Supplementary Materials to "Interval censored recursive forests"

Hunyong Cho
Department of Biostatistics,
University of North Carolina at Chapel Hill,

Nicholas P. Jewell
Department of Medical Statistics & Centre for Statistical Methodology,
London School of Hygiene & Tropical Medicine,

and

Michael R. Kosorok
Department of Biostatistics,
Department of Statistics and Operations Research,
University of North Carolina at Chapel Hill

May 19, 2021

1 Proof of GWRS consistency

We prove Theorem 1 (consistency of GWRS).

$$\begin{aligned}
&|W_n(S_n) - \theta(S_0)| \\
=&|W_n(S_n) - W_n(S_0)| + |W_n(S_0) - \theta(S_0)| \\
=&|(1A)| + |(1B)|,
\end{aligned}$$

where in what follows we show each term is $o_P(1)$.

$$\begin{aligned}
|(1A)| =&\left| \frac{1}{n_1 n_2} \sum_{i \in G_1} \sum_{j \in G_2} \zeta(I_{1,i}, I_{2,j} | X_{1,i}, X_{2,j}; S_n) - \zeta(I_{1,i}, I_{2,j} | X_{1,i}, X_{2,j}; S_n) \right| \\
=&\left| \frac{1}{n_1 n_2} \sum_{i \in G_1} \sum_{j \in G_2} \Pr(\mathring{T}_{1,i} < \mathring{T}_{2,i} | \mathring{T}_{1,i} \in I_{1,i}, \mathring{T}_{2,i} \in I_{2,i}, X_{1,i}, X_{2,i}; S_n) \right. \\
&\qquad + \frac{1}{2} \Pr(\mathring{T}_{1,i} = \mathring{T}_{2,i} | \mathring{T}_{1,i} \in I_{1,i}, \mathring{T}_{2,i} \in I_{2,i}, X_{1,i}, X_{2,i}; S_n) \\
&\qquad - \Pr(\mathring{T}_{1,i} < \mathring{T}_{2,i} | \mathring{T}_{1,i} \in I_{1,i}, \mathring{T}_{2,i} \in I_{2,i}, X_{1,i}, X_{2,i}; S_0) \\
&\qquad \left. - \frac{1}{2} \Pr(\mathring{T}_{1,i} = \mathring{T}_{2,i} | \mathring{T}_{1,i} \in I_{1,i}, \mathring{T}_{2,i} \in I_{2,i}, X_{1,i}, X_{2,i}; S_0) \right| \\
=&\frac{1}{n_1 n_2} \left| \sum_{i \in G_1} \sum_{j \in G_2} \int_0^\tau \{\check{S}_n(t | I_{1,i}, X_{1,i}) - \check{S}_0(t | I_{1,i}, X_{1,i})\} dS_n(t | I_{2,i}, X_{2,i}) \right. \\
&\qquad - \frac{1}{2} \{S_n(\tau | I_{1,i}, X_{1,i}) - \check{S}_0(\tau | I_{1,i}, X_{1,i})\} S_n(\tau | I_{2,i}, X_{2,i}) \\
&\qquad + \int_0^\tau \check{S}_0(t | I_{1,i}, X_{1,i}) d\{S_n(t | I_{2,i}, X_{2,i}) - S_0(t | I_{2,i}, X_{2,i})\} \\
&\qquad \left. - \frac{1}{2} S_0(\tau | I_{1,i}, X_{1,i}) \{S_n(\tau | I_{2,i}, X_{2,i}) - S_0(\tau | I_{2,i}, X_{2,i})\} \right| \\
\le&\frac{1}{n_1 n_2} \sum_{i \in G_1} \sum_{j \in G_2} \sup_t \left| \check{S}_n(t | I_{1,i}, X_{1,i}) - \check{S}_0(t | I_{1,i}, X_{1,i}) \right| \\
&\qquad + \sup_t \left| S_n(t | I_{2,i}, X_{2,i}) - S_0(t | I_{2,i}, X_{2,i}) \right|
\end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{n_1 n_2}\bigg| \int_0^\tau \sum_{i\in G_1}\{\check S_n(t|I_{1,i},X_{1,i})-\check S_0(t|I_{1,i},X_{1,i})\}d\sum_{j\in G_2} S_n(t|I_{2,i},X_{2,i})\\
&\qquad -\frac{1}{2}\sum_{j\in G_1}\{S_n(\tau|I_{1,i},X_{1,i})-\check S_0(\tau|I_{1,i},X_{1,i})\}\sum_{j\in G_2} S_n(\tau|I_{2,i},X_{2,i})\\
&\qquad +\int_0^\tau \sum_{j\in G_1}\check S_0(t|I_{1,i},X_{1,i})d\sum_{j\in G_2}\{S_n(t|I_{2,i},X_{2,i})-S_0(t|I_{2,i},X_{2,i})\}\\
&\qquad -\frac{1}{2}\sum_{j\in G_1} S_0(\tau|I_{1,i},X_{1,i})\sum_{j\in G_2}\{S_n(\tau|I_{2,i},X_{2,i})-S_0(\tau|I_{2,i},X_{2,i})\}\bigg|\\
\le&\sup_t\bigg|\frac{1}{n_1}\sum_{i\in G_1}\{\check S_n(t|I_{1,i},X_{1,i})-\check S_0(t|I_{1,i},X_{1,i})\}\bigg|\\
&\qquad +\sup_t\bigg|\frac{1}{n_2}\sum_{i\in G_2}\{S_n(t|I_{2,i},X_{2,i})-S_0(t|I_{2,i},X_{2,i})\}\bigg|.
\end{aligned}
$$

We further show that $\sup_{t\in[0,\tau]}\left|\frac{\mathbb{P}_n\check S_n(t|I,X)1(X\in G_l)}{\mathbb{P}_n 1(X\in G_l)}-\frac{\mathbb{P}_n\check S_0(t|I,X)1(X\in G_l)}{\mathbb{P}_n 1(X\in G_l)}\right| = o_P(1), l=1,2,$ where $\frac{\mathbb{P}_n\check S_n(t|I,X)1(X\in G_l)}{\mathbb{P}_n 1(X\in G_l)}=\frac{1}{n_1}\sum_{i\in G_1}\check S_n(t|I_{1,i},X_{1,i})$.

