Supplementary Materials

We complement the omitted proofs in Section A. We then provide the additional experimental results in Section B. Furthermore, in Section C, we give supplementary notes for some statements in the main text.

A. Proofs

A.1. Proof of Theorem 4.1

Theorem 4.1 (Provable Guarantee) *Let $\hat{w}(x)$ be an estimate of the weight $w(x)$. Assume that $\mathbb{E}[\hat{w}(X)|\mathcal{Z}_{tr}] = 1$ and $\mathbb{E}[w(X)] = 1$. Denote $\hat{C}_n^1(x)$ as the output band of Algorithm 1 with n calibration samples, then for a new data X', its corresponding survival time satisfies*

$$\lim_{n\to\infty} \mathbb{P}\left(T' \in \hat{C}_n^1(X')\right) \geq 1 - \alpha - \frac{1}{2}\mathbb{E}|\hat{w}(X) - w(X)|,$$

where the probability $\mathbb{P}$ on the left hand side is taken over $(X', T') \sim \mathcal{P}_X \times \mathcal{P}_{T|X}$, and all the expectation operators $\mathbb{E}$ are taken over $X \sim \mathcal{P}_{X|\Delta=1}$.

Firstly, consider a new sample $\left(\tilde{X}', \tilde{T}'\right)$ generated from $\tilde{P}_X \times P_{T|X}$, where we assume that

$$\mathrm{d}\tilde{P}_X(x) = \hat{w}(x)\,\mathrm{d}P_{X|\Delta=1}(x).$$

As a comparison, due to the definition of $w(x)$, we have

$$\mathrm{d}P_X(x) = w(x)\,\mathrm{d}P_{X|\Delta=1}(x).$$

We remark that $\tilde{P}_X(x)$, $P_X(x)$ are indeed distribution since we assume $\mathbb{E}w(X) = \mathbb{E}\hat{w}(X) = 1$, where the expectation is taken over $P_{X|\Delta=1}$.

We would first prove that the probability $\tilde{T}'$ falls in the derived confidence interval $\hat{C}_n^1(X')$ is larger than $1-\alpha$, where α is the given significance level. The intuition is that, since the derived $\hat{C}_n^1(X')$ is derived based on the estimated weight $\hat{w}(x)$, the sample $\tilde{T}'$ is then guaranteed to fall in the confidence interval with probability at least $1-\alpha$.

We derive that

$$\begin{aligned}
&\mathbb{P}\left(\tilde{T}' \in \hat{C}_n^1\left(\tilde{X}'\right)|\mathcal{Z}_{tr}\right)\\
=&\mathbb{P}\left(\tilde{T}' \leq T^u\left(\tilde{X}'\right)|\mathcal{Z}_{tr}\right)\\
=&\mathbb{P}\left(V\left(\tilde{X}', \tilde{T}'\right) \leq V\left(\tilde{X}', T^u\left(\tilde{X}'\right)|\mathcal{Z}_{tr}\right)\right)\\
\geq&\mathbb{P}\left(V\left(\tilde{X}', \tilde{T}'\right) \leq \text{Quantile}\left(1-\alpha; \sum_{i=1}^n \hat{p}_i\delta_{v_i} + \hat{p}_\infty\delta_\infty\right)|\mathcal{Z}_{tr}\right),
\end{aligned} \tag{8}$$

where the last inequality is due to the fact that $V\left(\tilde{X}', T^u\left(\tilde{X}'\right) \geq \text{Quantile}\right)(1-\alpha; \sum_{i=1}^n \hat{p}_i\delta_{v_i} + \hat{p}_\infty\delta_\infty)$. We also use the non-decreasing property of $V(X, T)$ on T. Furthermore, by Lemma 3, we can replace the δ_∞ in the quantile term by $\delta_{V\left(\tilde{X}', \tilde{T}'\right)}$, therefore,

$$\begin{aligned}
&\mathbb{P}\left(V\left(\tilde{X}', \tilde{T}'\right) \leq \text{Quantile}\left(1-\alpha; \sum_{i=1}^n \hat{p}_i\delta_{v_i} + \hat{p}_\infty\delta_\infty\right)|\mathcal{Z}_{tr}\right)\\
=&\mathbb{P}\left(V\left(\tilde{X}', \tilde{T}'\right) \leq \text{Quantile}\left(1-\alpha; \sum_{i=1}^n \hat{p}_i\delta_{v_i} + \hat{p}_\infty\delta_{V\left(\tilde{X}', \tilde{T}'\right)}\right)|\mathcal{Z}_{tr}\right).
\end{aligned} \tag{9}$$

Besides, we know from Lemma 4 that

$$V\left(\tilde{X}', \tilde{T}'\right)|\mathcal{Z}_{tr}, \mathcal{E}(V) \sim \sum_{i=1}^n \hat{p}_i\delta_{v_i} + \hat{p}_\infty\delta_{V\left(\tilde{X}', \tilde{T}'\right)},$$

therefore, we derive that

$$\begin{aligned}
&\mathbb{P}\left(V\left(\tilde{X}',\tilde{T}'\right) \leq \text{Quantile}\left(1-\alpha;\sum_{i=1}^{n}\hat{p}_i\delta_{v_i}+\hat{p}_\infty\delta_{V\left(\tilde{X}',\tilde{T}'\right)}\right)|\mathcal{Z}_{tr}\right)\\
=&\mathbb{E}_{\mathcal{E}}\mathbb{P}\left(V\left(\tilde{X}',\tilde{T}'\right) \leq \text{Quantile}\left(1-\alpha;\sum_{i=1}^{n}\hat{p}_i\delta_{v_i}+\hat{p}_\infty\delta_{V\left(\tilde{X}',\tilde{T}'\right)}\right)|\mathcal{Z}_{tr},\ \mathcal{E}\left(V\right)\right)\\
\geq&1-\alpha.
\end{aligned} \tag{10}$$

Combining the Equation 8, Equation 9 and Equation 10 leads to

$$\mathbb{P}\left(\tilde{T}'\in\hat{C}_n^1\left(\tilde{X}'\right)|\mathcal{Z}_{tr}\right)\geq 1-\alpha. \tag{11}$$

Furthermore, we need to transform the above results into random sample $(X',T')\sim \mathrm{d}P_X\times \mathrm{d}P_{T|X}$. This directly follows Lemma 5 that

$$\left|\mathbb{P}\left(T'\in\hat{C}_n^1\left(X'\right)|\mathcal{Z}_{tr}\right)-\mathbb{P}\left(\tilde{T}'\in\hat{C}_n^1\left(X'\right)|\mathcal{Z}_{tr}\right)\right|\leq d_{TV}\left(P_X\times P_{T|X},\tilde{P}_X\times P_{T|X}\right)=d_{TV}\left(P_X,\tilde{P}_X\right).$$

We can express the total-variation distance between Q_X and $\tilde{Q}_X$ as

$$d_{TV}\left(P_X,\tilde{P}_X\right)=\frac{1}{2}\int|\hat{w}\left(X\right)\mathrm{d}\mathcal{P}_{X|\Delta=1}\left(X\right)-w\left(X\right)\mathrm{d}\mathcal{P}_{X|\Delta=1}\left(x\right)|=\frac{1}{2}\mathbb{E}_{X\sim\mathcal{P}_{X|\Delta=1}}|\hat{w}\left(X\right)-w\left(X\right)|.$$

Combine the above results and take expectations on the training set, we conclude that

$$\begin{aligned}
&\mathbb{P}\left(T'\in\hat{C}_n^1\left(X'\right)\right)\\
=&\mathbb{E}_{\mathcal{Z}_{tr}}\mathbb{P}\left(T'\in\hat{C}_n^1\left(X'\right)|\mathcal{Z}_{tr}\right)\\
\geq&1-\alpha-\frac{1}{2}\mathbb{E}|\hat{w}\left(X\right)-w\left(X\right)|.
\end{aligned} \tag{12}$$

where the expectation is taken over the training set space and $X\sim\mathcal{P}_{X|\Delta=1}$

Technical Lemmas. In this part, we give some technical lemmas used in the proof. We first introduce Lemma 3 which is commonly used in conformal inference. By Lemma 3, we can use δ_∞ to replace $\delta_{V\left(\tilde{X}',\tilde{T}'\right)}$ without changing the probability.

