

History-Aware Conformal Prediction Sets for Censored Time-to-Event Outcomes

Yuyao Wang

Department of Biostatistics, Brown University
Department of Health Care Policy, Harvard Medical School

Joint work with:

Alexander W. Levis, University of Pennsylvania

Shu Yang, North Carolina State University

Larry Han, Brown University

Dynamic survival prediction

- Update the prediction of individual time-to-events as new information becomes available
- Classical methods: landmarking, joint modeling
- Recent developments in machine learning methods
 - ▶ Neural networks, e.g., Dynamic-DeepHit, ...
 - ▶ Random survival forests, SVM, ...
- Limitation: often lack reliable uncertainty quantification
→ limit their application in high stakes settings.

Conformal prediction for time-to-event outcomes

- Early developments focus on lower prediction bounds (LPB) under type-I censoring (Candes et al., 2023; Gui et al, 2024)
- Under the general right censoring setting:
 - censoring imputation (Sesia and Svetnik, 2025); **weighting** (Davidov et al., 2025)
 - EIF-based methods(Farina et al., 2025, Si & Qiu, 2025)
- Recent work on mixed-type prediction sets based on censoring status prediction (Holmes & Marandon, 2024), and two-sided prediction intervals (Yi et al., 2025; Qin et al., 2025).
- Existing conformal survival methods are all **static**
- Limitation: **not informative**

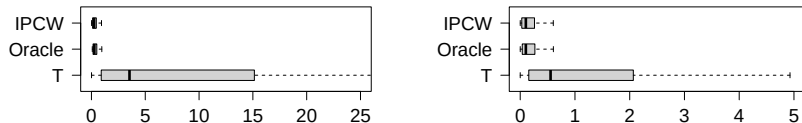


Figure: Boxplots of 90% LPBs under (left) simulation setup 1 in Gui et al. (2024), and (right) simulation setup 1 in Farina et al. (2025).

Notation and prediction target

- T : event time; C : censoring time.
- Z_t : covariates measured at t ; $\bar{Z}_t = \{Z_s : 0 \leq s \leq t\}$.
- Observed data: an i.i.d. sample $\mathcal{D} = \{O_i\}_{i=1}^n$, where $O = (X, \Delta, \bar{Z}_X)$, $X = \min(T, C)$, $\Delta = \mathbb{1}(T < C)$.

Prediction target

- For a fixed prediction time $\tau \geq 0$, construct a prediction set $\hat{\mathcal{C}}(\bar{Z}_{\tau, n+1})$, satisfying *survivor conditional coverage*:

$$\mathbb{P}\left\{T_{n+1} \in \hat{\mathcal{C}}(\bar{Z}_{\tau, n+1}) \mid T_{n+1} > \tau\right\} \geq 1 - \alpha.$$

- *Probably Asymptotically Approximately Correct* (PAAC) coverage:

With prob. $\geq 1 - \epsilon$ over $\mathcal{D}$,

$$\mathbb{P}\left\{T_{n+1} \in \hat{\mathcal{C}}(\bar{Z}_{\tau, n+1}) \mid T_{n+1} > \tau, \mathcal{D}\right\} \geq 1 - \alpha + o_p(1)$$

Assumptions

- **Conditional independent censoring given covariate history:**

$$\lambda_C(t|\bar{Z}_T, T) = \lambda_C(t|\bar{Z}_t)$$

- **Positivity:**

(i) $\mathbb{P}(T > \tau) > 0$;

(ii) there exists $\eta > 0$ s.t. $\mathbb{P}(C > T \mid T, \bar{Z}_T) > \eta$ a.s..

$$\lambda_C(t|\cdot) := \lim_{h \rightarrow 0^+} \mathbb{P}(t \leq C < t+h \mid \cdot, C \geq t, T \geq t)/h$$

