 0.1669 | 0.1669 | 0.1668 | 0.1668 |
| 5   | 0.1232 | 0.1390 | 0.1386 | 0.1393 | 0.1393 | 0.1389 | 0.1389 |
| 20  | 0.1499 | 0.1659 | 0.1654 | 0.1663 | 0.1663 | 0.1661 | 0.1661 |

The number of covariates $p = 5, 10, 20$, for each $p$, the first row is MAE, the second row is RMSE. SRF, proposed random forest-bases estimator; Naive.km, estimate based on Kaplan–Meier estimator without adjusting for the covariates; Naive.Cox, Cox regression based estimator; Lu.id, method of Tian et al. (2014) with identity link; Lu.exp, method of Tian et al. (2014) with exponential link; Wang.id, method of Wang and Schaubel (2018) with identity link; Wang.exp, method of Wang and Schaubel (2018) with exponential link.

| $L$ | SRF | Naive.Cox | Naive.km | Lu.id | Lu.exp | Wang.id | Wang.exp |
|-----|-----|-----------|----------|-------|--------|---------|---------|
| 3   | 0.6879 | 0.9247 | 0.9483 | 0.9286 | 0.9355 | 0.7630 | 0.7721 |
| 4   | 0.8258 | 0.8925 | 0.8967 | 0.8966 | 0.8983 | 0.8438 | 0.8455 |
| 5   | 1.2033 | 1.5450 | 1.5686 | 1.5704 | 1.5777 | 1.2862 | 1.3044 |
| 5   | 1.2403 | 1.3597 | 1.3648 | 1.3830 | 1.3817 | 1.2719 | 1.2752 |
| 5   | 1.7479 | 2.2107 | 2.2395 | 2.2467 | 2.2306 | 1.8251 | 1.8540 |
| 5   | 1.6761 | 1.8594 | 1.8655 | 1.8989 | 1.8858 | 1.7168 | 1.7193 |

The first row is MAE, the second row is RMSE. SRF, proposed random forest estimator; Naive.km, estimate based on Kaplan–Meier estimator without adjusting for the covariates; Naive.Cox, Cox regression based estimator; Lu.id, method of Tian et al. (2014) with identity link; Lu.exp, method of Tian et al. (2014) with exponential link; Wang.id, method of Wang and Schaubel (2018) with identity link; Wang.exp, method of Wang and Schaubel (2018) with exponential link.
of MAE and RMSE on the samples in the testing sets over the 10-fold cross-validation.

The results are shown in Table 7 and Figure 5 for $L = 3, 4, 5$ (see Supplementary Material for $L = 6, 7, 8$). There are 45.9, 31.2, 19.4, 11.8, 8.1, 4.4% of the observations larger than $L$ for $L = 3, 4, 5, 6, 7, 8$ correspondingly. For different choices of $L$, our proposed random forest based method dominates the other methods in MAE and RMSE. The methods of Tian et al. (2014) and Wang and Schaubel (2018) are based on parametric form of hRMST. Cox model is heavily dependent on the proportional hazard assumption, and the Kaplan–Meier approach does not take the covariates into account. We also notice that the method of Wang and Schaubel (2018) always performs better than the method of Tian et al. (2014), possibly due to the fact that the censoring mechanism in the data depends on the covariates.
6. DISCUSSION

In this paper, we have developed a non-parametric random forest-based method for estimation of hRMST. Compared with traditional Cox model, which gets hRMST estimates by transforming the estimated hazard functions, directly modeling hRMST would be more preferable for computation and feature importance analysis. The proposed estimator can relax the parametric assumptions imposed on the survival time used in Tian et al. (2014) and Wang and Schaubel (2018), and can achieve better prediction performance. We have derived the asymptotic distribution of the random forest estimator using IPCW approach, and presented a procedure based on bag of little bootstraps to obtain the variance of the estimator. Our simulation results and analysis of TCGA data sets have shown promising performance in predicting hRMST as compared to the other available methods, even when the dimension is high and the covariates include irrelevant variables. The method is implemented by R and C++, and is available at https://github.com/lny1019/SRF.