Lemma 3 (Equation(2) in Lemma 1 from Barber et al. (2019a).) *For random variables $v_i\in\mathbb{R}, i\in[n+1]$, let $p_i\in\mathbb{R}, i\in[n+1]$ be the corresponding weights summing to 1. Then for any $\beta\in[0,1]$, we have*

$$V\left(\tilde{X}',\tilde{T}'\right)\leq\text{Quantile}\left(\beta,\sum_{i=1}^{n}p_i\delta_{v_i}+p_\infty\delta_\infty\right)\Longleftrightarrow V\left(\tilde{X}',\tilde{T}'\right)\leq\text{Quantile}\left(\beta,\sum_{i=1}^{n}p_i\delta_{v_i}+p_\infty\delta_{V\left(\tilde{X}',\tilde{T}'\right)}\right). \tag{13}$$

We next introduce Lemma 4 which provides the distribution of the non-conformity score.

Lemma 4 (Equation(A.5) from Lei & Candès (2020).)

$$\left(V\left(\tilde{X}',\tilde{T}'\right)|\mathcal{E}\left(V\right)=\mathcal{E}\left(V^*\right),\mathcal{Z}_{tr}\right)\sim\sum_{i=1}^{n}\hat{p}_i\delta_{v_i^*}+\hat{p}_\infty\delta_{V\left(\tilde{X}',\tilde{T}'\right)}, \tag{14}$$

where $\mathcal{E}\left(V^*\right)$ is the unordered set of $V^*=\left(v_1^*,v_2^*,\ldots,v_n^*,V\left(\tilde{X}',\tilde{T}'\right)\right)$.

Thirdly, we introduce Lemma 5 which shows a basic property of the total variance distance.

Lemma 5 (Equation(10) from Berrett et al. (2020).) *Let $d_{TV}\left(Q_{1X},Q_{2X}\right)$ denote the total-variation distance between Q_{1X} and Q_{2X}, then*

$$d_{TV}\left(Q_{1X}\times P_{T|X},Q_{2X}\times P_{T|X}\right)=d_{TV}\left(Q_{1X},Q_{2X}\right). \tag{15}$$

A.2. Proof of Theorem 5.1

Theorem 5.1 (Lower Bound) *Let $\hat{w}(x)$ be an estimate of the weight $w(x)$, $\hat{q}_{\alpha_{lo}}(x), \hat{q}_{\alpha_{hi}}(x)$ be the quantile estimator returned by WCCI, and $H(X)$ be defined as Equation 7. Assume that $\mathbb{E}[\hat{w}(X)|\mathcal{Z}_{tr}] = 1$ and $\mathbb{E}[w(X)] = 1$, where all the expectation operators $\mathbb{E}$ are taken over $X \sim \mathcal{P}_{X|\Delta=1}$. Denote $\hat{C}_n^2(x)$ as the output band of Algorithm 2 with n calibration samples, and denote X' as the testing point.*

From the weight perspective, under assumptions (A1):

A1. $\mathbb{E}_{X|\Delta=1}|\hat{w}(X) - w(X)| \leq M_1$,
we have:

$$\lim_{n\to\infty} \mathbb{P}\left(T' \in \hat{C}_n^2(X')\right) \geq 1 - \alpha - \frac{1}{2}M_1.$$

From the quantile perspective, under assumptions (B1-B3):

B1. $H(X) \leq M_2$ *a.s. w.r.t.* X*;*
B2. *There exists* $\delta > 0$ *such that* $\mathbb{E}\hat{w}(X)^{1+\delta} < \infty$*;*
B3. *There exists* $\gamma, b_1, b_2 > 0$ *such that* $\mathbb{P}(T = t|X = x) \in [b_1, b_2]$ *uniformly over all (x,t) with* $t \in [q_{\alpha_{lo}}(x) - 2M_2 - 2\gamma, q_{\alpha_{lo}}(x) + 2M_2 + 2\gamma] \cup [q_{\alpha_{hi}}(x) - 2M_2 - 2\gamma, q_{\alpha_{hi}}(x) + 2M_2 + 2\gamma]$*,*
we have:

$$\lim_{n\to\infty} \mathbb{P}\left(T' \in \hat{C}_n^2(X')\right) \geq 1 - \alpha - b_2(2M_2 + \gamma) - \frac{16M_2}{(M_2 + \gamma)^2 b_1}.$$

The proof under (A1) directly follows the proof of Theorem 4.1. Firstly, for a new testing point $(X', T') \sim \mathcal{P}_X \times \mathcal{P}_{T|X}$, we have

$$\begin{aligned}
&\mathbb{P}\left(T' \in \hat{C}_n^2\left(X'\right)|X'\right) \\
=&\mathbb{P}\left(T' \in [\hat{q}_{\alpha_{lo}}\left(X'\right) - \eta, \hat{q}_{\alpha_{hi}}\left(X'\right) + \eta]|X'\right) \\
=&\mathbb{P}\left(\max\{T' - \hat{q}_{\alpha_{hi}}\left(X'\right), \hat{q}_{\alpha_{lo}}\left(X'\right) - T'\} \leq \eta|X\right) \\
\overset{i}{\geq}&\mathbb{P}\left(\max\{T' - q_{\alpha_{hi}}\left(X'\right), q_{\alpha_{lo}}\left(X'\right) - T'\} \leq \eta - H\left(X'\right)|X'\right) \\
\overset{ii}{\geq}&\mathbb{P}\left(\max\{T' - q_{\alpha_{hi}}\left(X'\right), q_{\alpha_{lo}}\left(X'\right) - T'\} \leq -\varepsilon - H\left(X'\right)|X'\right) - \mathbb{P}\left(\eta < -\varepsilon\right) \\
\overset{iii}{\geq}&\mathbb{P}\left(\max\{T' - q_{\alpha_{hi}}\left(X'\right), q_{\alpha_{lo}}\left(X'\right) - T'\} \leq -\varepsilon - H\left(X'\right)\mathbb{I}\left(H\left(X'\right) \leq \varepsilon\right)|X'\right) - \mathbb{I}\left(H\left(X'\right) > \varepsilon\right) - \mathbb{P}\left(\eta < -\varepsilon\right).
\end{aligned} \tag{16}$$

Equation (i) follows from Lemma 6, and Equation (ii) follows from:

$$\begin{aligned}
&\mathbb{P}\left(\max\{T' - q_{\alpha_{hi}}\left(X'\right), q_{\alpha_{lo}}\left(X'\right) - T'\} \leq -\varepsilon - H\left(X'\right)|X'\right) \\
&- \mathbb{P}\left(\max\{T' - \hat{q}_{\alpha_{hi}}\left(X'\right), \hat{q}_{\alpha_{lo}}\left(X'\right) - T'\} \leq \eta - H\left(X'\right)|X'\right) \\
\leq&\mathbb{P}\left(\eta - H\left(X'\right) < \max\{T' - q_{\alpha_{hi}}\left(X'\right), q_{\alpha_{lo}}\left(X'\right) - T'\} \leq -\varepsilon - H\left(X'\right)|X'\right) \\
\leq&\mathbb{P}\left(\eta - H\left(X'\right) < -\varepsilon - H\left(X'\right)|X'\right) \\
=&\mathbb{P}\left(\eta < -\varepsilon\right).
\end{aligned}$$

Equation (iii) can be derived simply based on the discussion on the value of $\mathbb{I}\left(H\left(X'\right) \leq \varepsilon\right)$.