$$
\begin{aligned}
&\sup_{t\in[0,\tau]}\left|\frac{\mathbb{P}_n\check S_n(t|I,X)1(X\in G_l)}{\mathbb{P}_n 1(X\in G_l)}-\frac{\mathbb{P}_n\check S_0(t|I,X)1(X\in G_l)}{\mathbb{P}_n 1(X\in G_l)}\right| && (1)\\
=&\sup_t\frac{\mathbb{P}_n\{(\check S_n(t|X,I)-\check S_0(t|X,I)\}1(X\in G_l)}{P1(X\in G_l)(1+o_P(1))} && (2)\\
\le&\sup_{t,x}|\mathbb{P}_n\{\check S_n(t|X,I)-\check S_0(t|X,I)\}|\lambda_l^{-1}(1+o_P(1))\\
\le&\sup_{t,x}\left|\mathbb{P}_n 1(t\in[L,R))\left\{\frac{\check S_0(t|X)-\check S_0(R|X)}{\check S_0(L|X)-\check S_0(R|X)}-\frac{\check S_n(t|X)-\check S_n(R|X)}{\check S_n(L|X)-\check S_n(R|X)}\right\}\right|\\
\le&\sup_{t,x}\left|\mathbb{P}_n\frac{1(t\in[L,R))}{\check S_0(L|X)-\check S_0(R|X)}\left\{\check S_0(t|X)-\check S_0(R|X)-\check S_n(t|X)+\check S_n(R|X)\right\}\right|\\
&+\sup_{t,x}\left|\mathbb{P}_n\frac{1(t\in[L,R))}{\check S_0(L|X)-\check S_0(R|X)}\left\{\check S_0(L|X)-\check S_0(R|X)-\check S_n(L|X)+\check S_n(R|X)\right\}\right|\\
\le&4\left|\sup_{t,x}\left\{\check S_n(t|x)-\check S_0(t|x)\right\}\right|\sup_t\mathbb{P}_n\frac{1(t\in[L,R))}{\check S_0(L|X)-\check S_0(R|X)}\\
\le&4\left|\sup_{t,x}\left\{\check S_n(t|x)-\check S_0(t|x)\right\}\right|(1+o_P(1))\\
=&o_P(1),
\end{aligned}
$$

where the last inequality is due to

$$
\begin{aligned}
&\mathbb{P}_n \frac{1(t \in [L, R))}{\check{S}_0(L|X) - \check{S}_0(R|X)} \\
=&\mathbb{P}_n \frac{1(t \in [L, R))}{S_0(L|X) - S_0(R|X)} && \text{for continuous } S_0, \\
=&\int \frac{1(l \leq t < r))}{\Pr(l \leq T < r|X = x)} d\mathbb{P}_n(l, r, x) \\
=&\int \frac{1(l \leq T < r)}{\Pr(l \leq T < r|X = x)} \frac{1(l \leq t < r)}{1(l \leq T < r)} d\mathbb{P}_n(T, l, r, x) \\
\leq&\sqrt{\int \frac{1(l \leq T < r)}{\Pr(l \leq T < r|X = x)} d\mathbb{P}_n(T, l, r, x)} \sqrt{\int \frac{1(l \leq t < r)}{1(l \leq T < r)} d\mathbb{P}_n(T, l, r, x)} \\
\leq&\sqrt{\int \underbrace{\int \frac{1(l \leq T < r)}{\Pr(l \leq T < r|X = x)} d\mathbb{P}_n(T, l, r|X = x)}_{=1+o_P(1)} d\mathbb{P}_{nX}(x)} \\
&\quad \times \sqrt{\int \frac{1(l \leq t < r)}{1(l \leq T < r)} d\mathbb{P}_n(T, l, r, x)} && \text{with the denominator} \leq 1 \text{ with probability 1.} \\
\leq&1 + o_P(1).
\end{aligned}
$$

Now we show $(1B) = o_P(1)$ to conclude proof of Theorem 1.

$$
\begin{aligned}
|(1B)| =&|W_n(S_0) - \theta(S_0)| \\
=&\Bigg| \int_0^\tau \left\{ \underbrace{\frac{1}{n_1} \sum_{i \in G_1}}_{=:\mathbb{P}_{n,1}} S_0(t|I_{1,i}, X_{1,i}) \right\} d \left\{ \underbrace{\frac{1}{n_2} \sum_{j \in G_2}}_{=:\mathbb{P}_{n,2}} S_0(t|I_{2,i}, X_{2,i}) \right\} \\
&\qquad - \frac{1}{2} \left\{ \frac{1}{n_1} \sum_{j \in G_1} S_0(\tau|I_{1,i}, X_{1,i}) \right\} \left\{ \frac{1}{n_2} \sum_{j \in G_2} S_0(\tau|I_{2,i}, X_{2,i}) \right\} \\
&- \int_0^\tau S_0(t|X \in G_1) dS_0(t|X \in G_2) + \frac{1}{2} S_0(\tau|X \in G_1) S_0(\tau|X \in G_2) \Bigg|
\end{aligned}
$$

$$
\begin{aligned}
=&\bigg|\int_0^\tau \{P_{n,1}S_0(t|I,X)-S_0(t|X\in G_1)\}\,dP_{n,2}S_0(t|I,X)\\
&\quad-\frac{1}{2}\{P_{n,1}S_0(\tau|I,X)-S_0(\tau|X\in G_1)\}\,P_{n,2}S_0(\tau|I,X)\\
&\quad+\int_0^\tau S_0(t|X\in G_1)d\{P_{n,2}S_0(t|I,X)-S_0(t|X\in G_2)\}\\
&\quad-\frac{1}{2}S_0(\tau|X\in G_1)\{P_{n,2}S_0(t|I,X)-S_0(\tau|X\in G_2)\}\bigg|\\
\le&\sum_{l=1}^{2}\sup_t\sup_t|P_{n,l}S_0(t|I,X)-S_0(t|X\in G_l)|\\
=&o_P(1).
\end{aligned}
$$

To see the last equality, we have

$$
\begin{aligned}
&\sup_t|P_{n,l}S_0(t|I,X)-S_0(t|X\in G_l)|\\
=&\sup_t\left|\frac{\mathbb{P}_nS_0(t|I,X)1(X\in G_l)}{\mathbb{P}_n1(X\in G_l)}-\frac{PS_0(t|I,X)1(X\in G_l)}{P1(X\in G_l)}\right|\\
\le&\sup_t\left|\frac{(\mathbb{P}_n-P)S_0(t|I,X)1(X\in G_l)}{\mathbb{P}_n1(X\in G_l)}\right|+\sup_t\frac{PS_0(t|I,X)1(X\in G_l)}{P1(X\in G_l)}\left|\frac{(\mathbb{P}_n-P)1(X\in G_l)}{\mathbb{P}_n1(X\in G_l)}\right|\\
=&o_P(1),
\end{aligned}
$$

by the law of large numbers and Slutsky's lemma.

2 Proof of GLR consistency

We prove Theorem 2 that states consistency of the GLR statistic.

Note that $LR_n(S_n)=g(\begin{pmatrix}Y_1(\cdot;S_n)\\Y_2(\cdot;S_n)\end{pmatrix})$ and $\rho(S_0)=g(\begin{pmatrix}S_0(\cdot|G_1)\\S_0(\cdot|G_2)\end{pmatrix})$, where g is a continuous map from $D_{[0,1]}[0,\tau]$ to $\mathbb{R}_+$ and $D_{[0,1]}[0,\tau]$ is the space of cadlag (right-continuous with left-hand limits) functions bounded by 0 and 1 with support $[0,\tau]$. The continuity of g can be shown without difficulty using convergence theorems for integration maps (see, e.g., Proposition 7.27 of Kosorok [2007]). If we show $\sup_{t,l}|Y_l(t;S_n)-S_0(t|G_l)|\to_p 0$, by the functional continuous mapping theorem, $LR_n(S_n)\to_p\rho(S_0)$. Thus, it remains to show $\sup_{t,l}|Y_l(t|S_n)-S_0(t|G_l)|\to_p 0$.

