Estimation and Inference of Causal Effects for Spatially Clustered Survival Data: A Non-parametric Bayesian Approach

Durbadal Ghosh

May 6, 2025

Florida Cancer Registry

- **Data:**
  - 76,106 breast cancer patients from 67 counties in FL
  - Spatial clustering by county
  - Time-to-death (right-censored)
  - Treatment: delay ($>$90 days) vs no delay
- **Goal:**
  - Estimate average and county-level causal effects
  - Adjust for patient and county-level confounders
  - Address spatial clustering
- **Our Approach:**
  - Bayesian non-parametric model using BART
  - Handles spatial clustering, censoring, and confounding

Clustering, Spatial Effects, and BART in Causal Inference

- **Why Account for Clustering and Spatial association?**
  - Cluster members share hidden traits.
  - Ignoring structure risks biased inference.
  - Nearby units face similar environments.
  - County-specific treatment effects are crucial.
- **Why BART?**
  - Captures complex, nonlinear effects.
  - Provides uncertainty quantification
  - Promising in causal settings.

Our Contributions

- **Novel Integration:**
  - Combines BART with spatial random effects
  - Tailored for censored and clustered survival data
- **Methodological Advantages:**
  - Complex treatment–outcome patterns in presence of confounding
  - Spatial correlation via CAR priors
  - Doubly robust two-stage procedure
  - Full posterior for County-specific effects

Notation and Definitions

- $i = 1, \ldots, K$: clusters; $j = 1, \ldots, n_i$: subjects; $N = \sum_i n_i$
- Treatment: $z_{ij} \in \{0, 1\}$
- Confounders: $\mathbf{x}_{ij}$ (individual), $\mathbf{v}_i$ (cluster-level)
- Survival time: T_{ij}, censoring time: C_{ij}
- Observed: $y_{ij} = \min(T_{ij}, C_{ij})$, $\delta_{ij} = \mathbb{1}(T_{ij} < C_{ij})$
- Counterfactuals: $T_{ij}(1), T_{ij}(0), C_{ij}(1), C_{ij}(0)$

Causal Estimands

- **Counterfactual Survival Functions:**
  - $S^{(z)}(t \mid \mathbf{x}, \mathbf{v}) = P(T(z) \geq t \mid \mathbf{x}, \mathbf{v})$
  - $S^{(z)}(t) = \mathbb{E}_{\mathbf{x},\mathbf{v}}[S^{(z)}(t \mid \mathbf{x}, \mathbf{v})]$
- **Conditional Estimands:**
  - $\Delta^{\text{CACE}}(\mathbf{x}, \mathbf{v}) = \mathbb{E}[T(1) - T(0) \mid \mathbf{x}, \mathbf{v}]$
  - $\Delta^{\text{CRACE}}(t, \mathbf{x}, \mathbf{v}) = \mathbb{E}[T(1) \wedge t - T(0) \wedge t \mid \mathbf{x}, \mathbf{v}]$
- **Marginal Estimands:**
  - **SPCE:** $\Delta^{\text{SPCE}}(t) = S^{(1)}(t) - S^{(0)}(t)$
  - **ACE:** $\Delta^{\text{ACE}} = \mathbb{E}_{\mathbf{x},\mathbf{v}}[T(1)] - \mathbb{E}[T(0)]$
  - **RACE:** $\Delta^{\text{RACE}}(t) = \mathbb{E}_{\mathbf{x},\mathbf{v}}[T(1) \wedge t - T(0) \wedge t]$

Causal Assumptions

(A1) SUTVA: For any two subjects j, j' in clusters i, i',

$$T_{ij}(z_{ij}, z_{i'j'}) = T_{ij}(z_{ij})$$

(A2) Consistency:

$$T_{ij} = T_{ij}(1)\mathbb{1}(Z_{ij} = 1) + T_{ij}(0)\mathbb{1}(Z_{ij} = 0)$$

(A3) Unconfoundedness:

$$T_{ij}(z) \perp\!\!\!\perp Z_{ij} \mid \mathbf{x}_{ij}, \mathbf{v}_i, \quad z = 0, 1$$

(A4) Positivity:

$$e_{ij} = e(\mathbf{x}_{ij}, \mathbf{v}_i) = P(Z_{ij} = 1 \mid \mathbf{x}_{ij}, \mathbf{v}_i)$$

is bounded away from 0 and 1.