By Assumption (B3), since $-\varepsilon - H\left(X'\right)\mathbb{I}\left(H\left(X'\right) \leq \varepsilon\right) \leq -2\varepsilon$, when $\varepsilon \leq M_2 + \gamma$, we have

$$\begin{aligned}
&\mathbb{P}\left(\max\{T' - q_{\alpha_{hi}}\left(X'\right), q_{\alpha_{lo}}\left(X'\right) - T'\} \leq -\varepsilon - H\left(X'\right)\mathbb{I}\left(H\left(X'\right) \leq \varepsilon\right)|X'\right) \\
\geq&\mathbb{P}\left(\max\{T' - q_{\alpha_{hi}}\left(X'\right), q_{\alpha_{lo}}\left(X'\right) - T'\} \leq 0|X'\right) - b_2\left(\varepsilon + H\left(X'\right)\mathbb{I}\left(H\left(X'\right) \leq \varepsilon\right)\right) \\
\geq&\mathbb{P}\left(\max\{T' - q_{\alpha_{hi}}\left(X'\right), q_{\alpha_{lo}}\left(X'\right) - T'\} \leq 0|X'\right) - b_2\left(\varepsilon + H\left(X'\right)\right) \\
\geq&1 - \alpha - b_2\left(\varepsilon + H\left(X'\right)\right).
\end{aligned} \tag{17}$$

Combining Eqn 16 with Eqn 17 and taking expectations over X', we have:

$$\begin{aligned}
&\mathbb{P}\left(T' \in \hat{C}_n^2\left(X'\right)\right) \\
\geq& 1-\alpha-b_2\left(\varepsilon+\mathbb{E}H\left(X'\right)\right)-\mathbb{P}\left(H\left(X'\right)>\varepsilon\right)-\mathbb{P}\left(\eta<-\varepsilon\right) \\
\overset{(i)}{\geq}& 1-\alpha-b_2\left(2M_2+\gamma\right)-\mathbb{P}\left(\eta<-\left(M_2+\gamma\right)\right).
\end{aligned} \tag{18}$$

The Equation (i) holds by taking $\varepsilon = M_2+\gamma$. Due to Assumption (B1), we have $\mathbb{E}H\left(X'\right) \leq M_2$ and $\mathbb{P}\left(H\left(X'\right)>\varepsilon\right)=0$. Note that $\varepsilon = M_2+\gamma$ does not break the condition of Equation 17.

We next show that $\lim_{n\to\infty}\mathbb{P}\left(\eta<-\left(M_2+\gamma\right)\right) \leq \frac{16M_2}{(M_2+\gamma)^2 b_1}$. Firstly, by Lemma 7, we have

$$\lim_{n\to\infty}\mathbb{P}\left(\eta<-\left(M_2+\gamma\right)\right) \leq \frac{16}{\left(M_2+\gamma\right)^2 b_1}\mathbb{E}[\hat{w}\left(X\right)H\left(X\right)] \leq \frac{16M_2}{\left(M_2+\gamma\right)^2 b_1}\mathbb{E}[\hat{w}\left(X\right)] = \frac{16M_2}{\left(M_2+\gamma\right)^2 b_1}.$$

The inequality is due to Assumption (B1) and $\mathbb{E}[\hat{w}\left(X\right)]=1$.

The proof is done.

Technical Lemmas. In this part, we give some technical lemmas used in the proof. The following Lemma 6 shows that the difference between non-conformity score under true quantile q and non-conformity score under estimated quantile $\hat{q}$ is upper bounded by $H\left(X'\right)$.

Lemma 6 *Under notations in Theorem 5.1, we have* $|\max\{T'-\hat{q}_{\alpha_{hi}}\left(X'\right),\hat{q}_{\alpha_{lo}}\left(X'\right)-T'\} - \max\{T'-q_{\alpha_{hi}}\left(X'\right),q_{\alpha_{lo}}\left(X'\right)-T'\}| \leq H\left(X'\right)$

Proof. We investigate the results via situations on the operator $\max$.

If $\max\{T'-\hat{q}_{\alpha_{hi}}\left(X'\right),\hat{q}_{\alpha_{lo}}\left(X'\right)-T'\} = T'-\hat{q}_{\alpha_{hi}}\left(X'\right)$ and $\max\{T'-q_{\alpha_{hi}}\left(X'\right),q_{\alpha_{lo}}\left(X'\right)-T'\} = T'-q_{\alpha_{hi}}\left(X'\right)$, the conclusion follows the definition of $H\left(X'\right)$.

If $\max\{T'-\hat{q}_{\alpha_{hi}}\left(X'\right),\hat{q}_{\alpha_{lo}}\left(X'\right)-T'\} = T'-\hat{q}_{\alpha_{hi}}\left(X'\right)$ and $\max\{T'-q_{\alpha_{hi}}\left(X'\right),q_{\alpha_{lo}}\left(X'\right)-T'\} = q_{\alpha_{lo}}\left(X'\right)-T'$, we have

$$\begin{aligned}
&\max\{T'-\hat{q}_{\alpha_{hi}}\left(X'\right),\hat{q}_{\alpha_{lo}}\left(X'\right)-T'\} - \max\{T'-q_{\alpha_{hi}}\left(X'\right),q_{\alpha_{lo}}\left(X'\right)-T'\} \\
=&\left(T'-\hat{q}_{\alpha_{hi}}\left(X'\right)\right)-\left(q_{\alpha_{lo}}\left(X'\right)-T'\right) \\
\leq&\left(T'-\hat{q}_{\alpha_{hi}}\left(X'\right)\right)-\left(T'-q_{\alpha_{hi}}\left(X'\right)\right) \\
=&q_{\alpha_{hi}}\left(X'\right)-\hat{q}_{\alpha_{hi}}\left(X'\right) \\
\leq&H\left(X'\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\max\{T'-\hat{q}_{\alpha_{hi}}\left(X'\right),\hat{q}_{\alpha_{lo}}\left(X'\right)-T'\} - \max\{T'-q_{\alpha_{hi}}\left(X'\right),q_{\alpha_{lo}}\left(X'\right)-T'\} \\
=&\left(T'-\hat{q}_{\alpha_{hi}}\left(X'\right)\right)-\left(q_{\alpha_{lo}}\left(X'\right)-T'\right) \\
\geq&\left(\hat{q}_{\alpha_{lo}}\left(X'\right)-T'\right)-\left(q_{\alpha_{lo}}\left(X'\right)-T'\right) \\
=&\hat{q}_{\alpha_{lo}}\left(X'\right)-q_{\alpha_{lo}}\left(X'\right) \\
\geq&-H\left(X'\right).
\end{aligned}$$

Therefore, we have

$$|\max\{T'-\hat{q}_{\alpha_{hi}}\left(X'\right),\hat{q}_{\alpha_{lo}}\left(X'\right)-T'\} - \max\{T'-q_{\alpha_{hi}}\left(X'\right),q_{\alpha_{lo}}\left(X'\right)-T'\}| \leq H\left(X'\right).$$

The other two situations are derived similarly.

