

# Bayesian survival time prediction

## for high-dimensional causal effects estimation using regression trees

Tijn Jacobs\*, Wessel van Wieringen, Stéphanie van der Pas  
Department of Mathematics, Vrije Universiteit Amsterdam



### Problem

Estimating **causal effects** for survival data in settings with many more predictors than observations ( $p > n$ ) is challenging.

We require a method to predict the survival times:

$$\mathbb{E}[T \mid A = a, X = x],$$

given  $A$  (treatment) and  $X \in \mathbb{R}^p$  (covariates) that captures nonlinear relationships and interaction effects. Additionally, the method should adapt to different levels of sparsity while retaining all variables in the model.

### Results

We performed a simulation study with the following settings:

- Sample size of  $n = 100$  observations
- Censoring rate is ca. 30%
- Noise level is ca. 20%
- Covariates  $X \in \mathbb{R}^p$  are drawn independently from  $U[0,1]$

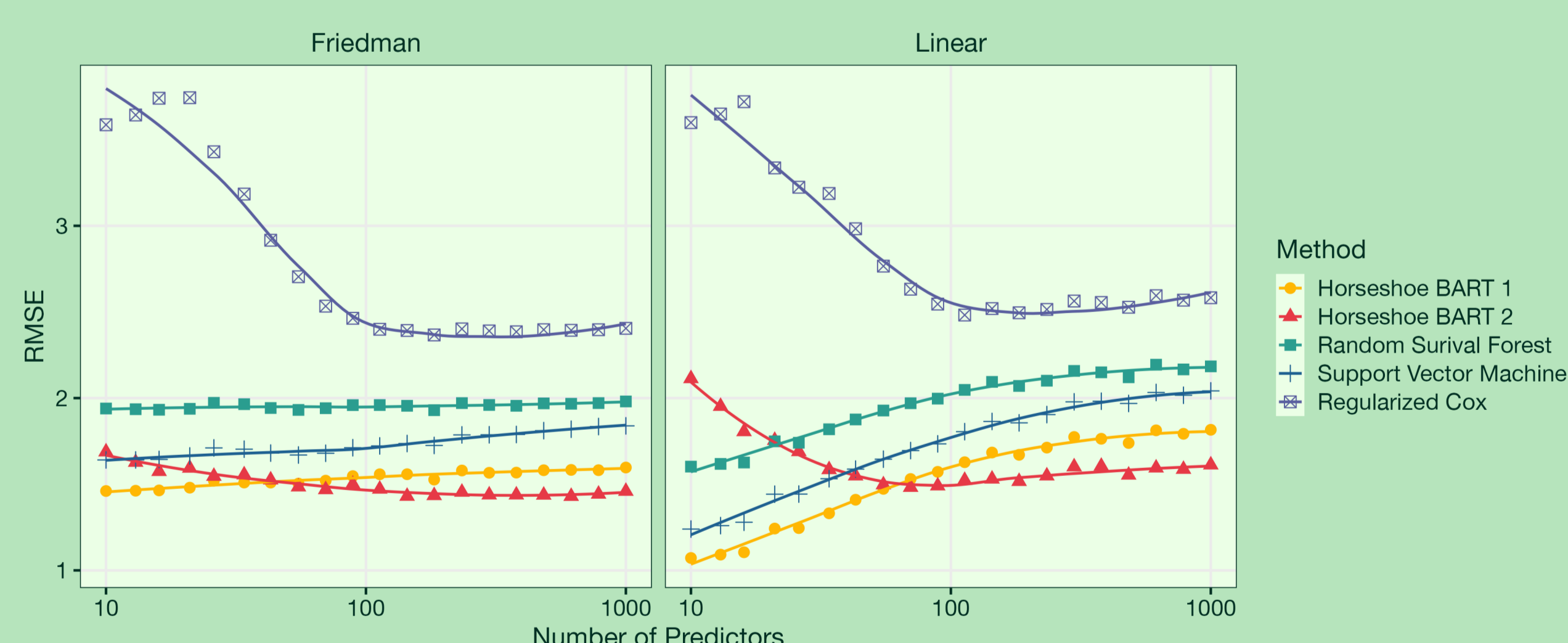
We consider two data-generating processes:

1. Linear model with 10% of nonzero coefficients:

$$\log(T) = \beta^T X + \epsilon.$$

2. Nonlinear model with interactions terms (**Friedman function**):

$$\log(T) = 10 \sin(\pi x_1 x_2) + 20(x_3 - 1/2)^2 + 10x_4 + 5x_5 + \epsilon.$$



Methods are compared using the root mean square error (RMSE), evaluated on a held-out set of size 1000.