Let $\lambda_l = \lim_{n\to\infty} \lambda_{n,l}, l = 1, 2$. For each $l = 1, 2$, we have the following decomposition.

$$
\begin{aligned}
& Y_l(t; S_n) - S_0(t|G_l) \\
=& \frac{\mathbb{P}_n S_n(t|X, I)1(X \in G_l)}{\mathbb{P}_n 1(X \in G_l)} - \frac{P S_0(t|X)1(X \in G_l)}{P1(X \in G_l)} \\
=& \frac{(\mathbb{P}_n - P)S_n(t|X, I)1(X \in G_l)}{\mathbb{P}_n 1(X \in G_l)} && (2A) \\
& \quad + \frac{P\{(S_n(t|X, I) - S_0(t|X, I)\}1(X \in G_l)}{\mathbb{P}_n 1(X \in G_l)} && (2B) \\
& \quad + \frac{P\{S_0(t|X, I) - S_0(t|X)\}1(X \in G_l)}{\mathbb{P}_n 1(X \in G_l)} && (2C) \\
& \quad + PS_0(t|X)1(X \in G_l) \left\{ \frac{1}{\mathbb{P}_n 1(X \in G_l)} - \frac{1}{P1(X \in G_l)} \right\}. && (2D)
\end{aligned}
$$

In (2A), since $S_n(t|X, I)$ is a stochastic process that is monotone in t, by Lemma 9.10 of Kosorok [2007], this process $\{S_n(t|X, I) : t\}$ has VC-dimension 2 and, thus, is a Glivenko-Cantelli class. Also since any finite number of fixed sets form a Glivenko-Cantelli class and any collection of elementwise products of Glivenko-Cantelli classes that are bounded are again a Glivenko-Cantelli class, $\{S_n(t|X, I)1(X \in G_l) : t, l = 1, 2\}$ is Glivenko-Cantelli. Thus by the Glivenko-Cantelli Theorem,

$$
\sup_t |(1A)| \leq \{(\mathbb{P}_n - P)S_n(t|X, I)1(X \in G_l)\} \lambda_l^{-1}(1 + o_P(1)) \to 0,
$$

where we used the fact that $\mathbb{P}_n 1(X \in G_l) = \lambda_l(1 + o_P(1))$ and the numerator being asymptotically bounded by twice the denominator in absolute values.

$$
\begin{aligned}
\sup_t |(2B)| =& \sup_t \frac{P\{(S_n(t|X,I) - S_0(t|X,I)\}1(X \in G_l)}{P1(X \in G_l)}(1 + o_P(1)) \qquad (3)\\
\leq& \sup_t |P\{S_n(t|X,I) - S_0(t|X,I)\}|\lambda_l^{-1}(1 + o_P(1))\\
\leq& \sup_t \left| P1(t \in [L,R)) \left\{ \frac{S_0(t|X) - S_0(R|X)}{S_0(L|X) - S_0(R|X)} - \frac{S_n(t|X) - S_n(R|X)}{S_n(L|X) - S_n(R|X)} \right\} \right|\\
\leq& \sup_t \left| P \frac{1(t \in [L,R))}{S_0(L|X) - S_0(R|X)} \{S_0(t|X) - S_0(R|X) - S_n(t|X) + S_n(R|X)\} \right|\\
&+ \sup_t \left| P \frac{1(t \in [L,R))}{S_0(L|X) - S_0(R|X)} \{S_0(L|X) - S_0(R|X) - S_n(L|X) + S_n(R|X)\} \right|\\
\leq& 4 \left| \sup_{t,x} \{S_n(t|x) - S_0(t|x)\} \right| \sup_t P \frac{1(t \in [L,R))}{S_0(L|X) - S_0(R|X)}\\
\leq& 4 \left| \sup_{t,x} \{S_n(t|x) - S_0(t|x)\} \right|\\
=& o_P(1),
\end{aligned}
$$

where the last inequality is due to

$$
\begin{aligned}
& P \frac{1(t \in [L,R))}{S_0(L|X) - S_0(R|X)}\\
=& \int \frac{1(l \leq t < r))}{\Pr(l \leq T < r|X = x)} dP(l,r,x)\\
=& \int \frac{1(l \leq T < r)}{\Pr(l \leq T < r|X = x)} \frac{1(l \leq t < r)}{1(l \leq T < r)} dP(T,l,r,x)\\
\leq& \sqrt{\int \frac{1(l \leq T < r)}{\Pr(l \leq T < r|X = x)} dP(T,l,r,x)} \sqrt{\int \frac{1(l \leq t < r)}{1(l \leq T < r)} dP(T,l,r,x)}\\
\leq& \sqrt{\int \underbrace{\int \frac{1(l \leq T < r)}{\Pr(l \leq T < r|X = x)} dP(T,l,r|X = x)}_{=1} dP_X(x)}\\
& \times \sqrt{\int \frac{1(l \leq t < r)}{1(l \leq T < r)} dP(T,l,r,x)} \qquad \text{with the denominator} \leq 1 \text{ with probability 1.}\\
\leq& 1.
\end{aligned}
$$

$$
\begin{aligned}
\sup_t |(2C)| =& \sup_t \left\{ P\{S_0(t|X,I) - S_0(t|X)\}1(X \in G_l) \right\} \lambda_l^{-1}(1 + o_P(1)) \\
\leq& \sup_t \left\{ P\{S_0(t|X,I) - S_0(t|X)\} \right\} \lambda_l^{-1}(1 + o_P(1)) \\
=& 0,
\end{aligned}
$$

where the last equality is from the fact that both $S_0(t|X,I)$ and $S_0(t|X)$ can be written as expectation with only difference in the conditioning argument that are marginalized out by the population average operator P (the double expectation). Finally,

$$
\begin{aligned}
\sup_t |(2D)| =& \sup_t PS_0(t|X)1(X \in G_l) \left\{ \frac{1}{\mathbb{P}_n 1(X \in G_l)} - \frac{1}{P1(X \in G_l)} \right\} \\
\leq& P1(X \in G_l) \left\{ \frac{1}{\mathbb{P}_n 1(X \in G_l)} - \frac{1}{P1(X \in G_l)} \right\} \\
\leq& \frac{P1(X \in G_l) - \mathbb{P}_n 1(X \in G_l)}{\mathbb{P}_n 1(X \in G_l)} \\
=& o_P(1).
\end{aligned}
$$

Therefore, the desired result holds.

3 Proof of uniform consistency of interval censored recursive forests

3.1 Overview of the proof of Theorem 3

It suffices to prove the theorem for a single iteration, because, for a large sample, terminal node size becomes arbitrarily small with the potential splitting bias being eliminated and as a result recursion does not add to bias reduction.

We borrow the strategy used in Cho et al. [2020] in establishing the uniform consistency of random survival forests, which uses empirical process theory for right censored data. There is a unique challenge in applying the approach to interval censored survival regression problems—namely the identifiability issue. In Cho et al. [2020], the Z-estimator theorem (Theorem 2.10) in Kosorok [2007] could be used without such an issue, since the self-consistency algorithm gives a unique solution for right-censored data. However, for interval censored data, the self-consistency algorithm may not identify the global maximum of the likelihood. Thus, a careful handling of the identifiability condition is required.