We next introduce Lemma 7 which provides the upper bounds of the term $\lim_{n\to\infty}\mathbb{P}\left(\eta<-\varepsilon\right)$.

Lemma 7 *Under the assumptions in Theorem 5.1, the following inequality holds.*

$$\lim_{n\to\infty}\mathbb{P}\left(\eta<-\varepsilon\right) \leq \frac{16}{\varepsilon^2 b_1}\mathbb{E}[\hat{w}\left(X\right)H\left(X\right)].$$

Proof. By combining Equation (A.10), (A.11), (A.12), (A.13) in Lei & Candès (2020), and apply the Assumption (B2) we have

$$\lim_{n\to\infty} \mathbb{P}\left(\eta < -\varepsilon\right) \leq \mathbb{P}\left(\sum_{i=1}^{n} \hat{w}\left(X_i\right) H\left(X_i\right) \geq \frac{\varepsilon^2 b_1 n}{16}\right).$$

And by Markov inequality, we have

$$\mathbb{P}\left(\sum_{i=1}^{n} \hat{w}\left(X_i\right) H\left(X_i\right) \geq \frac{\varepsilon^2 b_1 n}{16}\right) \leq \frac{16}{\varepsilon^2 b_1 n} \sum_{i=1}^{n} \mathbb{E}[\hat{w}\left(X_i\right) H\left(X_i\right)] = \frac{16}{\varepsilon^2 b_1} \mathbb{E}[\hat{w}\left(X\right) H\left(X\right)].$$

The proof is done.

A.3. Proof of Theorem 5.2

Theorem 5.1 (Lower Bound) *Let $\hat{w}(x)$ be an estimate of the weight $w(x)$, $\hat{q}_{\alpha_{lo}}(x), \hat{q}_{\alpha_{hi}}(x)$ be the quantile estimator returned by WCCI, and $H(X)$ be defined as Equation 7. Assume that $\mathbb{E}[\hat{w}(X)|\mathcal{Z}_{tr}] = 1$ and $\mathbb{E}[w(X)] = 1$, where all the expectation operators $\mathbb{E}$ are taken over $X \sim \mathcal{P}_{X|\Delta=1}$. Denote $\hat{C}_n^2(x)$ as the output band of Algorithm 2 with n calibration samples, and denote X' as the testing point.*

From the weight perspective, under assumptions (A1):

A1. $\mathbb{E}_{X|\Delta=1}|\hat{w}(X) - w(X)| \leq M_1$,
we have:

$$\lim_{n\to\infty} \mathbb{P}\left(T' \in \hat{C}_n^2(X')\right) \geq 1 - \alpha - \frac{1}{2}M_1.$$

From the quantile perspective, under assumptions (B1-B3):

B1. *$H(X) \leq M_2$ a.s. w.r.t. X;*
B2. *There exists $\delta > 0$ such that $\mathbb{E}\hat{w}(X)^{1+\delta} < \infty$;*
B3. *There exists $\gamma, b_1, b_2 > 0$ such that $\mathbb{P}(T = t|X = x) \in [b_1, b_2]$ uniformly over all (x,t) with $t \in [q_{\alpha_{lo}}(x) - 2M_2 - 2\gamma, q_{\alpha_{lo}}(x) + 2M_2 + 2\gamma] \cup [q_{\alpha_{hi}}(x) - 2M_2 - 2\gamma, q_{\alpha_{hi}}(x) + 2M_2 + 2\gamma]$,*
we have:

$$\lim_{n\to\infty} \mathbb{P}\left(T' \in \hat{C}_n^2(X')\right) \geq 1 - \alpha - b_2(2M_2 + \gamma) - \frac{16M_2}{(M_2 + \gamma)^2 b_1}.$$

We now show the proof of Theorem 5.2. For a new testing point $(X', T') \sim \mathcal{P}_X \times \mathcal{P}_{T|X}$, we have

$$\begin{aligned}
&\mathbb{P}\left(T' \in \hat{C}_n^2\left(X'\right)|X'\right) \\
=&\mathbb{P}\left(T' \in [\hat{q}_{\alpha_{lo}}\left(X'\right) - \eta, \hat{q}_{\alpha_{hi}}\left(X'\right) + \eta]|X'\right) \\
=&\mathbb{P}\left(\max\{T' - \hat{q}_{\alpha_{hi}}\left(X'\right), \hat{q}_{\alpha_{lo}}\left(X'\right) - T'\} \leq \eta|X\right) \\
\overset{i}{\leq}&\mathbb{P}\left(\max\{T' - q_{\alpha_{hi}}\left(X'\right), q_{\alpha_{lo}}\left(X'\right) - T'\} \leq \eta + H\left(X'\right)|X'\right) \\
\overset{ii}{\leq}&\mathbb{P}\left(\max\{T' - q_{\alpha_{hi}}\left(X'\right), q_{\alpha_{lo}}\left(X'\right) - T'\} \leq \eta_{q,w} + \varepsilon + H\left(X'\right)|X'\right) + \mathbb{P}\left(\eta - \eta_{q,w} > \varepsilon\right) \\
\overset{iii}{\leq}&\mathbb{P}(\max\{T' - q_{\alpha_{hi}}\left(X'\right), q_{\alpha_{lo}}\left(X'\right) - T'\} \leq \eta_{q,w} + \varepsilon \\
&+ H\left(X'\right)\mathbb{I}\left(H\left(X'\right) \leq \varepsilon\right)|X') + \mathbb{I}\left(H\left(X'\right) > \varepsilon\right) + \mathbb{P}\left(\eta - \eta_{q,w} > \varepsilon\right).
\end{aligned} \tag{19}$$

where we denote $\eta_{q,w} = \text{Quantile}\left(1-\alpha; \sum_{i=1}^{n} p_i \delta_{V_i^*} + p_\infty \delta_{V_\infty^*}\right)$, and obviously, $\eta_{q,w} = 0$ by the definition of the non-conformity score, where V_i^* is calculated based on $q_{\alpha_{lo}}(\cdot)$ and $q_{\alpha_{hi}}(\cdot)$.

Equation (i) follows from Lemma 6, and Equation (ii) follows from:

$$\begin{aligned}
&\mathbb{P}\left(\max\{T' - q_{\alpha_{hi}}\left(X'\right), q_{\alpha_{lo}}\left(X'\right) - T'\} \leq \eta + H\left(X'\right)|X'\right) \\
&- \mathbb{P}\left(\max\{T' - \hat{q}_{\alpha_{hi}}\left(X'\right), \hat{q}_{\alpha_{lo}}\left(X'\right) - T'\} \leq \eta_{q,w} + \epsilon + H\left(X'\right)|X'\right) \\
\leq&\mathbb{P}\left(\eta_{q,w} + \epsilon + H\left(X'\right) < \max\{T' - q_{\alpha_{hi}}\left(X'\right), q_{\alpha_{lo}}\left(X'\right) - T'\} \leq \eta + H\left(X'\right)|X'\right) \\
\leq&\mathbb{P}\left(\eta_{q,w} + \epsilon + H\left(X'\right) < \eta + H\left(X'\right)|X'\right) \\
=&\mathbb{P}\left(\eta - \eta_{q,w} > \varepsilon\right).
\end{aligned}$$

Equation (iii) can be derived simply based on the discussion on the value of $\mathbb{I}\left(H\left(X'\right)\leq\varepsilon\right)$.