The proposed method can be used to estimate the heterogeneous treatment effects in randomized clinical trials when the outcome is censored. One can simply apply the method separately to the treated group and the placebo group and take the difference. However, for the observational studies, one needs to account for the fact that the treatment assignments might not be completely at random. Wager and Athey (2015) developed a non-parametric causal forest for estimating heterogeneous treatment effects that extends Breiman’s random forest algorithm. In the potential outcomes framework with non-confounding, they showed that causal forest are pointwise consistent for the true treatment effect and have an asymptotically Gaussian and centered sampling distribution. For the observational studies with censored survival outcomes, it is also possible to combine the methods proposed here and the method of Wager and Athey (2015) in order to estimate the treatment effect on the restricted mean survival time.

The proposed methods can also be extended to take into account possible competing risk. This can be done by introducing an additional inverse probability weight (IPCW) to differentiate the non-informative censoring and competing risk censoring. In this case, the estimation equation \( \psi \) function with covariates history \( \tilde{X} = \tilde{x} \) under true \( G_C \) and \( G_R \) becomes

\[
\tilde{\nu}_L(\tilde{x}, Z^L, \delta^L) = \frac{1}{1 - G_C(Z^L|X = x)} \frac{1}{1 - G_R(Z^L|\tilde{X} = \tilde{x})} \delta^L(Z^L - \mu),
\]

where under competing risk scenario, \( \delta^L = 1\{T^L \leq \Delta < \delta_R\} \). The method proposed in this paper can be automatically adapted to the competing risk case and the asymptotic normality result can be derived similarly.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

ML and HL developed the ideas and the methods together, analyzed the real data sets, and wrote the manuscript. ML implemented the methods and performed the numerical analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by NIH grants GM123056 and GM129781.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.587378/full#supplementary-material

REFERENCES

Akbari, R., Ng, P. K. S., Werner, H. M., Shahmoradgoli, M., Zhang, F., Ju, Z., et al. (2015). Corrigendum: a pan-cancer proteomic perspective on the Cancer Genome Atlas. Nat. Commun. 6:5852. doi: 10.1038/ncomms5852
Andersen, P. K., and Gill, R. D. (1982). Cox’s regression model for counting processes: a large sample study. Ann. Stat. 10, 1100–1120. doi: 10.1214/aos/1176349767
Athey, S., Tibshirani, J., and Wager, S. (2018). Generalized Random Forests. Technical report. Stanford, CA: Stanford University.
Biau, G. (2012). Analysis of a random forests model. J. Mach. Learn. Res. 13, 1063–1095.
Biau, G., Devroye, L., and Lugosi, G. (2008). Consistency of random forests and other averaging classifiers. J. Mach. Learn. Res. 9, 2015–2033.
Biau, G., and Scornet, E. (2016). A random forest guided tour. Test 25, 197–227. doi: 10.1007/s11749-016-0481-7
Breiman, L. (2001). Random forests. Mach. Learn. 45, 5–32. doi: 10.1023/A:1010933404324
Breiman, L. (2004). Consistency for a Simple Model of Random Forests. Technical report 670. Statistics Department, University of California at Berkeley.
Chen, P. Y., and Tsaiats, A. A. (2001). Causal inference on the difference of restricted mean lifetime between two groups. Biometrics 57, 1030–1038. doi: 10.1111/j.0006-341X.2001.1030.x
Cox, D. (1972). Regression models and life-tables. J. R. Stat. Soc. Ser. B 34, 187–220. doi: 10.1111/j.2517-6161.1972.tb00899.x
Cox, D. (1975). Partial likelihood. Biometrika 62, 269–276. doi: 10.1093/biomet/62.2.269
Cutler, D., Edwards, T. C., Beard, K., Cutler, A., Hess, K., Gibson, J., and Lawler, J. (2007). Random forests for classification in ecology. Ecology 8811, 2783–2792.
Davison, A., and Hinkley, D. V. (1997). Bootstrap Methods and Their Application. Cambridge University Press. Available online at: https://www.cambridge.org/core/books/bootstrap-methods-and-their-application/
Liu and Li