The main technique used to guarantee identifiability is to restrict the class of candidate survival functions to those which satisfy the identifiability condition given data. If one can show that the unrestricted class of the estimating equations is Glivenko-Cantelli [Kosorok, 2007, van der Vaart and Wellner, 2013], the resulting theoretical property—uniform convergence of empirical processes—is inherited to smaller, restricted classes. This is true even if the restriction is done in a data-dependent fashion, since any subset of a Glivenko-Cantelli class is also a Glivenko-Cantelli class. In this way, the desired result, or uniform consistency, can be established.

Noting that NPMLEs have uniform-over-time consistency [Groeneboom and Wellner, 1992] in the non-regression context and that the unique NPMLEs can be estimated through the iterative convex minorant (ICM) algorithms [Groeneboom, 1991, Jongbloed, 1998, Wellner and Zhan, 1997], the problem now reduces to incorporating the identifiability restriction into the estimating equation and extending the uniform consistency results to the regression context.

3.2 The Z-estimator framework

Now we give a detailed proof of Theorem 3 with an introduction to some basic notation for self-consistency equations for non-regression settings and extend the notation to regression settings. The self-consistency equation, without covariates, for case-II censoring with two monitoring times can be expressed as

$$\mathbb{P}_n \psi_{S,t}^{(m)} = 0 \quad \forall t \in [0, \tau], \tag{4}$$

where we put superscript (m) to denote that this is for marginal, or non-regression, settings, $\psi_{S,t}^{(m)} \equiv \eta_1 \frac{S(t)-S(U)}{1-S(U)} \vee 0 \; + \; \left(\eta_2 \frac{S(t)-S(V)}{S(U)-S(V)} \vee 0\right) \wedge 1 \; + \eta_3 \frac{S(t)}{S(V)} \wedge 1 - S(t)$, U and V are the ordered monitoring times, $\eta_1 = 1(T \leq U), \eta_2 = 1(U < T \leq V)$, $\eta_3 = 1 - \eta_1 - \eta_2$, $\wedge$ and $\vee$ are the minimum and the maximum operators, and $\mathbb{P}_n$ is the empirical measures of given data of size n. $\mathbb{P}_n$ is, at the same time, used to denote the sample average operator such that, given a function $f : \mathcal{X} \mapsto \mathbb{R}$ that maps the sample space to the real space, $\mathbb{P}_n f = \int f(x) d\mathbb{P}_n(x) = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$, where X_i is the ith random entry of the data, or $X_i = (U_i, V_i, \eta_{1,i}, \eta_{2,i})$ in this specific problem.

The following lemma is a restatement of Theorem 2.10 of Kosorok [2007]. In the survival regression setting, we let $\Psi : \Theta \mapsto \mathbb{L}$ be a map between two normed spaces, where Θ is the space of all marginal survival functions with time ranging over $[0, \tau]$, $\mathbb{L}$ is a normed space of right-coninuous-over-time functions with support $[0, \tau]$ and range $[-1, 1]$, $\|\cdot\|_{\mathbb{L}}$ denotes the uniform norm over $[0, \tau]$, Ψ is a fixed map, and Ψ_n is a data-dependent map.

Lemma A1 (Consistency of Z-estimators)**.** *Let $\Psi(S_0) = 0$ for some $S_0 \in \Theta$, and assume $\|\Psi(S_n)\|_{\mathbb{L}} \to 0$ implies $\|S_n - S_0\|_{\mathbb{L}} \to 0$ for any sequence $\{S_n\} \in \Theta$. Then, if $\|\Psi_n(\hat{S}_n)\|_{\mathbb{L}} \to 0$ in probability for some sequence of estimators $\hat{S}_n \in \Theta$ and $\sup_{S \in \Theta} \|\Psi_n(S) - \Psi(S)\|_{\mathbb{L}} \to 0$ in probability, $\|\hat{S}_n - S_0\|_{\mathbb{L}} \to 0$ in probability, as $n \to \infty$.*

Now we adapt the lemma to address the identifiability issue by introducing a necessary and sufficient condition [Gentleman and Geyer, 1994] for an NPMLE S to be unique:

$$\mathbb{P}_n \phi_{S,t}^{(m)} \leq 1 \quad \forall t \in [0, \tau], \tag{5}$$

where $\phi_{S,t}^{(m)} = \eta_1 \frac{1(t \leq U)}{1 - S(U)} + \eta_2 \frac{1(U < t \leq V)}{S(U) - S(V)} + \eta_3 \frac{1(t > V)}{S(V)}$. This condition guarantees that the NPMLE $\hat{S}$ that satisfies $\mathbb{P}_n \phi_{S,t}^{(m)} \leq 1$ is the global maximum and thus identifies the true S_0 at its limit given self-consistency. Thus, if we restrict the space to $\Theta_n = \{S : \sup_{t \in [0,\tau]} \mathbb{P}_n \phi_{S,t}^{(m)} \leq 1\}$, within the resticted space, only the unique NPMLE $S = \hat{S}_n$ satisfies the estimating equation (4). This space Θ_n is adaptively defined as it depends on a specific data set. This sequence of spaces always exists, because, given data, the NPMLE can be uniquely estimated via the ICM algorithm. Now the lemma is adapted in the following corollary to reflect the restriction and to be further used for the regression setting.

Corollary A1 (Consistency of Z-estimators)**.** *Let Θ be a class of all covariate-conditional survival functions $S : [0, \tau] \times \mathcal{X} \mapsto [0, 1]$ and let $\Psi : \Theta \mapsto \mathbb{L}$ where $\mathbb{L}$ is some normed space* *of functions $S : [0,\tau] \times \mathcal{X} \mapsto [-1,1]$. (i) Let $\Psi(S_0) = 0$ for some $S_0 \in \Theta$. (ii) Assume that there exists a sequence of subclasses Θ_n such that for any sequence $\{S_n \in \Theta_n\}$, $\|\Psi(S_n)\|_{\mathbb{L}} \to 0$ implies $\|S_n - S_0\|_{\mathbb{L}} \to 0$. (iii) Further assume that $\|\Psi_n(\hat{S}_n)\|_{\mathbb{L}} \to 0$ in probability for some sequence of estimators $\{\hat{S}_n \in \Theta_n\}$. Then, if (iv) $\sup_{S\in\Theta_n} \|\Psi_n(S) - \Psi(S)\|_{\mathbb{L}} \to 0$ in probability, $\|\hat{S}_n - S_0\|_{\mathbb{L}} \to 0$ in probability, as $n \to \infty$.*

We first define the regression-version estimating equations, Ψ and Ψ_n, as below.

$$\Psi(S) = \frac{P\psi_{S,t}\delta_x}{P\delta_x} \equiv P_{\cdot|x}\psi_{S,t},$$
$$\Psi_n(S) = \frac{\mathbb{P}_n\psi_{S,t}k_x}{\mathbb{P}_n k_x} \equiv \mathbb{P}_{n,\cdot|k_x}\psi_{S,t},$$

where

$$\psi_{S,t} = \eta_1 \frac{S(t|X) - S(U|X)}{1 - S(U|X)} \vee 0 \; + \; \eta_2 \frac{S(t|X) - S(V|X)}{S(U|X) - S(V|X)} \vee 0 \wedge 1 \; + \eta_3 \frac{S(t|X)}{S(V|X)} \wedge 1 - S(t|X), \tag{6}$$

P is the population version of $\mathbb{P}_n$ so that, $Pf = \int f(x)dP(x)$, $\delta_x = I(\cdot = x)$ is the unnormalized Dirac measure, and k_x the unnormalized forest kernel. To be more specific, $k_x = \frac{1}{n_{\text{tree}}} \sum_{b=1}^{n_{\text{tree}}} 1(x \in L_b(x))$, where n_{tree} is the number of trees in the forest, $L_b(x)$ is the terminal node of the bth tree of the forest that contains the point x. We use the term 'unnormalized' to mean that they are not multiplied by the sample (or the population) size. By using the subscripts $\cdot \mid x$ and $\cdot \mid k_x$ we denote the conditional probability measures, where the latter is a probability measure weighted by the kernel k_x.