By Assumption (C4), since $\varepsilon+H\left(X'\right)\mathbb{I}\left(H\left(X'\right)\leq\varepsilon\right)\leq 2\varepsilon$, when $\varepsilon\leq M_2+\gamma$, we have

$$\begin{aligned}
&\lim_{n\to\infty}\mathbb{P}\left(\max\{T'-q_{\alpha_{hi}}\left(X'\right),q_{\alpha_{lo}}\left(X'\right)-T'\}\leq\eta_{q,w}+\varepsilon+H\left(X'\right)\mathbb{I}\left(H\left(X'\right)\leq\varepsilon\right)|X'\right)\\
\leq&\lim_{n\to\infty}\mathbb{P}\left(\max\{T'-q_{\alpha_{hi}}\left(X'\right),q_{\alpha_{lo}}\left(X'\right)-T'\}\leq\eta_{q,w}|X'\right)+b_2\left(\varepsilon+H\left(X'\right)\mathbb{I}\left(H\left(X'\right)\leq\varepsilon\right)\right)\\
\leq&\lim_{n\to\infty}\mathbb{P}\left(\max\{T'-q_{\alpha_{hi}}\left(X'\right),q_{\alpha_{lo}}\left(X'\right)-T'\}\leq 0|X'\right)+b_2\left(\varepsilon+H\left(X'\right)\right)\\
\leq&1-\alpha+b_2\left(\varepsilon+H\left(X'\right)\right).
\end{aligned}\tag{20}$$

The last inequality is from Lemma 2. Combining Eqn 19 with Eqn 20 and taking expectations over X', we have:

$$\begin{aligned}
&\lim_{n\to\infty}\mathbb{P}\left(T'\in\hat{C}_n^2\left(X'\right)\right)\\
\leq&1-\alpha+b_2\left(\varepsilon+\mathbb{E}H\left(X'\right)\right)+\mathbb{P}\left(H\left(X'\right)>\varepsilon\right)+\lim_{n\to\infty}\mathbb{P}\left(\eta-\eta_{q,w}>\varepsilon\right).
\end{aligned}$$

By taking $\varepsilon=M_2'+M_1'/K$, and plugging in Assumption (B1), the above equation is equivalent to

$$\lim_{n\to\infty}\mathbb{P}\left(T'\in\hat{C}_n^2\left(X'\right)\right)\leq 1-\alpha+b_2\left(2M_2'+M_1'/K\right)+\lim_{n\to\infty}\mathbb{P}\left(\eta-\eta_{q,w}>M_2'+M_1'/K\right).\tag{21}$$

The left is to show that

$$\lim_{n\to\infty}\mathbb{P}\left(\eta-\eta_{q,w}>M_2'+M_1'/K\right)=0.\tag{22}$$

First we denote $\eta_q=\text{Quantile}\left(1-\alpha;\sum_{i=1}^n\hat{p}_i\delta_{V_i^*}+\hat{p}_\infty\delta_{V_\infty^*}\right)$ where $\hat{p}_i$ is the estimator of p_i. Equation 22 holds by applying Lemma 8 and Lemma 9.

The proof is done.

Technical Lemmas. In this part, we give some technical lemmas used in the proof. We first prove the following Lemma 8 which gives an upper bound of $|\eta-\eta_q|$.

Lemma 8 *Under assumptions of Theorem 5.2,* $|\eta-\eta_q|\leq M_2'$

Proof. Notice that η is the quantile of distribution $\sum_{i=1}^n\hat{p}_i\delta_{V_i}+\hat{p}_\infty\delta_{V_\infty}$ and η_q is the quantile of distribution $\sum_{i=1}^n\hat{p}_i\delta_{V_i^*}+\hat{p}_\infty\delta_{V_\infty^*}$, where V_i is calculated based on $\hat{q}_{\alpha_{hi}},\hat{q}_{\alpha_{lo}}$.

When $\eta=\eta_q$, the conclusion directly follows. When $\eta\neq\eta_q$, notice that the two distribution share the same weight $\hat{w}$, there must exist a V' and V'' such that

$$\begin{aligned}
V'^*\leq\eta,\,V'\geq\eta_q\\
V''^*\geq\eta,\,V''\leq\eta_q.
\end{aligned}$$

This leads to the following two inequalities by Assumption (C2) and Lemma 7.

$$\begin{aligned}
\eta-\eta_q\geq V'^*-V'\geq -H\left(X\right)\geq -M_2'\\
\eta-\eta_q\leq V''^*-V''\leq H\left(X\right)\leq M_2'
\end{aligned}\tag{23}$$

Therefore,

$$|\eta-\eta_q|\leq M_2'.$$

The proof is done.

We next prove Lemma 9 which provides an upper bound of $|\eta_q-\eta_{q,w}|$.

Lemma 9 *Under assumptions of Theorem 5.2,* $|\eta_q-\eta_{q,w}|\leq M_1'/K$

Proof. WLOG, assume $\eta_{q,w} - \eta_q = \Delta > 0$. Denote $F(\cdot) = \sum_{i=1}^n p_i \delta_{V_i^*} + p_\infty \delta_{V_\infty^*}$ and $G(\cdot) = \sum_{i=1}^n \hat{p}_i \delta_{V_i^*} + \hat{p}_\infty \delta_{V_\infty^*}$. By the definition of $\eta_q, \eta_{q,w}$, we have

$$F(\eta_{q,w}) = G(\eta_q) = 1 - \alpha.$$

Therefore, on the one hand, by Assumption (C1) and Assumption $\mathbb{E}w(X) = \mathbb{E}\hat{w}(X) = 1$

$$\begin{aligned} & F(\eta_{q,w}) - F(\eta_q) \\ =& G(\eta_q) - F(\eta_q) \\ \leq& \sup_t G(\eta_q) - F(\eta_q) \\ \leq& \sup_S \left| \sum_{i \in S} w(X_i) - \hat{w}(x_I) \right| \\ \leq& M_1'. \end{aligned} \tag{24}$$

On the other hand, by Assumption (C3)

$$\begin{aligned} & F(\eta_{q,w}) - F(\eta_q) \\ =& F(\eta_q + \Delta) - F(\eta_q) \\ \geq& K\Delta. \end{aligned} \tag{25}$$

Combining Equation 24 and Equation 25 leads to

$$\eta_{q,w} - \eta_q \leq M_1'/K.$$

The conclusion directly follows.

Combining Lemma 8 and Lemma 9 leads to the upper bound of $|\eta - \eta_{q,w}|$.

B. Supplemental Experiment Results

All codes are available at https://github.com/thutzr/Cox.

B.1. Experiment Process

Data Pre-processing. There are both numerical features and categorical features in our datasets. We normalize numerical features on both training set and test set.

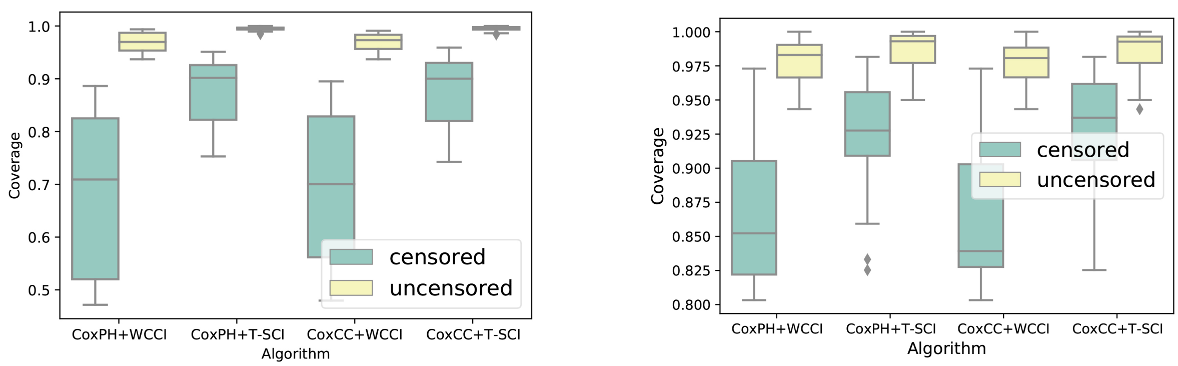
Process. In each run, the experiment runs as follows:

