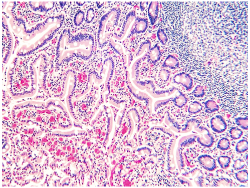
# Conformalized Survival Analysis for General Right-Censored Data

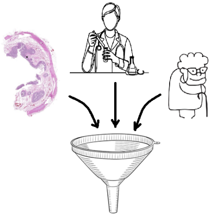
Hen Davidov, Shai Feldman, Gil Shamai, Ron Kimmel,
Yaniv Romano

# Motivation

- How long will a patient survive?
- How long until a prisoner reoffends?
- How long until a machine fails?



# Survival Analysis



$X$

$T$ = time to event (e.g. survival)

$C$ = time to censorship 

We want to reliably predict $T$ from $X$.

We obseve $\mathcal{D} = \{X_i, \tilde{T}_i = \min(T_i, C_i), e_i = \mathbb{I}[T_i < C_i]\}_{i \in [n]}$.

# Current Solutions

## Assumptions

1. Conditionally independent censoring: $C \perp T \,|\, X$ (Identifiability)
2. IID Data: $\{(X_i, T_i, C_i)\}_{i=1}^{n}$.

## Methods to estimate $T|X$

- Non-parametric models: Kaplan-Meier, Nelson-Aalen, etc.
- Semi-parametric models: Cox, AFT, etc.
- Recent Deep Learning methods.

## Limitations

Model misspecification, opacity, and lack of validity.

# PAC-type LPB

**Definition**

$\hat{L}(x)$ is a **Marginally calibrated PAC-type Lower Predictive Bound (LPB)** at level $\alpha \in (0,1)$ with tolerance $\delta \in (0,1)$ if, with probability at least $1-\delta$ over the realization of $\mathcal{D}$,

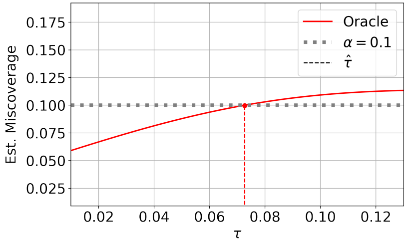
$$\mathbb{P}(T_{\text{test}} \geq \hat{L}(X_{\text{test}}) \mid \mathcal{D}) \geq 1-\alpha.$$

*"The survival time of a patient with the given medical records is at least 24 months with 90% confidence."*

# General approach

We adapt the approach by Gui et al.[1,2]:

1. Train $\{\hat{q}_\tau(x)\}$ to est. the $\tau$ quantile of $T \mid X = x$.
2. For each $\tau$, estimate $\mathbb{P}(T \leq \hat{q}_\tau)$ using a miscoverage est.
3. Choose $\hat{L}(x) = \hat{q}_{\hat{\tau}}(x)$, where $\hat{\tau}$ is chosen as ilustrated:




**Key Challenge**

How do we construct miscoverage estimator if we don't always observe $T$?

---

[1] Gui, Yu, et al. "Conformalized survival analysis with adaptive cut-offs." *Biometrika* 111.2 (2024): 459-477.

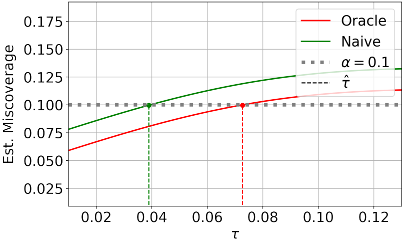
[2] This research also builds on the work of Candès, Emmanuel, Lihua Lei, and Zhimei Ren. "Conformalized survival analysis." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 85.1 (2023): 24-45.

# Naive solution

- Key observation - $\tilde{T} \leq T$.
- The Naive miscoverage estimator is the prop. of points for which $\tilde{T}_i < \hat{q}_\tau(X_i)$.

**Theorem (informal)**

The LPB produced by the Naive solution is a valid PAC-type LPB for $T$.




**Limitation**

Too conservative when there are very small $C$s, resulting in uninformative LPBs.

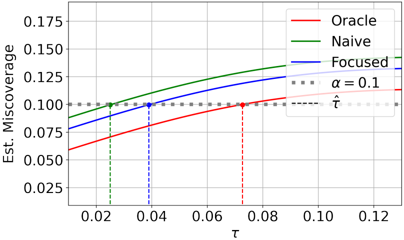
# First proposal - Focused Calibration

- How can we utilize our the fact that we know whether $T$ is observed ($e = 1$) for each data point?
- The Focused miscoverage estimator is the prop. of uncensored points for which $T_i = \tilde{T}_i < \hat{q}_\tau(X_i)$, reweighted to account for the distribution shift.

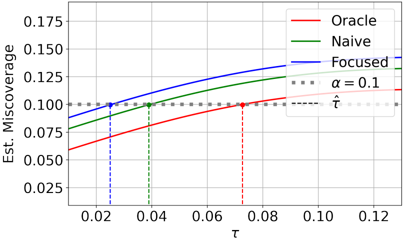
**Theorem (informal)**

Focused Calibration yields a valid, doubly robust PAC-type lower prediction bound for $T$.