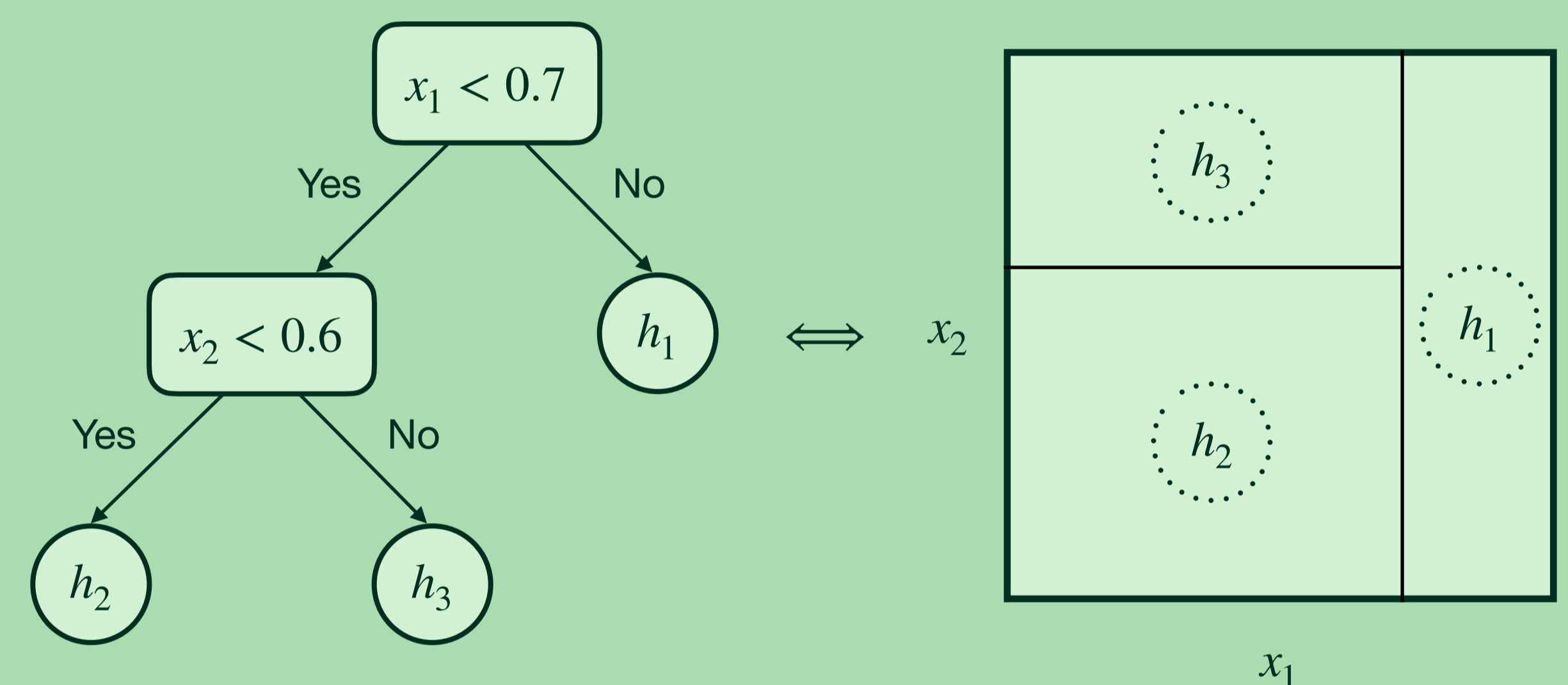
### Model

The survival time is modelled using an *accelerated failure time (AFT)* model:

$$\log(T) = \sum_{j=1}^m f(X, a; \theta_j) + \epsilon,$$

with a *sum-of-trees* regression function and  $\epsilon \sim \mathcal{N}(0, \sigma^2)$ .