Survival Random Forest

ED2FD043579F27952363566DC09CBDD6A. doi: 10.1017/CBO9780511802843

Denil, M., Matheson, D., and De Freitas, N. (2014). “Narrowing the gap: random forests in theory and in practice,” in Proceedings of The 31st International Conference on Machine Learning, 665–673. Available online at: http://proceedings.mlr.press/v32/denil14.html

Diaz-Uriarte, R., and Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. BMC Bioinformatics 7:3. doi: 10.1186/1471-2105-7-3

Erfon, B. (1980). The Jackknife, the Bootstrap, and Other Resampling Plans. Available online at: https://statistics.stanford.edu/research/jackknife-bootstrap-and-other-resampling-plans

Fang, E. X., Ning, Y., and Liu, H. (2017). Testing and confidence intervals for high dimensional proportional hazards model. J. R. Stat. Soc. Ser. B 79, 1415–1437. doi: 10.1111/rssb.12224

Friedberg, R., Tibshirani, J., Athey, S., and Wager, S. (2018). Local linear forests. J. Comput. Graph. Stat. 1–25. doi: 10.1080/10618600.2020.1831930

Gill, R. D., and Gill, R. D. (1984). Understanding Cox’s regression model: a martingale approach. J. Am. Stat. Assoc. 79, 441–447. doi: 10.1080/01621459.1984.10478069

Hastie, T., Tibshirani, R. and Friedman, J. (2001). The Elements of Statistical Learning. New York, NY: Springer New York Inc. Available online at: https://www.bibsonomy.org/bibtex/2f58afc5c9793fcc8ad8389824e57984c/sb3000. doi: 10.1007/978-0-387-84858-7

Huang, J., Sun, T., Ying, Z., Yu, Y., and Zhang, C. H. (2013). Oracle inequalities for high dimensional proportional hazards model. J. R. Stat. Soc. Ser. B 79, 1415–1437. doi: 10.1111/rssb.12224

Ishwaran, H., Kogalur, U. B. (2011). Consistency of random survival forests. Stat. Probab. Lett. 80, 1056–1064. doi: 10.1016/j.spl.2010.02.020

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., and Lauer, M. S. (2008). Random survival forests. Ann. Appl. Stat. 2, 841–860. doi: 10.1214/08-AOAS169

Meinshausen, N. (2006). Quantile regression forests. J. Mach. Learn. Res. 7, 983–999.

Mentch, L., and Hooker, G. (2016). Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. J. Mach. Learn. Res. 17, 261–264.

Royston, P., and Parmar, M. K. R. (2013). Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med. Res. Methodol. 13:152. doi: 10.1186/1471-2288-13-152

Sexton, J., and Laake, P. (2009). Standard errors for bagged and random forest estimators. Comput. Stat. Data Anal. 53, 801–811. doi: 10.1016/j.csda.2008.07.007

Steingrimsson, J. A., Diao, L., and Strawderman, R. L. (2019). Censoring unbiased regression trees and ensembles. J. Am. Stat. Assoc. 114, 370–383. doi: 10.1080/01621459.2017.1407775

Zucker, D. M. (1998). Restricted mean life with covariates: modification and extension of a useful survival analysis method. Biometrics 54, 702–709. doi: 10.1111/j.1541-0420.1998.02843.00073.x

Zucker, D. M. (1998). Restricted mean life with covariates: modification and extension of a useful survival analysis method. Biometrics 54, 702–709. doi: 10.1111/j.1541-0420.1998.02843.00073.x

Copyright © 2021 Liu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.