Calibration condition

- **Nested candidate prediction sets:** e.g., $\mathcal{C}_{\tau,\theta}(\bar{z}_\tau; \mathcal{A}) = (\hat{q}_{1-\theta}(\bar{z}_\tau), \hat{q}_\theta(\bar{z}_\tau))$, $\theta \in [0.5, 1]$.
- **Calibration condition in censoring-free data:**

$$\mathbb{E} \left[\mathbb{1}(T > \tau) \left\{ \mathbb{1}(T \in \mathcal{C}_{\tau,\theta}(\bar{Z}_\tau; \mathcal{A})) - (1 - \alpha) \right\} \right] \geq 0.$$

- **With censored data, by IPCW** (Robins & Rotnitzky, 1992; Rotnitzky & Robins, 2005):

$$U(\theta; G, \mathcal{A}) = \frac{\Delta \mathbb{1}(X > \tau)}{G(X|\bar{Z}_X)} \left[\mathbb{1}\{X \in \mathcal{C}_{\tau,\theta}(\bar{Z}_\tau; \mathcal{A})\} - (1 - \alpha) \right].$$

- **Calibration condition in the observed data (censored):**

$$\mathbb{E} \{ U(\theta; G, \mathcal{A}) \} \geq 0.$$

$$G(t|\bar{Z}_t) = \exp\{-\int_0^t \lambda_C(u|\bar{Z}_u) du\} = \mathbb{P}(C > t | \bar{Z}_t); \quad \lambda_C(t|\cdot) = \lim_{h \rightarrow 0^+} \mathbb{P}(t \leq C < t+h | \cdot, C \geq t, T \geq t)/h$$

History-aware conformal prediction sets (HAPS)

Split-conformal algorithm:

- 1 Randomly split $\mathcal{D}$ (50:50) into a training set $\mathcal{D}_{\text{tr}}$ and a calibration set $\mathcal{D}_{\text{cal}}$.
- 2 Fit the prediction model $\hat{\mathcal{A}}$ on $\mathcal{D}_{\text{tr}}$, and construct nested candidate sets $\{\mathcal{C}_{\tau, \theta}\}_{\theta \in \Theta}$.
- 3 On $\mathcal{D}_{\text{tr}}$, estimate the nuisance parameter G , and denote the estimate as $\hat{G}$.
- 4 On $\mathcal{D}_{\text{cal}}$, compute $\hat{\theta} = \inf \left\{ \theta \in \Theta : \sum_{i \in \mathcal{D}_{\text{cal}}} U_i(\theta; \hat{G}, \hat{\mathcal{A}}) \geq 0 \right\}$.
- 5 **Output:** prediction set $\hat{\mathcal{C}}(\bar{Z}_{\tau, n+1}) = \mathcal{C}_{\tau, \hat{\theta}}(\bar{Z}_{\tau, n+1}; \hat{\mathcal{A}})$.

HAPS coverage

Theorem 1 (HAPS coverage)

For any $\epsilon \in (0, 1)$, there exists $K > 0$ such that, *with prob.* $\geq 1 - \epsilon$ over $\mathcal{D}$,

$$\begin{aligned} \mathbb{P}(T_{n+1} \in \hat{\mathcal{C}}(\bar{Z}_{\tau, n+1}) \mid T_{n+1} > \tau, \mathcal{D}) \\ \geq 1 - \alpha - s_{\tau}^{-1} \eta^{-2} \|\hat{G} - G\|_2 - s_{\tau}^{-1} \eta^{-1} \left(\sqrt{\frac{1}{2} \log \frac{1}{\epsilon}} + K \right) \frac{1}{\sqrt{|\mathcal{D}_{cal}|}}. \end{aligned}$$

- Achieves PAAC type coverage if $\|\hat{G} - G\|_2 \rightarrow \infty$
- Sensitive to censoring estimation error

Doubly robust extensions

HAPS-DR: doubly robust post-processing

$$\hat{C}_{\text{DR}}(\bar{Z}_\tau) = \hat{C}(\bar{Z}_\tau) \cup (\hat{q}_{\alpha/2}(\bar{Z}_\tau), \hat{q}_{1-\alpha/2}(\bar{Z}_\tau)).$$

- Achieves PAAC coverage if either $\hat{G}$ or $(\hat{q}_{\alpha/2}, \hat{q}_{1-\alpha/2})$ is consistent.