- We first randomly split 80% data as the training set while the rest is splitted randomly into calibration set and test set.
- Following (Chen, 2020), we randomly sample 100 data points, which are used to construct prediction intervals, from test set. Denote these points as $\mathcal{X}_{\text{centers}}$.
- For each point $x_0 \in \mathcal{X}_{\text{centers}}$:
  - We sample 100 data points in test set with respect to sampling probability proportional to $K(x, x_0)$, where $K(\cdot)$ is the Gaussian kernel.
  - For these 100 test points, we use algorithm 1 and 2 to calculate the predicted survival interval with respect to the given confidence level α and check if the true survival time of each point is included in the predicted interval. The fraction of points that are covered in the calculated confidence interval is the empirical coverage. And the difference of upper confidence band and its lower counterpart is interval length. In our experiments, the upper interval band is likely to be infinity sometimes. We truncate those upper bands to the maximum duration of the according dataset.

We run the above procedure for different confidence level αs. Results show that our algorithms are robust and effective for different αs. In our experiments, α is chosen to be $0.6, 0.7, 0.8, 0.9, 0.95$,respectively.

Table 3. Model Comparison on METABRIC

Method	Total		Censored		Uncensored		Interval Length	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
Cox Reg.	0.996	0.003	/	/	/	/	319.84	2.78
Random Survival Forest (Ishwaran et al., 2008)	0.997	0.002	/	/	/	/	341.73	4.82
Nnet-Survival (Gensheimer & Narasimhan, 2019)	0.978	0.008	0.554	0.017	0.980	0.003	19.85	0.34
MTLR (Yu et al., 2011)	0.990	0.005	0.990	0.008	0.990	0.005	306.08	6.44
CoxPH (Katzman et al., 2018)	0.995	0.003	0.994	0.005	0.995	0.005	340.20	6.85
CoxCC (Katzman et al., 2018)	0.996	0.004	0.994	0.005	0.998	0.004	344.26	6.93
CoxPH+WCCI	0.977	0.011	0.948	0.021	0.985	0.009	334.14	5.33
CoxPH+T-SCI	0.986	0.009	0.970	0.016	0.989	0.008	340.67	2.91
CoxCC+WCCI	0.973	0.014	0.942	0.027	0.980	0.012	334.28	5.36
CoxCC+T-SCI	0.986	0.008	0.971	0.015	0.990	0.007	340.80	2.98
CoxPH+WCCI(unweigted)	0.946	0.031	0.910	0.055	0.972	0.021	254.56	8.38
CoxPH+T-SCI(unweighted)	0.958	0.063	0.932	0.063	0.977	0.021	261.18	9.26
CoxCC+WCCI(unweigted)	0.946	0.031	0.904	0.053	0.968	0.019	254.35	8.27
CoxCC+T-SCI(unweighted)	0.958	0.063	0.926	0.061	0.975	0.022	261.17	8.87
Kernel (Chen, 2020)	0.981	0.025	0.997	0.010	0.971	0.038	337.66	20.39



(a) SUPPORT

(b) METABRIC

Figure 7. Empirical Coverage of Censored and Uncensored Data

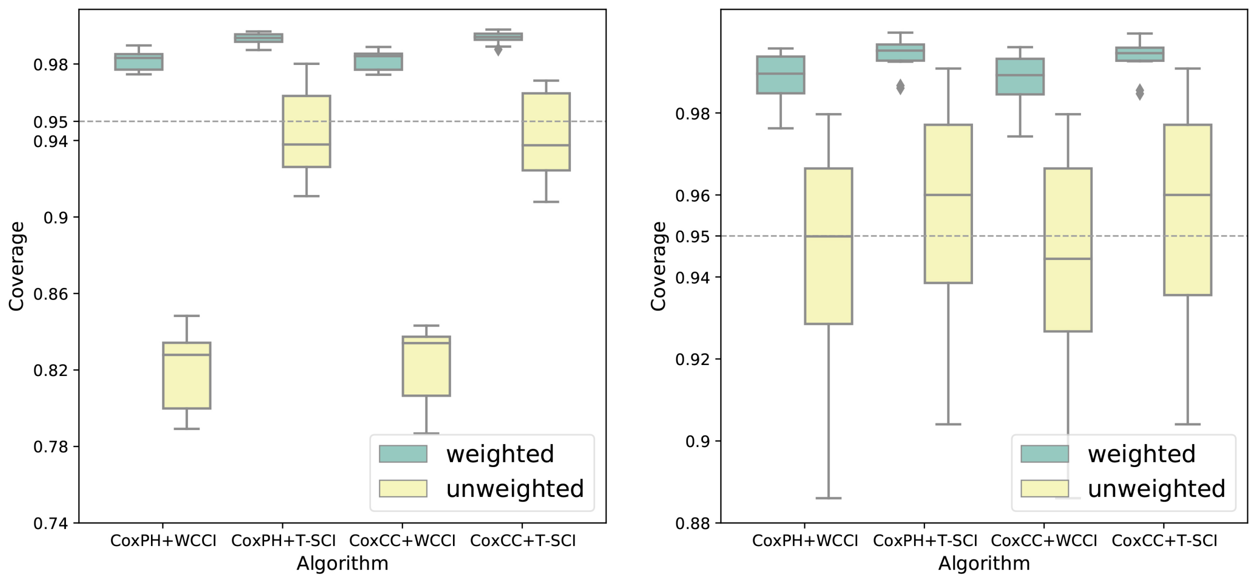
B.2. Model and Hyperparameters

We use pycox and PyTorch to implement CoxCC,CoxPH and neural network model respectively. Then we combine them together to Cox-MLP models. We implement a neural network model with three hidden layers, where each layer has 32 hidden nodes. Between each two layers, we use ReLU as the activation function. We apply batch normalization and dropout which drop 10% nodes at one epoch. Adam is chosen to be the optimizer. In the training process, we feed 80% data as training data and the rest as validation data. Note that here the total data set (training data + validation data) is the training data mentioned in the main article. Each batch contains 128 data points. We train the network for 512 epochs and the trained model is used as the MLP part in our Cox-MLP model.

B.3. Supplemental Results

All the empirical coverage and predicted interval length of different models are listed in Table 1 (Dataset: RRNLNPH) Table 2 (Dataset: SUPPORT) and Table 3 (Dataset: METABRIC). The results in Table 2 and Table 3 are similar to those of RRNLNPH.

Prediction on censored data and uncensored data are compared in Figure 7. Performance on censored data are much better than that of uncensored data. The standard deviation is also larger on uncensored data than censored data. This is consistent with our intuition as we do not have exact information of censored data.



(a) SUPPORT

(b) METABRIC

Figure 8. Empirical Coverage of Weighted and Unweighted Models

Comparison on weighted and unweighted models are shown in Figure 8. Performance of unweighted models are much weaker than weighted models. This shows the effectiveness of weighted conformal inference.

We also compare performance on different α's on both SUPPORT and METABRIC. Results similar with RRNLNPH are shown in Figure 9, Figure 10, and Figure 11. Note that we split the weighted and unweighted version to make the figure more clear. The empirical coverage of our algorithms exceed the given confidence level and the standard deviation of prediction is relatively low.

