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

Yuyao Wang[1], Alexander W. Levis[2], Shu Yang[3], Larry Han[1]

[1]Department of Public Health and Health Sciences, Northeastern University
[2]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania
[3]Department of Statistics, North Carolina State University

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 common 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***:
  – calibration at time origin using only baseline covariates;
  – marginal coverage guarantees over the entire population.
- **Limitation: not informative.**

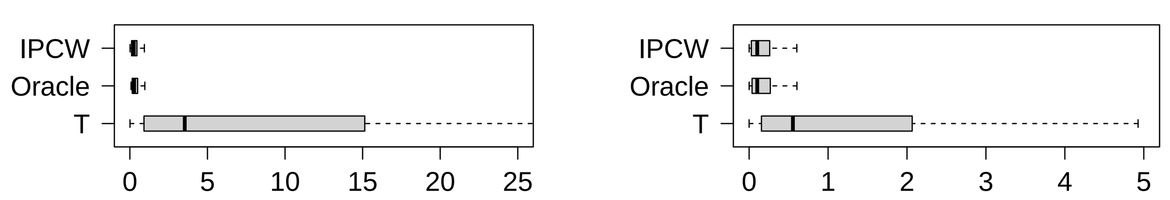

Figure: Boxplots of 90% LPBs under (left) simulation setup 1 in Gui et al. (2024), and (right) simulation setup 1 in Farina et al. (2025).

Dynamic survival prediction

- Predict individual event times T for those with $T > \tau$, using covariate histories till τ.
- Modern dynamic prediction algorithms: approaches based on super-learner, random survival forests, neural networks, support vector machines.
- **Lack uncertainty quantification - limiting use in practice**

Notation and assumptions

- T: event time; C: censoring time.
- Z_t: covariates measured at t; $\bar{Z}_t = \{Z_s : 0 \le s \le t\}$.
- Observed data: an i.i.d. sample $\mathcal{D} = \{X, \Delta, \bar{Z}_X\}_{i=1}^n$, where $X = \min(T, C)$, $\Delta = \mathbb{1}(T < C)$.
- $\lambda_C(t|\cdot) = \lim_{h\to 0^+} \mathbb{P}(t \le C < t + h \mid \cdot, C \ge t, T \ge t)/h$
- $G(t|\bar{Z}_t) = \exp\{-\int_0^t \lambda_C(u|\bar{Z}_u)\,du\} = \mathbb{P}(C > t \mid \bar{Z}_t)$
- $s_\tau = \mathbb{P}(T > \tau)$

Prediction target

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

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

- ***Probably Asymptotically Approximately Correct* (PAAC)** coverage: with prob. $\ge 1 - \epsilon$ over $\mathcal{D}$,

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

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..

History-Aware Prediction Sets (HAPS)

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

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

With censored data, apply IPCW (Robins & Rotnitzky, 1992; Rotnitzky & Robins, 2005):

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

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 parameters G, and denote the estimate as $\hat{G}$.
4. On $\mathcal{D}_{\text{cal}}$, compute $\hat{\theta} = \inf\Big\{\theta \in \Theta : \sum_{i\in\mathcal{D}_{\text{cal}}} U_i(\theta; \hat{G}, \hat{\mathcal{A}}) \ge 0\Big\}$.
5. **Output:** prediction set $\hat{\mathcal{C}}(\bar{Z}_{\tau,n+1}) = \mathcal{C}_{\tau,\hat{\theta}}(\bar{Z}_{\tau,n+1}; \hat{\mathcal{A}})$.

PAAC-type survivor conditional coverage

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

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

Two doubly robust extensions

HAPS-DR: doubly robust post-processing

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

- Achieve 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,

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

where $h_\tau(u, \bar{Z}_u; \theta, \mathcal{A}) = \mathbb{E}\Big(\mathbb{1}(T > \tau)\Big[\mathbb{1}\{T \in \mathcal{C}_{\tau,\theta}(\bar{Z}_\tau; \mathcal{A})\} - (1 - \alpha)\Big]\Big|\, \bar{Z}_u, T \ge u\Big)$,
$N_C(t) = \mathbb{1}(X \le t, \Delta = 0)$, $M_C(t) = N_C(t) - \int_0^t \mathbb{1}(X \ge u)\lambda_C(u|\bar{Z}_u)du$,