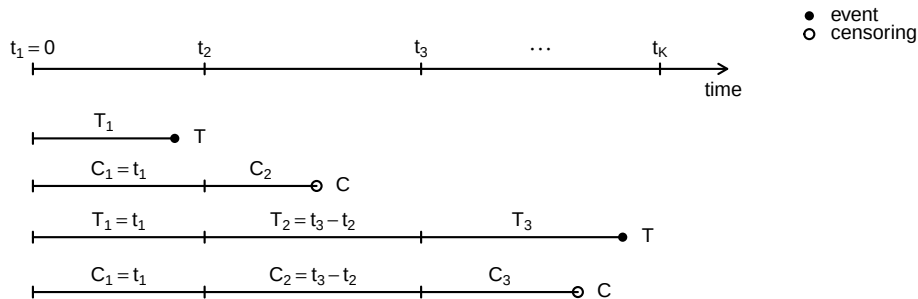
HAPS-A: use augmented IPCW (AIPCW) estimating function,

$$U_{\text{AIPCW}} = U(\theta; G, \mathcal{A}) + \int_0^X h_\tau(u, \bar{Z}_u; \theta, \mathcal{A}) \frac{dM_C(u; G)}{G(u | \bar{Z}_u)}.$$

where $h_\tau(u, \bar{Z}_u; \theta, \mathcal{A}) = \mathbb{E}(\mathbb{1}(T > \tau) [\mathbb{1}\{T \in \mathcal{C}_{\tau, \theta}(\bar{Z}_\tau; \mathcal{A})\} - (1 - \alpha)] | \bar{Z}_u, T \geq u)$,
 $N_C(t) = \mathbb{1}(X \leq t, \Delta = 0)$, $M_C(t; G) = N_C(t) - \int_0^t \mathbb{1}(X \geq u) dG(u | \bar{Z}_u) / G(u | \bar{Z}_u)$.

Nuisance parameter estimation

Estimation of h_τ in settings where time-varying covariates only change at fixed visit times.



Coverage theory for HAPS-A: product rate definitions

$$\begin{aligned}r_1(\theta) &= \int_0^\infty \left\| \hat{h}_\tau(u, \bar{Z}_u; \theta, \hat{\mathcal{A}}) - h_\tau(u, \bar{Z}_u; \theta, \mathcal{A}) \right\|_2 \cdot \left\| \hat{\lambda}_C(u | \bar{Z}_u) - \lambda_C(u | \bar{Z}_u) \right\|_2 du, \\r_2(\theta) &= \left\| \hat{h}_\tau(X, \bar{Z}_X; \theta, \hat{\mathcal{A}}) - h_\tau(X, \bar{Z}_X; \theta, \mathcal{A}) \right\|_2 \cdot \left\| \hat{\Lambda}_C(X | \bar{Z}_X) - \Lambda_C(X | \bar{Z}_X) \right\|_2 \\&\quad + \int_0^\infty \left\| \Lambda_C(u | \bar{Z}_u) - \hat{\Lambda}_C(u | \bar{Z}_u) \right\|_2 \cdot \left\| \frac{\partial}{\partial u} \hat{h}_\tau(u, \bar{Z}_u; \theta, \hat{\mathcal{A}}) - \frac{\partial}{\partial u} h_\tau(u, \bar{Z}_u; \theta, \mathcal{A}) \right\|_2 du, \\r_3(\theta) &= \left\| \hat{\Lambda}_C - \Lambda_C \right\|_{\text{sup},2} \cdot \left\| \hat{h}_\tau(\cdot, \cdot; \theta, \hat{\mathcal{A}}) - h_\tau(\cdot, \cdot; \theta, \mathcal{A}) \right\|_{\text{sup},2}.\end{aligned}$$