Note that for ease of theoretical exposition, we assume that the terminal node prediction of random survival forests is given by the NPMLEs of observations weighted by the forest kernels, instead of averaging the NPMLEs of each tree. This kernel weighting approach is often taken in the literature [Athey et al., 2019, Yao et al., 2019] and is equivalent to the average of tree predictions for non-censored mean outcomes. Although for censored data, these two approaches are not equal in general, they can be shown equivalent for $B >> 1$, as the former can be seen as the average of estimates of random subsamples from the kernel-weighted population from which the latter is estimated.

The identifiability condition (5) in the marginal context is now replaced by

$$\frac{\mathbb{P}_n \phi_{S,t} k_x}{\mathbb{P}_n k_x} \leq 1 \quad \forall t \in [0, \tau], x \in \mathbb{R}^d. \tag{7}$$

Note that the kernel k not only depends on given data but also has extra randomness due to subsampling, random variable subsetting, and/or random cut-off selection. In other words, given data, k is formed as a result of a realized partition of the trees or the forests. Similarly to the restriction done in the marginal setting, the class Θ of covariate-conditional survival curves can also be restricted to $\Theta_{n,k}$ given data and a specific partition (or the kernel k). In other words, $\Theta_{n,k} = \{S : S \in \Theta, \sup_{t \in [0,\tau], x \in \mathcal{X}} \frac{\mathbb{P}_n \phi_{S,t} k_x}{\mathbb{P}_n k_x} \leq 1\}$.

Hence Theorem 3 will follows if we can show that the conditions of Corollary A1 hold. First, that (i) $\Psi(S_0) = 0$ for some $S_0 \in \Theta$ is trivial. The second condition, (ii) existence of a restricted set Θ_n with which $\|\Psi(S_n)\|_{\mathbb{L}} \to 0$ implies $\|S_n - S_0\|_{\mathbb{L}} \to 0$ for any sequence $\{S_n \in \Theta_{n,k}\}$, can be shown to be satisfied by verifying the assumptions of Lemma 2 in Section 3.3. The third condition, (iii), is met, since the kernel-weighed NPMLE is the solution to $\|\Psi_n(\hat{S})n)\|_{\mathbb{L}}$. The last condition, (iv), is checked in Section 3.4 below.

3.3 Uniform identifiability

We introduce additional notation for Lemma 2. Let $\mathcal{Q}$ denote the space of all survival functions on $S : [0, \tau] \mapsto [0, 1]$, $S_0 : \mathcal{X} \mapsto \mathcal{Q}$ denote the true survival functions, and $\mathcal{Q}_0 = \{S_0(x) : x \in \mathcal{X}\}$ be the collection of S_0's. Let $\Phi : \mathcal{Q} \times \mathcal{Q} \mapsto \mathbb{R}$ be the function that takes $S_1, S_2 \in \mathcal{Q}$ and computes the supremum over $[0, \tau]$ of the absolute value of a certain estimating equation, where S_2 is the true survival function and S_1 is the candidate survival functions.

Assumption A1 (Closed covariate space, compact and continuous true survival space)**.** *(i) $\mathcal{X}$ is closed, (ii) $\mathcal{Q}_0$ is compact with respect to the uniform norm on $\mathcal{Q}$, and (iii) for all sequence $\{x_n\} \in \mathcal{X}$ such that $x_n \to x_1$, $\|S_0(x_n) - S_0(x_1)\|_\infty \to 0$.*

Assumption A2 (local identifiability)**.** *For every sequence $S_n^* \in D$ (and also in Θ_n), and every seqeunce $\{x_n\} \in H : x_n \to x_1$, we have that $\Phi(S_n^*, S_0(x_n)) \to 0 \Rightarrow \|S_n^* - S_0(x_1)\|_\infty \to 0$.*

Lemma A2 (uniform identifiability)**.** *Assume Assumptions 1–2. Suppose* $\forall x \in \mathcal{X}$*,* $S_n(x)$ *is a sequence* $\in \mathcal{Q}$ *and suppose* $\sup_{x\in\mathcal{X}} \Phi(S_n(x), S_0(x)) \to 0$*. Then* $\sup_{x\in\mathcal{X}} \|S_n(x) - S_0(x)\|_\infty \to 0$*.*

Proof. Assume that $\sup_{x\in\mathcal{X}} \Phi(S_n(x), S_0(x)) \to 0$ but $\sup_{x\in\mathcal{X}} \|S_n(x) - S_0(x)\|_\infty \not\to 0$. Then there exists a subsequence n' and an associated sequence $x_{n'}$ such that $\|S_{n'}(x_{n'}) - S_0(x_{n'})\|_\infty \to c > 0$. Also, there exists, for this subsequence, n'' such that $x_{n''} \to x_1$ for some $x_1 \in \mathcal{X}$ (by compactness of $\mathcal{Q}_0$).

By Assumption 2, with $S^*_{n''} = S_{n''}(x_{n''})$, we obtain that $\Phi(S^*_{n''}, S_0(x_{n''})) \to 0 \Rightarrow \|S_{n''}(x_{n''}) - S_0(x_{n''})\|_\infty \to 0$. This is a contradiction. Thus, the conclusion of the lemma holds. □

Assumptions 2 and 3 (Lipschitz continuity and bounded and closed covariate space) are sufficient for Assumption A1. Specifically, (ii) is obtained from the Ascoli-Arzelá Theorem. If we show that the interval censored recursive forest satisfies Assumption 2, the result of Lemma 2 holds. While this identifiability result is valid with Φ function, the second condition of Corollary A1 which relies on Ψ function, or the first term of the Φ function, is always satisfied within the restricted class Θ_n. Such a sequence Θ_n of spaces exists, because, given data, the NPMLE can be uniquely estimated via the ICM algorithm.