# When is Focused better than Naive?




Focused calibration is powerful for data with **many** early censorship events.

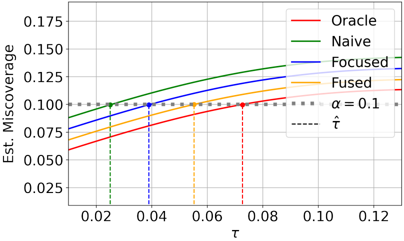



But the Naive solution is preferable for data with **few** early censorship events.

See Proposition 3.1 for rigorus characterisation.

# Can we Fuse the two methods?

- The Fused method dynamically intrapolates between Focused and Naive methods, by dropping samples with early censorship and ratining those with high $C$.
- The Fused miscoverage estimator is the prop. of remaining points that hold $\tilde{T}_i < \hat{q}_\tau(X_i)$, again reweighted.

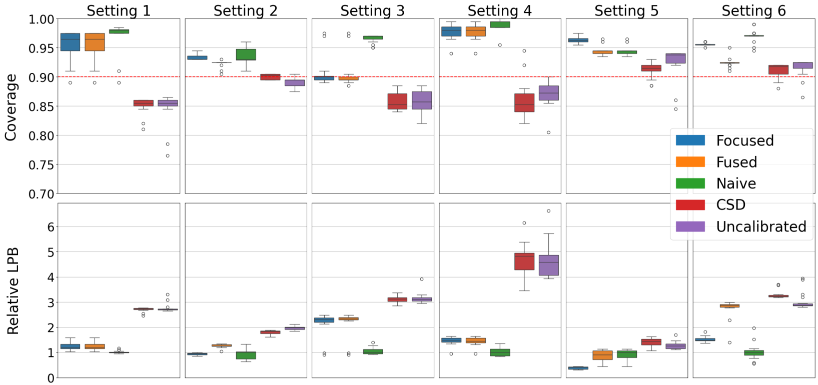



**Theorem (informal)**

Fused Calibration yields a valid doubly robust PAC-type lower prediction bound for $T$.

# Synthetic experiments

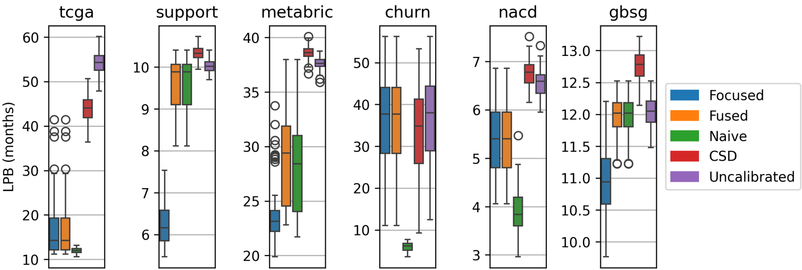
Evaluated LPB size & coverage on 6 synthetic datasets.[3]




[3] Qi, Shi-ang, Yakun Yu, and Russell Greiner. "Conformalized survival distributions: A generic post-process to increase calibration." *arXiv preprint arXiv:2405.07374 (2024)*.

# Real data experiments

Evaluated LPB size on 6 real datasets.




# Thank you!

# Formal definitions of the miscoverage estimators

- The Naive miscoverage estimator:

$$\hat{\alpha}(\tau) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left\{ \tilde{T}_i < \hat{q}_\tau(X_i) \right\},$$

- The Focused miscoverage estimator:

$$\hat{\alpha}(\tau) = \frac{\sum_{i=1}^{n} \hat{w}(X_i)\, \mathbb{I}\{e_i = 1\}\, \mathbb{I}\{\tilde{T}_i < \hat{q}_\tau(X_i)\}}{\sum_{i=1}^{n} \hat{w}(X_i)\, \mathbb{I}\{e_i = 1\}},$$

with $\hat{w}(X_i) = \frac{1}{\hat{\mathbb{P}}(e_i=1|X_i)}$ correcting for shift in $X$.

- The Fused miscoverage estimator:

$$\hat{\alpha}(\tau) = \frac{\sum_{i=1}^{n} \hat{w}_\tau(X_i)\, \mathbb{I}\{\hat{s}_\tau(X_i) = 1 \text{ or } e_i = 1\}\, \mathbb{I}\{\tilde{T}_i < \hat{q}_\tau(X_i)\}}{\sum_{i=1}^{n} \hat{w}_\tau(X_i)\, \mathbb{I}\{\hat{s}_\tau(X_i) = 1 \text{ or } e_i = 1\}},$$

where $\hat{w}_\tau(X_i) = \frac{1}{\hat{\mathbb{P}}(\hat{s}_\tau(X_i)=1 \text{ or } e_i=1|X_i)}$.