Recall that $\mathcal{D}$ is the observed sample. For a random function $\hat{f}(u, \bar{z}_u)$ estimated from $\mathcal{D}$, define $\|\hat{f}(u, \bar{Z}_u)\|_2 = \mathbb{E}\{|\hat{f}(u, \bar{Z}_u)|^2 \mid \mathcal{D}\}^{1/2}$, $\|\hat{f}\|_{\text{sup},2} = \mathbb{E}\{\sup_u |\hat{f}(u, \bar{Z}_u)|^2 \mid \mathcal{D}\}^{1/2}$, $\Lambda_C(t|\bar{Z}_t) = \int_0^t \lambda_C(t|\bar{Z}_u) du$.

Coverage theory for HAPS-A

Theorem 2 (HAPS-A coverage)

For any $\epsilon \in (0, 1)$, there exists a constant $K > 0$ such that with probability $\geq 1 - \epsilon$ over $\mathcal{D}$,

$$\mathbb{P}(T \in \hat{\mathcal{C}}_A(\bar{Z}_T) \mid T > \tau, \mathcal{D}) > 1 - \alpha - s_\tau^{-1} \eta^{-1} \cdot \sup_{\theta \in \Theta} \left[\min \left\{ r_1(\theta), r_2(\theta) + (1 - \eta) r_3(\theta) \right\} \right] \\ - s_\tau^{-1} c_\eta \left(\sqrt{\frac{1}{2} \log \frac{1}{\epsilon}} + K \right) \frac{1}{\sqrt{|\mathcal{D}_{cal}|}},$$

where $c_\eta = 2 \max(\alpha, 1 - \alpha)(2\eta^{-1} - 1)$.

- Achieves PAAC coverage if either $\hat{G}$ or $\hat{h}_\tau$ is consistent.
- Rate double robustness: nuisance estimation errors affect the coverage gap only by the product errors between $\hat{G}$ and $\hat{h}_\tau$.

Simulation

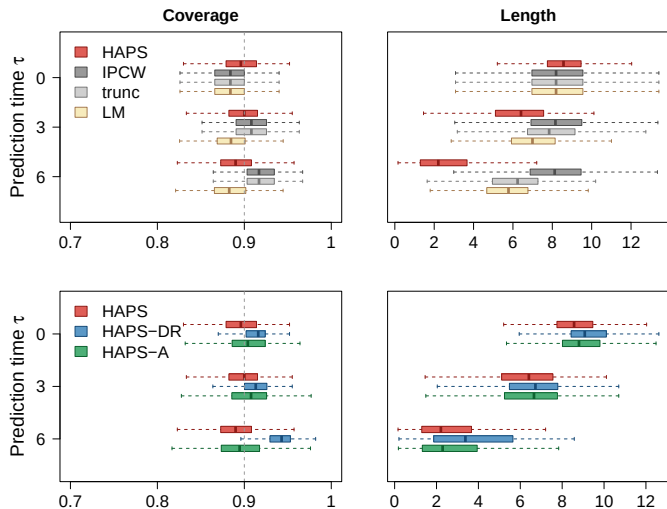


Figure: 90% prediction intervals under 200 Monte Carlo replications, with Dynamic-DeepHit prediction model and RSF/XGB-XGB nuisance models. Sample size $n = 1000$; test data with size 500 and no censoring. Relative to baseline methods, the median interval length is reduced by **5%–22%** at $\tau = 3$ and **34%–68%** at $\tau = 6$.

Simulation - disentangle the effects of time-varying covariates

Table: Coverage (with standard errors) and median interval length at $\tau = 3$ and $\tau = 6$ for 90% prediction intervals.