3.4 Consistency of the estimating function

Finally, we show how the last condition is fulfilled. We decompose the quantity into three components and bound the error.

$$
\begin{aligned}
&\sup_{S\in\Theta_n} \|\Psi_n(S)) - \Psi(S)\|_{\mathbb{L}} \\
\leq & \sup_{S\in\Theta_n, t\in[0,\tau], x\in\mathcal{X}} \left| \frac{\mathbb{P}_n \psi_{S,t} k_x}{\mathbb{P}_n k_x} - \frac{P\psi_{S,t}k_x}{Pk_x} \right| + \sup_{S\in\Theta_n, t\in[0,\tau], x\in\mathcal{X}} \left| \frac{P\psi_{S,t}k_x}{Pk_x} - \frac{P\psi_{S,t}\delta_x}{P\delta_x} \right| \\
\leq & \underbrace{\sup_{S\in\Theta_n, t\in[0,\tau], x\in\mathcal{X}} \left| \frac{(\mathbb{P}_n - P)\psi_{S,t}k_x}{\mathbb{P}_n k_x} \right|}_{=(3A)} + \underbrace{\sup_{S\in\Theta_n, t\in[0,\tau], x\in\mathcal{X}} \left| P\psi_{S,t}k_x \{ \frac{1}{\mathbb{P}_n k_x} - \frac{1}{Pk_x} \} \right|}_{=(3B)} \\
& \quad + \underbrace{\sup_{S\in\Theta_n, t\in[0,\tau], x\in\mathcal{X}} \left| \frac{P\psi_{S,t}k_x}{Pk_x} - \frac{P\psi_{S,t}\delta_x}{P\delta_x} \right|}_{=(3C)}.
\end{aligned}
$$

We use empirical process theory to bound the error of $(3A)$. The class of functions $\{\psi_{S,t} : S \in \Theta_n, t \in [0,\tau]\}$ can be shown to be a Donsker class. To see this, notice that each of the four terms in (6) is a monotone stochastic process and, thus, is a VC class according to Lemma 9.10 of Kosorok [2007]. As a finite sum of VC classes is a VC class and a VC class endowed with a bounded envelope—in this case $F = 1$—is Donsker, the class of $\psi_{S,t}$ functions is a Donsker class. Next, since k_x can be shown to be a Donsker class by Proposition 6 (Bounded entropy integral of the tree and forest kernels) of Cho et al. [2020] and is bounded above by 1, the class of their products $\{\psi_{S,t}\delta_x : S \in \Theta_n, t \in [0,\tau]\}$ is again Donsker. Consequently, $\sup_{S\in\Theta_n, t\in[0,\tau]} \left| (\mathbb{P}_n - P)\psi_{S,t}k_x \right| = O_P(n^{-1/2})$. Meanwhile since the denominator $\mathbb{P}_n k_x \asymp n^{\beta-1}$ by the assumption of terminal node size, we have $(3A) = o_P(1)$

$(3B) = o_P(1)$ can also be shown similarly.

$$
\begin{aligned}
(3B) =& \sup_{S\in\Theta_n, t\in[0,\tau], x\in\mathcal{X}} \left| P\psi_{S,t}k_x \{ \frac{1}{\mathbb{P}_n k_x} - \frac{1}{Pk_x} \} \right| \\
=& \sup_{S\in\Theta_n, t\in[0,\tau], x\in\mathcal{X}} \left| \underbrace{\frac{P\psi_{S,t}k_x}{\mathbb{P}_n k_x}}_{=O_P(1)} \frac{(\mathbb{P}_n - P)k_x}{\mathbb{P}_n k_x} \right| \\
=& O_P(1) O_P(n^{-1/2-(\beta-1)}) = o_P(1).
\end{aligned}
$$

Finally, we show $(3C) = o_P(1)$. We first note that

$$\begin{aligned}
&P_{\cdot\,|k_x}\psi_{S,x} \\
&= 1 - G_U(t|k_x) - \int_{u=t}^{\infty}\{1 - S_0(u|k_x)\}\frac{1 - S(t|x)}{1 - S(u|x)}dG_U(u|k_x) \\
&+ \int_{u=0}^{t}\int_{v=t}^{\infty} S_0(u|k_x)dG(u,v|k_x) + \int_0^t \frac{S(t|x)}{S(v|x)}S_0(v|k_x)dG_V(v|k_x) \\
&- \int_{u=0}^{t}\int_{v=t}^{\infty}\frac{S(u|x) - S(t|x)}{S(u|x) - S(v|x)}(S_0(u|k_x) - S_0(v|k_x))dG(u,v|k_x) - S(t|x) \\
&= 1 - G_U(t|k_x) - S(t|x) \\
&- \int_{u=t}^{\infty} R_1(t,u,x)\{1 - S_0(u|k_x)\}dG_U(u|k_x) \\
&+ \int_{u=0}^{t}\int_{v=t}^{\infty} S_0(u|k_x)dG(u,v|k_x) + \int_0^t R_2(t,v,x)S_0(v|k_x)dG_V(v|k_x) \\
&- \int_{u=0}^{t}\int_{v=t}^{\infty} R_3(t,u,v,x)(S_0(u|k_x) - S_0(v|k_x))dG(u,v|k_x),
\end{aligned}$$

where $0 \le R_1(t,u,x), R_2(t,v,x), R_3(t,u,v,x) \le 1$ are decreasing in u and increasing in v. Thus,

$$\begin{aligned}
(3C) &= \sup_{S\in\Theta, x\in\mathcal{X}} |P_{\cdot\,|k_x}\psi_{S,x} - P_{\cdot\,|x}\psi_{S,x}| \\
&= \sup_{S\in\Theta, x\in\mathcal{X}} \underbrace{\left|\int_{u=t}^{\infty} R_1(t,u,x)\{1 - S_0(u|k_x)\}dG_U(u|k_x) - \int_{u=t}^{\infty} R_1(t,u,x)\{1 - S_0(u|x)\}dG_U(u|x)\right|}_{=(3C1)} \\
&\quad + \sup_{S\in\Theta, x\in\mathcal{X}} \underbrace{\left|\int_{u=0}^{t}\int_{v=t}^{\infty} S_0(u|k_x)dG(u,v|k_x) - \int_{u=0}^{t}\int_{v=t}^{\infty} S_0(u|x)dG(u,v|x)\right|}_{=(3C2)} \\
&\quad + \sup_{S\in\Theta, x\in\mathcal{X}} \underbrace{\left|\int_0^t R_2(t,v,x)S_0(v|k_x)dG_V(v|k_x) - \int_0^t R_2(t,v,x)S_0(v|x)dG_V(v|x)\right|}_{=(3C3)} \\
&\quad + \sup_{S\in\Theta, x\in\mathcal{X}} \underbrace{\begin{aligned}&\left|\int_{u=0}^{t}\int_{v=t}^{\infty} R_3(t,u,v,x)(S_0(u|k_x) - S_0(v|k_x))dG(u,v|k_x)\right. \\ &\quad \left. - \int_{u=0}^{t}\int_{v=t}^{\infty} R_3(t,u,v,x)(S_0(u|x) - S_0(v|x))dG(u,v|x)\right|\end{aligned}}_{=(3C4)}
\end{aligned}$$

We show $\sup_{S\in\Theta, x\in\mathcal{X}}(3C1) = o_P(1)$. Then $(3C2)$–$(3C4)$ can be shown to be $o_P(1)$ using similar arguments.