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

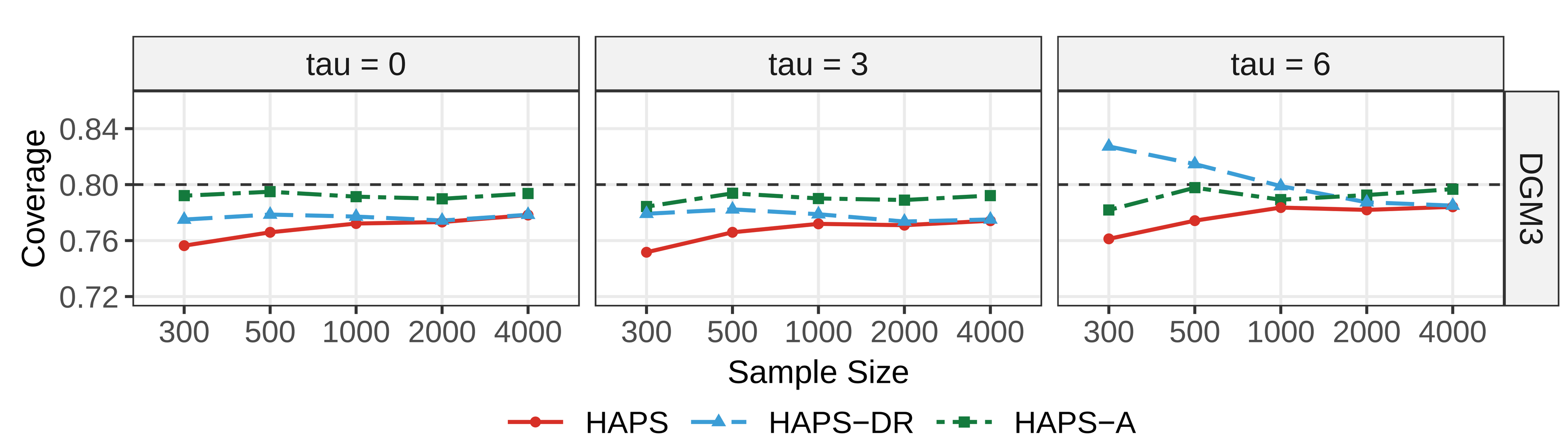

Figure: Mean coverage and median interval length of 80% prediction intervals from HAPS, HAPS-DR, and HAPS-A across sample sizes, under 200 Monte Carlo replications.

Simulation

– Sample size $n = 1000$; test data with size 500 and no censoring.
– DGM: $B \sim Bernoulli(0.5)$, $L_k \sim AR(1)$; censoring follows a mixture of uniform distribution and Weibull distribution depending on covariates.
– Censoring rate: 32.1%