Method	$\tau = 3$		$\tau = 6$	
	Cov (SE)	Len	Cov (SE)	Len
HAPS	.899 (.0020)	7.31	.899 (.0017)	2.35
IPCW-Ztau	.917 (.0024)	7.81	.903 (.0018)	2.37
HAPS-Z0	.895 (.0022)	8.32	.900 (.0019)	7.04

Simulation – advantage of rate DR

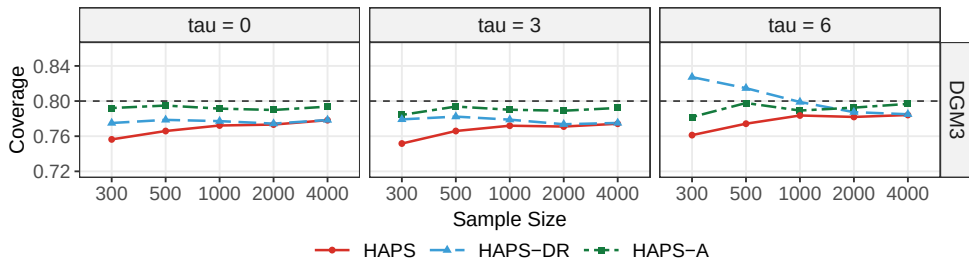


Figure: Mean coverage and median interval length of 80% prediction intervals, with RSF prediction model and RSF/XGB-XGB nuisance models, under 200 Monte Carlo replications.

Applications: colon cancer data

- 929 patients; censoring rate: 51.3%.
- Baseline covariates: age (years), sex (male/female), treatment group (observation, Levamisole, or Levamisole+5-FU), number of positive lymph nodes (more than 4 vs. not)
- Time-varying covariate: recurrence (a 0–1 process).

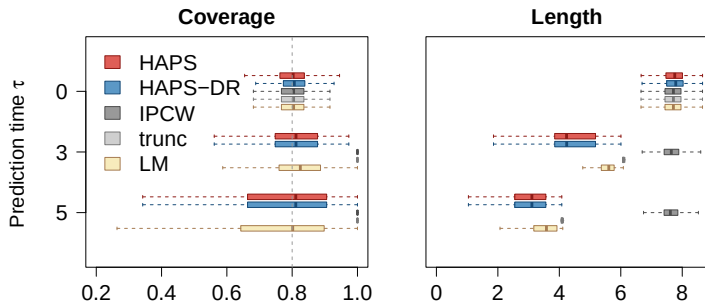


Figure: 80% prediction intervals with Dynamic-DeepHit prediction model for the colon cancer data across 200 random splits. Relative to baseline methods, the median interval length is reduced by **22%–44%** at $\tau = 3$ years and **13%–60%** at $\tau = 5$ years.

Applications: PBC data

- 312 patients; censoring rate: 55.1%.
- Baseline covariates: age
- Time-varying covariates: $\log(\text{bilirubin})$, $\log(\text{prothrombin time})$, $\log(\text{albumin})$, and edema score (0/0.5/1).

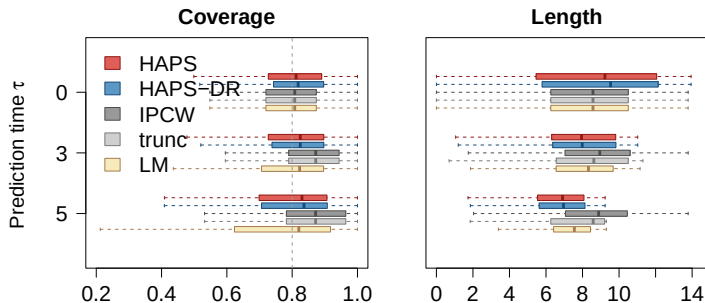


Figure: 80% prediction intervals with Dynamic-DeepHit prediction model for the PBC data across 200 random splits. Relative to baseline methods, the median interval length is reduced by **9%–22%**.

Discussion

- We develop HAPS:
 - ▶ A conformal framework for constructing prediction intervals for individual event times using covariate histories;
 - ▶ Substantially shorter prediction intervals than baseline methods while maintaining close-to-nominal coverage.
- HAPS, HAPS-DR, and HAPS-A, each have advantages and disadvantages.
- Preprint: [arXiv:2605.06581](https://arxiv.org/abs/2605.06581)
- Email: yuyao_wang1@brown.edu