Each tree recursively partitions the covariate space, assigning a parameter value to each region. A two-dimensional example of a single tree:



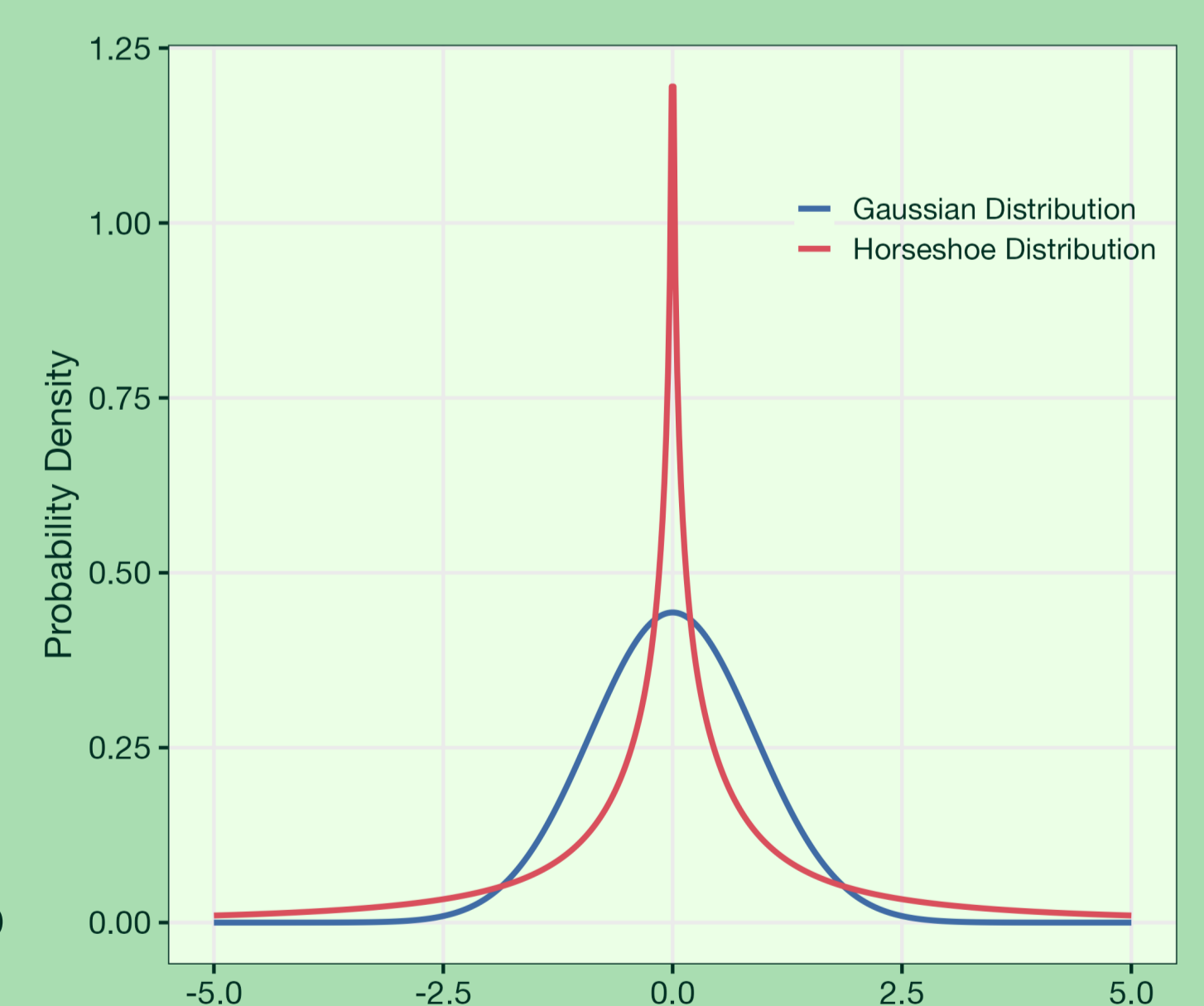
For each tree, we model the step heights  $h$  using a **horseshoe prior**, which provides adaptive shrinkage. This prior is placed on the distribution of  $h$  as follows:

$$\begin{aligned} h_j \mid \lambda_j^2, \tau^2, \sigma^2 &\sim \mathcal{N}(0, \lambda_j^2 \tau^2 \sigma^2), \\ \lambda_j^2 &\sim C^+(0, 1), \\ \tau^2 &\sim C^+(0, 1), \end{aligned}$$

where  $C^+$  denotes the half-Cauchy distribution.

This allows for global shrinkage via  $\tau$  and for local signals to be picked up via the  $\lambda$ 's.

We use a non-informative inverse gamma prior for  $\sigma^2$  while the trees follow a BART prior. Posterior samples are drawn using an efficient reversible jump Markov chain Monte Carlo (RJ-MCMC) algorithm.



### Conclusion

Our method effectively captures nonlinearity, interactions, and adapts to sparsity, outperforming alternatives in high-dimensional survival prediction while retaining all variables for causal effect estimation.