$$
\begin{aligned}
(3C1) \leq & \left| \int_{u=t}^{\infty} R_1(t,u,x)\{1 - S_0(u|k_x)\} dG_U(u|k_x) - \int_{u=t}^{\infty} R_1(t,u,x)\{1 - S_0(u|x)\} dG_U(u|k_x) \right| \\
& + \left| \int_{u=t}^{\infty} R_1(t,u,x)\{1 - S_0(u|x)\} dG_U(u|k_x) - \int_{u=t}^{\infty} R_1(t,u,x)\{1 - S_0(u|x)\} dG_U(u|x) \right| \\
= & \left| \int_{u=t}^{\infty} \underbrace{R_1(t,u,x)}_{\in[0,1]} \{S_0(u|x) - S_0(u|k_x)\} dG_U(u|k_x) \right| \\
& + \left| \int_{u=t}^{\infty} \underbrace{R_1(t,u,x)\{1 - S_0(u|x)\}}_{\in[0,1]} d\{G_U(u|k_x) - G_U(u|x)\} \right| \\
\leq & \int_{u=t}^{\infty} \sup_{u' \in [0,\infty)} |S_0(u'|x) - S_0(u'|k_x)| dG_U(u|k_x) \\
& + \left| \Big[R_1(t,u,x)\{1 - S_0(u|x)\}\{G_U(u|k_x) - G_U(u|x)\} \Big]_{u=t}^{\infty} \right| \\
& + \left| \int_{u=t}^{\infty} \{G_U(u|k_x) - G_U(u|x)\} d\Big[R_1(t,u,x)\{1 - S_0(u|x)\} \Big] \right| \\
\leq & \sup_{u \in [0,\infty)} \Big| S_0(u|x) - S_0(u|k_x) \Big| + \Big| G_U(t|k_x) - G_U(t|x) \Big| \\
& + \int_{u=t}^{\infty} \sup_{u' \in [0,\infty)} \Big| G_U(u'|k_x) - G_U(u'|x) \Big| d\Big[\underbrace{R_1(t,u,x)\{1 - S_0(u|x)\}}_{\text{increasing in } u, \text{ bounded by 0 and 1.}} \Big] \\
\leq & \sup_{u \in [0,\infty)} \Big| S_0(u|x) - S_0(u|k_x) \Big| + \Big| G_U(t|k_x) - G_U(t|x) \Big| + \sup_{u \in [0,\infty)} \Big| G_U(u|k_x) - G_U(u|x) \Big| \\
& \leq \sup_{x \in \mathcal{X}} \sup_{x' \in k_x} L_S \|x - x'\|_1 + 2 \sup_{x \in \mathcal{X}} \sup_{x' \in k_x} L_G \|x - x'\|_1 \\
& = o_P(1).
\end{aligned}
$$

The last inequality comes from Assumption 2 (Lipschitz continuity) and the subsequent equations result from Assumptions 5 (shrinking terminal node) and 4 (random and regular splits). The derivation for a unit hypercube with a bounded density is given in the proof of Theorem 3 of Wager and Walther [2015].

4 Computational cost

Computation of ICRF is affected by the choice of the splitting and splitting rules, the sample size, and the bandwidth in kernel-smoothing. We discuss the effects of choice of the rules and the sample size from the simulations done in Section 5. Larger bandwidths require more computation having a linear relationship with the number of operations.

First, we discuss the computational cost of ICRF with respect to the splitting and prediction rules. We use the Scenario 1 current status data ($M = 1$) with sample size of $n = 300$, 10 forest iterations, and 300 simulation replicates. First, different splitting rules do not make noticeable differences than having different prediction rules as can be seen in Figure A1. This is because the computationally expensive NPMLEs should be obtained for all n/n_{min} terminal nodes. The computation time for NPMLE is almost a quadratic function of sample size due to its $O(nk)$ EM-iterations and $O(n \log_2 n)$ steps for preprocessing (sorting and indexing), where n is the sample size which is n_{min} in our application and k is the number of Turnbull intervals [Anderson-Bergman, 2017]. Assuming $k \simeq cn_{\text{min}}$ for some constant $c > 0$ and $n_{\text{min}} = n^{\beta}$, the total computational burden of quasi-honest ICRF is $O(n_{\text{fold}} n^{1+\beta})$. In contrast, the exploitative ICRF implements the NPMLE computation only once at the initial step, saving the constant n_{fold} and the set-up cost of $n^{1-\beta}$ many NPMLE calculations.

Compared to existing splitting rules, SWRS and SLR, the new splitting rules cost slightly more computationally. However, as mentioned above, the prediction rule is the predominant determinant of computation over the splitting rule. Given a pair of two samples of sizes $n_l, l = 1, 2$ and k time points of evaluation for numerical integration, the computational burden of GWRS and GLR is $O(kn_1n_2)$ and $O(k(n_1 + n_2))$, respectively. Although GWRS and GLR do not make a large difference in computation time in Figure A1, for large samples, GWRS may be computationally more burdensome.

Figure A2 illustrates the trend of computation time in terms of sample size. The trend is mildly superlinear supporting the total burden of $O(n^{\gamma})$ for some $1 < \gamma \leq 2$.

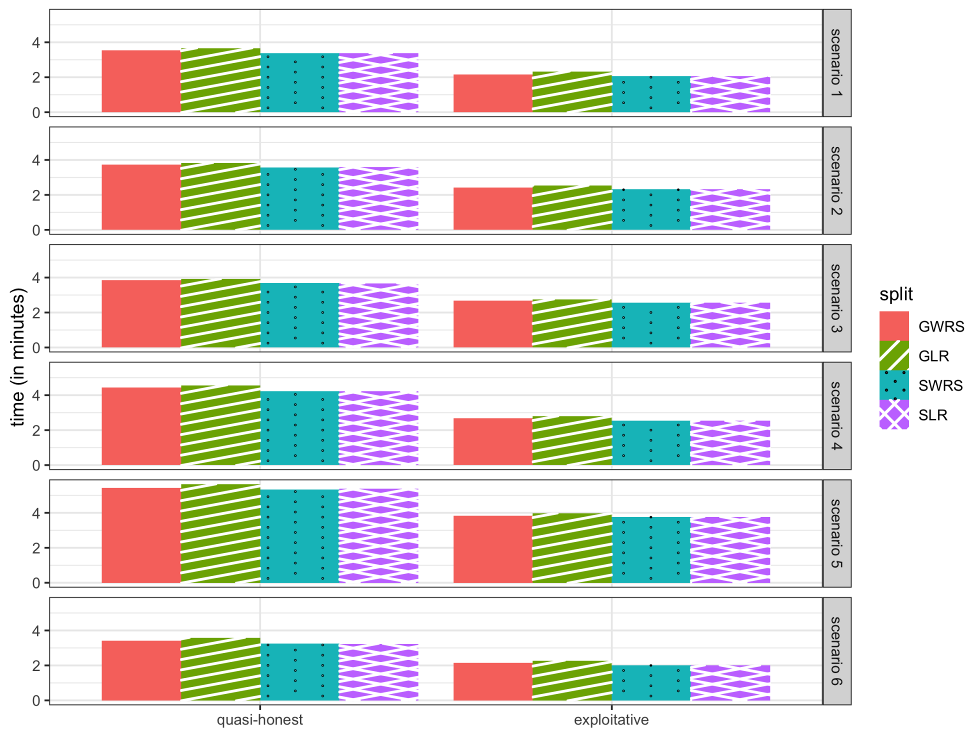

Figure A1: Running time for fitting ICRFs with different splitting and prediction rules.

5 The National Longitudinal Mortality Study data analysis

We use the NLMS dataset with six years of follow-up recorded around April 2002. The dataset includes 745,162 subjects with their time to mortality, demographic information such as age, sex, and race, socioeconomic information such as income and housing tenure, and other covariates. The censoring rate of this dataset is very high (97% survived six years), as this is a general population, and only administrative censoring is observed. We narrow our focus to the aged population (age ≥ 80 in years) with complete covariate records ($n = 3,630$). The proportion of missing data is 20.7% for the whole data and 65.9% for the aged group data. Thus, it should be noted that this data analysis is for performance comparison among the methods and that the results obtained from this regression analysis are limited to the tracked population. The administrative censoring rate is 69.6%, and the distribution of the data is summarized in Table A1.