Notice that in Table 2 and Table 3, censored coverage in CoxPH and CoxCC performs well (around 0.992 in Table 2 and 0.004 in Table 3). However, we emphasize that any coverage type could happen due to a lack of theoretical guarantee (extremely large, extremely small, or highly unbalanced, etc.). The censored coverage of CoxPH happens to be large (0.992) in Table 2 (dataset: SUPPORT) and Table 3 (dataset: METABRIC), just like it happens to be small in Table 1 (0.554, dataset: RRNLNPH). As a comparison, our newly proposed method (T-SCI) is guaranteed to return a nearly perfect guarantee (the whole coverages are larger than 0.95), which is more stable.

C. Supplementary Notes

We make some supplementary notes in this section.

C.1. Stability of Non-conformity Score

We state in Section 4 that the non-conformity score is usually more stable when it is single-peak. A multi-peak situation implies that samples with different covariate may have different coverage, namely,

$$\mathbb{P}(T \in C_n(X)|X = x_1) \neq \mathbb{P}(T \in C_n(X)|X = x_2),$$

where x_1, x_2 belong to different peaks. Therefore, we describe the above formula as "unstable" since it provides distinct coverage for distinct groups, although the overall coverage (for the population) is still $1 - \alpha$.

The multi-peak phenomenon comes from the fact that we calculate the $1 - \alpha$ quantile of V_i based on all samples. For example, consider a two-peak distribution where the first group has $1 - \alpha$ populations. Then the algorithm returns zero coverage (probability equal to zero) for the second group and returns one coverage (probability equal to one) for the first group, which causes instability.

C.2. The Effect of Censoring

Figure 1 illustrates that ignoring and deleting censoring will indeed cause bias and inefficiency. We may also consider a more straightforward case to calculate the sample mean of a dataset. However, the dataset contains censoring issues, where we clip the data to a constant C when the value is larger than C. If we ignore the censoring phenomenon, the new sample mean is smaller than the expected value, leading to *bias*. If we delete the censoring phenomenon, the new sample mean is also smaller since we delete samples with large values (those deleted samples are always larger than the constant C). Besides, it leads to *inefficiency* since we use fewer samples during the inference.

C.3. Strong Ignorability Assumption

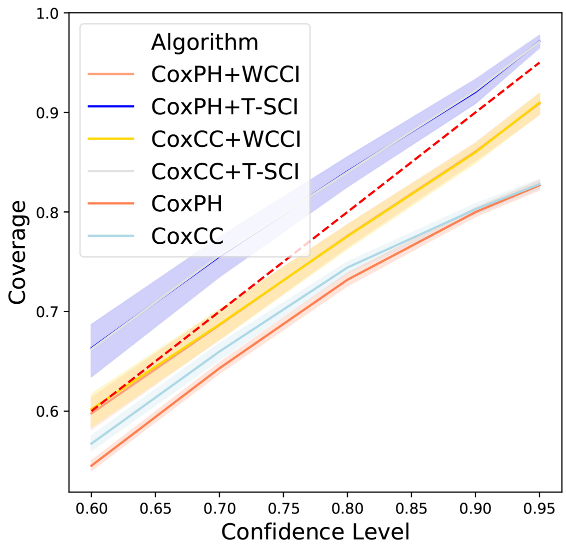
Strong ignorability assumption stands for

$$T \perp \Delta \mid X.$$

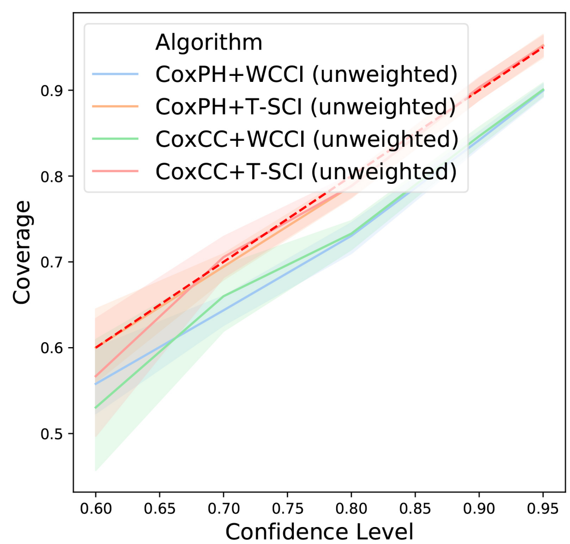
Note that strong ignorability assumption directly leads to the fact:

$$\mathcal{P}_{T|X,\Delta} = \mathcal{P}_{T|X}.$$

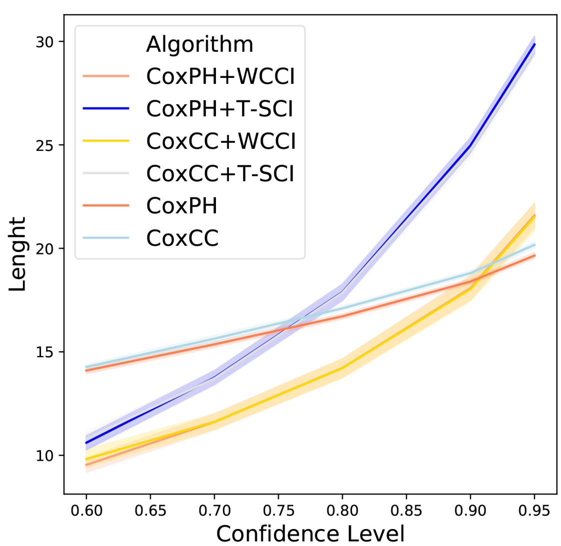
since $\mathcal{P}_{T|X,\Delta} = \mathcal{P}_{T,\Delta|X}/\mathcal{P}_{\Delta|X} = \mathcal{P}_{T|X}$. Therefore, we can apply weighted conformal inference which requires a covariate shift. A similar idea could be found in Lei & Candès (2020).



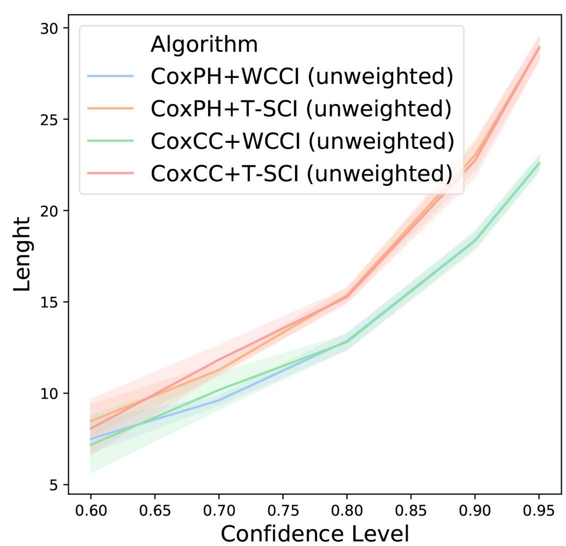
(a) Coverage Comparison(weighted)



(b) Coverage Comparison(unweighted)

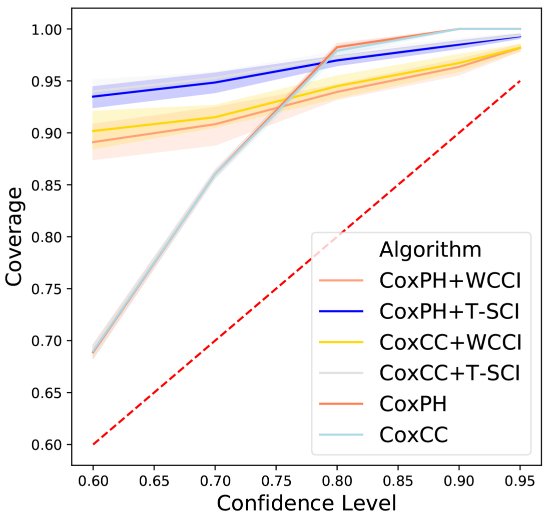


(c) Interval Length Comparison(weighted)