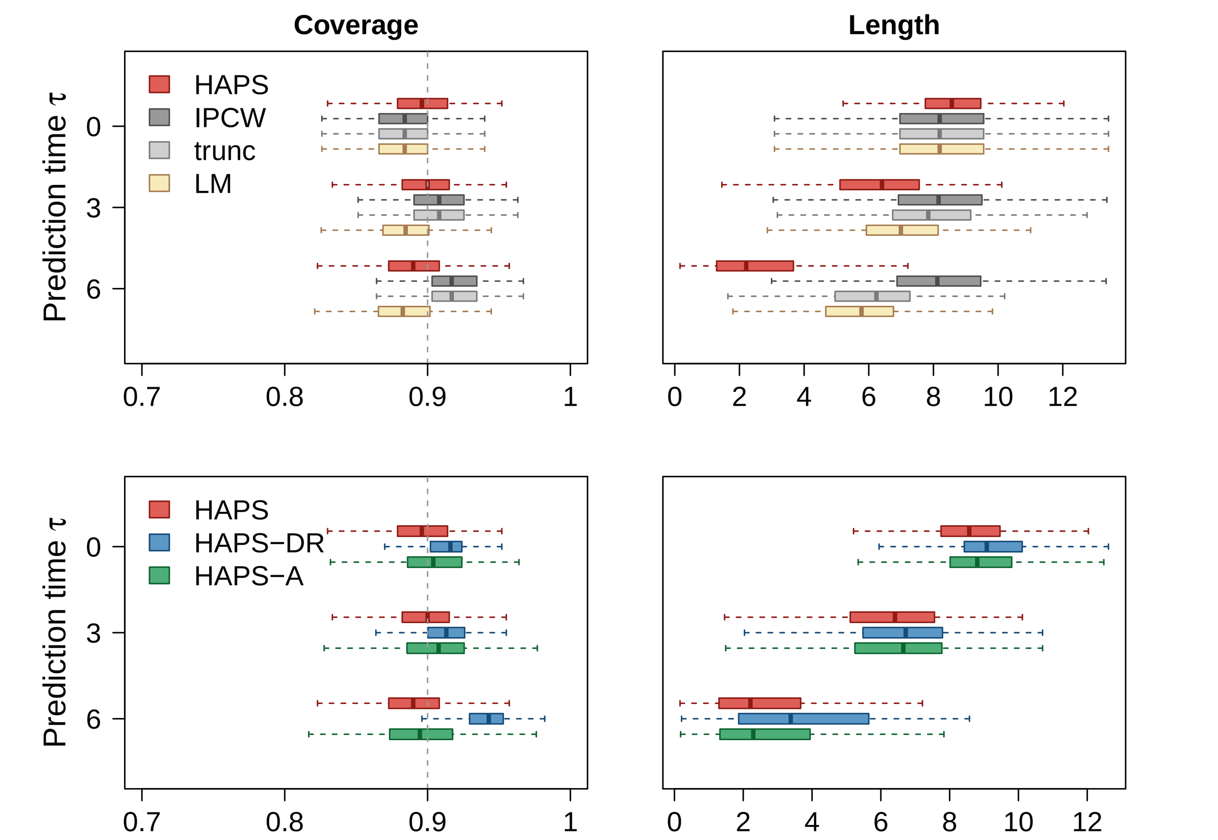

Figure: 90% prediction intervals under 200 Monte Carlo replications, with Dynamic-DeepHit prediction model and RSF/XGB-XGB nuisance models.

- Relative to baseline methods, the median interval length is reduced by **5%–22%** at $\tau = 3$ and **34%–68%** at $\tau = 6$.

Application: Colon cancer data

- Colon cancer: 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).

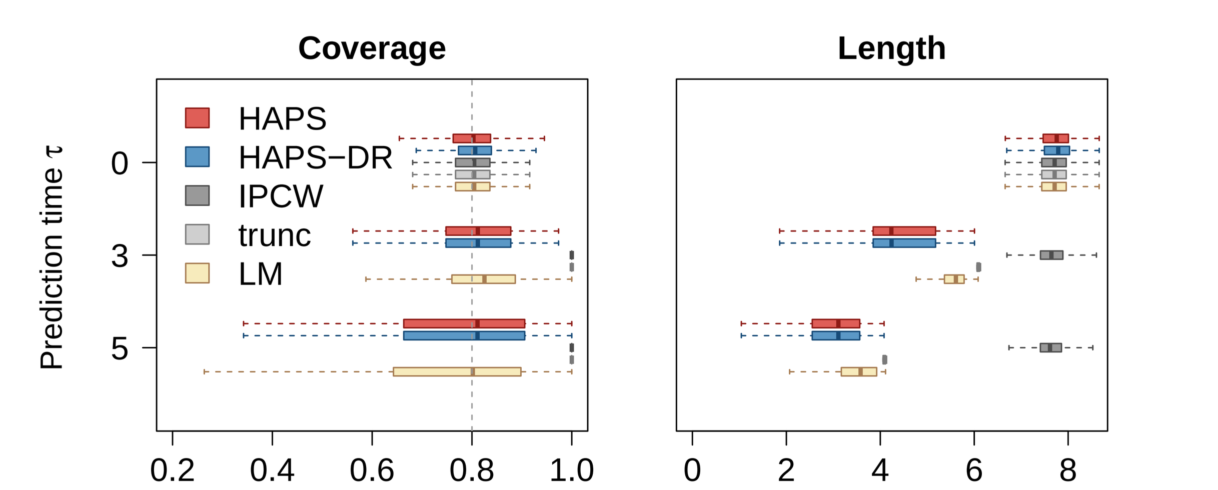

Figure: 80% prediction intervals with Dynamic-DeepHit prediction model across 200 random splits for colon cancer data.

- Relative to baseline methods, the median interval length is reduced by **22%–44%** at $\tau = 3$ and **13%–60%** at $\tau = 5$.

References and Contact Information

- Preprint: arXiv:2605.06581
- GitHub: https://github.com/wangyuyao98/HAPS
- Email: yuya.wang@northeastern.edu