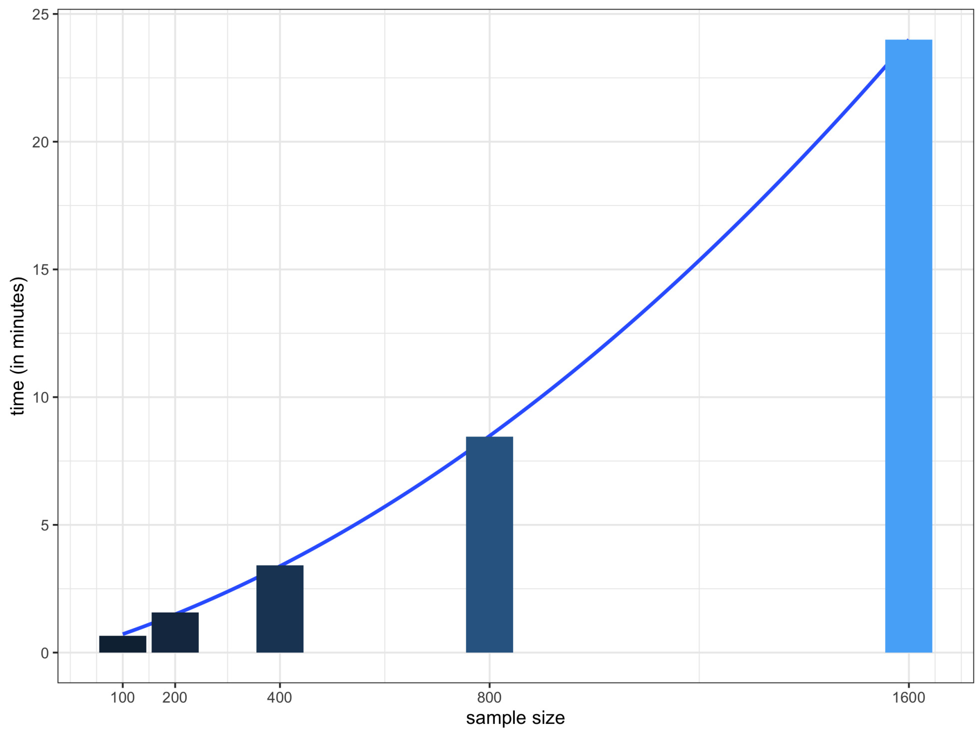

Figure A2: Running time for fitting ICRFs with different sample sizes. The blue curve is a quadratic line that minimizes the mean squared error.

Since the observed failure time is sparse after the follow-up time of 1500, we set $\tau = 1500$. We induce current status censoring where the monitoring time is dependent on age and number of households. The monitoring time is randomly drawn from the model

$$C \sim N(1000 + 100(10 - X_{\text{age}}/10 + X_{\#\text{ households}}), 300^2) \vee 0 \wedge \tau.$$

The analysis framework is largely the same for the avalanche data analysis, except that with the large sample size, the terminal node size is allowed to be larger ($n_{min} = 20$ for random forests and $n_{min} = 40$ for trees).

References

Clifford Anderson-Bergman. An efficient implementation of the emicm algorithm for the interval censored npmle. *Journal of Computational and Graphical Statistics*, 26(2):463–467, 2017.

Susan Athey, Julie Tibshirani, Stefan Wager, et al. Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178, 2019.

Hunyong Cho, Shannon T Holloway, and Michael R Kosorok. Multi-stage optimal dynamic treatment regimes for survival outcomes with dependent censoring. *arXiv preprint arXiv:2012.03294*, 2020.

Robert Gentleman and Charles J Geyer. Maximum likelihood for interval censored data: Consistency and computation. *Biometrika*, 81(3):618–623, 1994.

Piet Groeneboom. Nonparametric maximum likelihood estimators for interval censoring and deconvolution. 1991.

Piet Groeneboom and Jon A Wellner. *Information bounds and nonparametric maximum likelihood estimation*, volume 19. Springer Science & Business Media, 1992.

Geurt Jongbloed. The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational and Graphical Statistics*, 7(3):310–321, 1998.

Michael R Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer Science & Business Media, 2007.

Aad W van der Vaart and Jon Wellner. *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media, 2013.

Stefan Wager and Guenther Walther. Adaptive concentration of regression trees, with application to random forests. *arXiv preprint arXiv:1503.06388*, 2015.

Jon A Wellner and Yihui Zhan. A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *Journal of the American Statistical Association*, 92(439):945–959, 1997.

Weichi Yao, Halina Frydman, and Jeffrey S Simonoff. An ensemble method for interval-censored time-to-event data. *Biostatistics*, 2019.

variable	min	1Q	median	mean	3Q	max
failure time	2	315	633	694.9	1003	1739
censoring	censored 69.6%					
age (years)	81	83	85	84.8	86	90
number of households	1	1	2	1.9	2	13
adjusted weight	14	251	432	471	633	1982
sex	male 34.8%, female 65.2%					
race	white 85.6%, black 10.1%, others and unknown 4.4%					
Hispanic	Mexican 3.4%, other Hispanics 3.6%, non-Hispanics 93.0%					
relationship	A. 25.6%; B. 46.7%; C. 13.6%; D. 13.2%; E. 0.9					
adjusted income	1. 6.6%; 2. 11.4%; 3. 14.1%; 4. 11.6%; 5. 8.5%; 6. 12.6%; 7. 8.6%; 8. 5.7%; 9. 4.1%; 10. 3.5%; 11. 3.8%; 12. 2%; 13. 3.0%; 14. 4.6					
social security number	present 56.9%					
housing tenure	owner 76.4; rent 21.1; non-cash rent 2.5					
health in general	A. 5.1; B. 17.1; C. 32.5; D. 27.8; E. 17.4"					
health insurance type	A. 0.6; B. 75.2; C. 7.1; D. 7.5; E. 9.5					
urban	urban 73.1; rural 26.9					
citizenship	native citizen born in mainland US 88.3%; others 11.7%					

Table A1: The NLMS data. Failure time, non-censored time in days; Relationship, relationship to the reference person (A: reference person with other relatives in household, B. reference person with no other relatives in household, C. spouse of reference person, D. other relative of reference person, E. non-relative of reference person); Adjusted income, 1. $<$ \$5,000, 2. $<$ \$7,499, 3. $<$ \$10,000, 4. $<$, 4. $<$ \$12,500, 5. $<$ \$15,000, 6. $<$ \$20,000, 7. $<$ \$25,000, 8. $<$ \$30,000, 9. $<$ \$35,000, 10. $<$ \$40,000, 11. $<$ \$50,000, 12. $<$ \$60,000, 13. $<$ \$75,000, 14. $\geq$ \$75,000; health in general, A. Excellent, B. Very good, C. Good, D. Fair, E. Poor; health insurance type, A. Medicare, B. Medicaid, C. governmental healthcare (ChampUS, ChampVA, etc), D. employer-based, E. private non-employer-based.