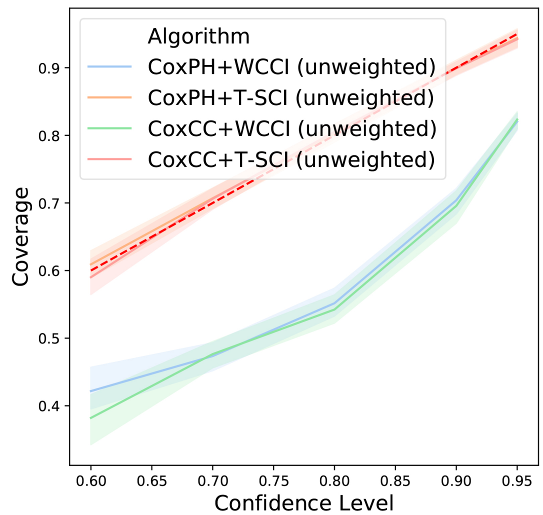


(d) Interval Length Comparison(unweighted)

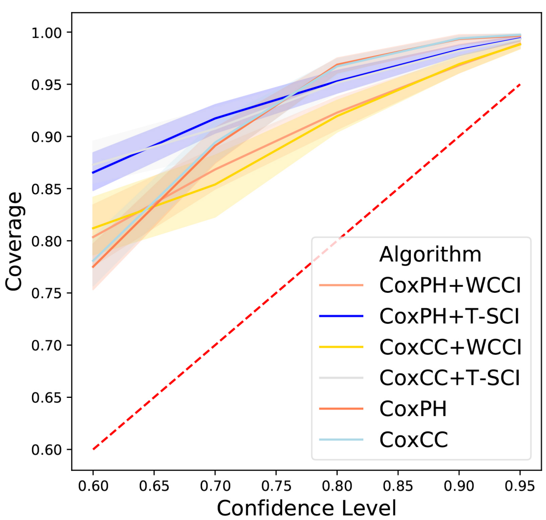
Figure 9. Different Model's Empirical Coverage of Different α (RRNLNPH).



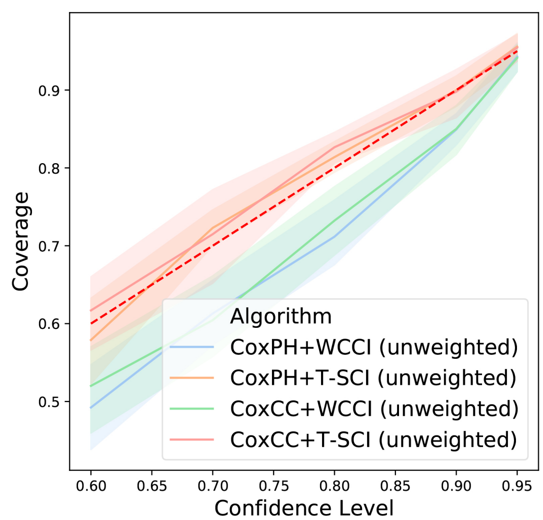
(a) Coverage Comparison(weighted)



(b) Coverage Comparison(unweighted)

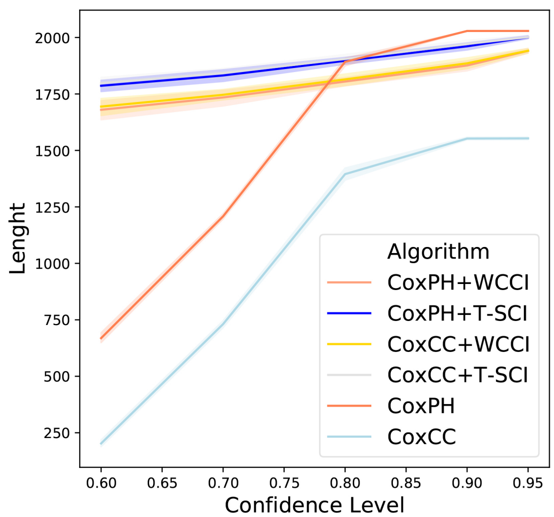


(c) Interval Length Comparison(weighted)

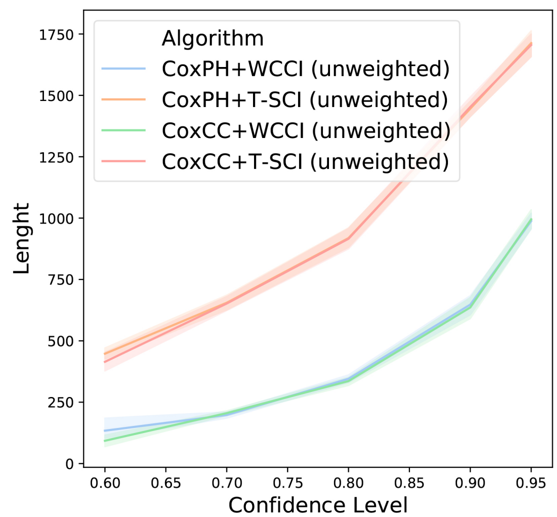


(d) Interval Length Comparison(unweighted)

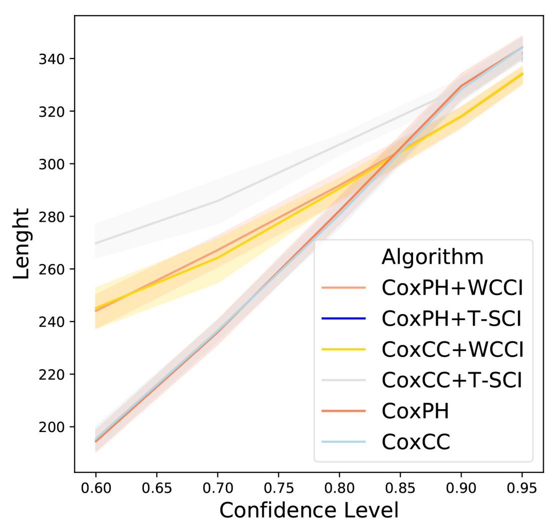
Figure 10. Different Model's Empirical Coverage of Different α (SUPPORT)



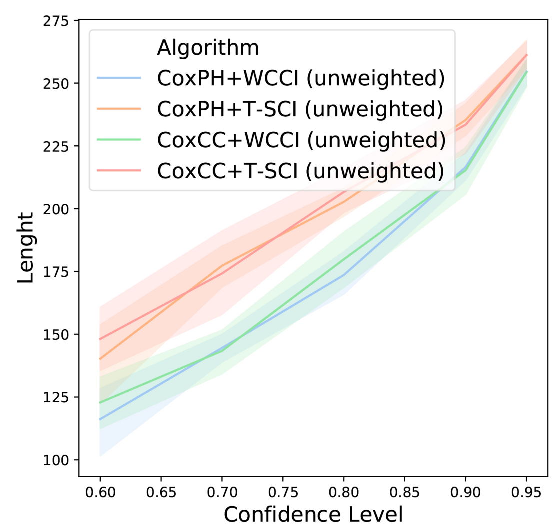
(a) Coverage Comparison(weighted)



(b) Coverage Comparison(unweighted)



(c) Interval Length Comparison(weighted)



(d) Interval Length Comparison(unweighted)

Figure 11. Different Model's Empirical Coverage of Different α (METABRIC)