(A5) Covariate-dependent Censoring:

$$T_{ij}(z) \perp\!\!\!\perp C_{ij}(z) \mid \mathbf{x}_{ij}, \mathbf{v}_i, Z_{ij} = z, \quad z = 0, 1$$

Model for Propensity Score

$P[Z_{ij} = 1 \mid \mathbf{x}_{ij}, \mathbf{v_i}] = e_{ij}$ (Propensity score)
Two-stage implementation:

1. Estimate PS $\hat{e}_{ij}$
2. Plug in $\hat{e}_{ij}$ into the survival model. (Doubly Robust)

- **Advantages of Two-Stage Approach:**
  - Issue with selection bias: $\Delta(z) = E[T(z) \mid Z = z] - E[T(z)]$
  - Provides double robustness: consistent if either model is correctly specified
  - Avoids feedback between treatment and outcome models
  - Computationally efficient: separate estimation of PS and outcome models

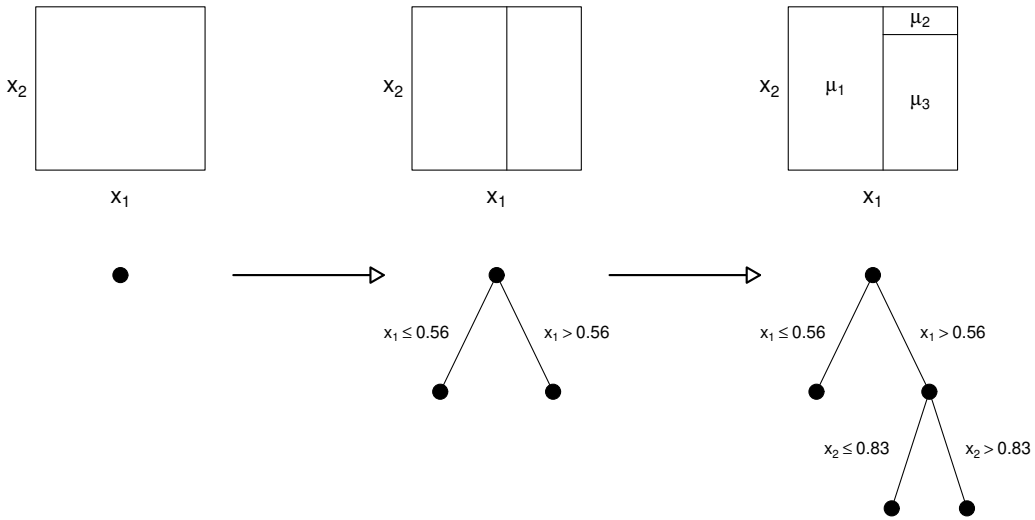
Model for Propensity Score:
$e_{ij} = \Phi(b_1(\mathbf{x}_{ij}, \mathbf{v}_i))$, where $b_1(\cdot) \sim BART$

Bayesian Additive Regression Trees (BART)

- $b(\mathbf{x}) = \sum_{j=1}^{J} g(\mathbf{x}; \tau_j, \mathcal{M}_j)$

  τ_j : topology and splitting rules of tree k

  $\mathcal{M}_j = (\mu_{j1}, \ldots, \mu_{jm_j})$: the set of predictions associated with the m_j terminal nodes of the tree j



SoftBART: Adds smoothness to BART

The Outcome Model

- **Model:**

$$\log T_{ij} = b_2\Big(z_{ij}, \mathbf{x}_{ij}, \mathbf{v}_i, \hat{e}_{ij}\Big) + W_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2).$$

- **BART:**

$$b_2(\cdot) \sim SoftBART$$

- **Spatial Effects:** CAR prior on $\mathbf{W}$:

$$p(\mathbf{W}) \propto \exp\left(-\frac{1}{2\sigma^2_{\mathbf{W}}}\mathbf{W}^\top(D - \rho A)\mathbf{W}\right)$$

- **Censoring:** Truncated normal latent variables .

Simulation Setup

- $n = 3000$ subjects, $K = 15$ clusters (variable sizes)
- Confounders: $X_1, X_2 \sim U(0, 1)$
- Propensity score:

$$e = \text{expit}(0.3X_1 - 0.2X_2 + 0.5X_1X_2)$$

- Treatment: $Z \sim \text{Bernoulli}(e)$
- Censoring: $C \sim \text{Exp}(0.05)$, $y = \min(T, C)$, $\delta = \mathbb{1}(T \leq C)$

Outcome Models and Censoring

Model 1:

$$\log T = b(X, Z) + \mathbf{W} + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

- **Model 2:**

$$\log T = b(X, Z) + \mathbf{W} \cdot \left(1 + 0.5\, Z\, X_1\right) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

- **True Function:**

$$b(X, Z) = \sin(\pi X_1) + \ln(1 + X_2^2) + 2Z\,(X_1 X_2) + (X_1^2)Z.$$

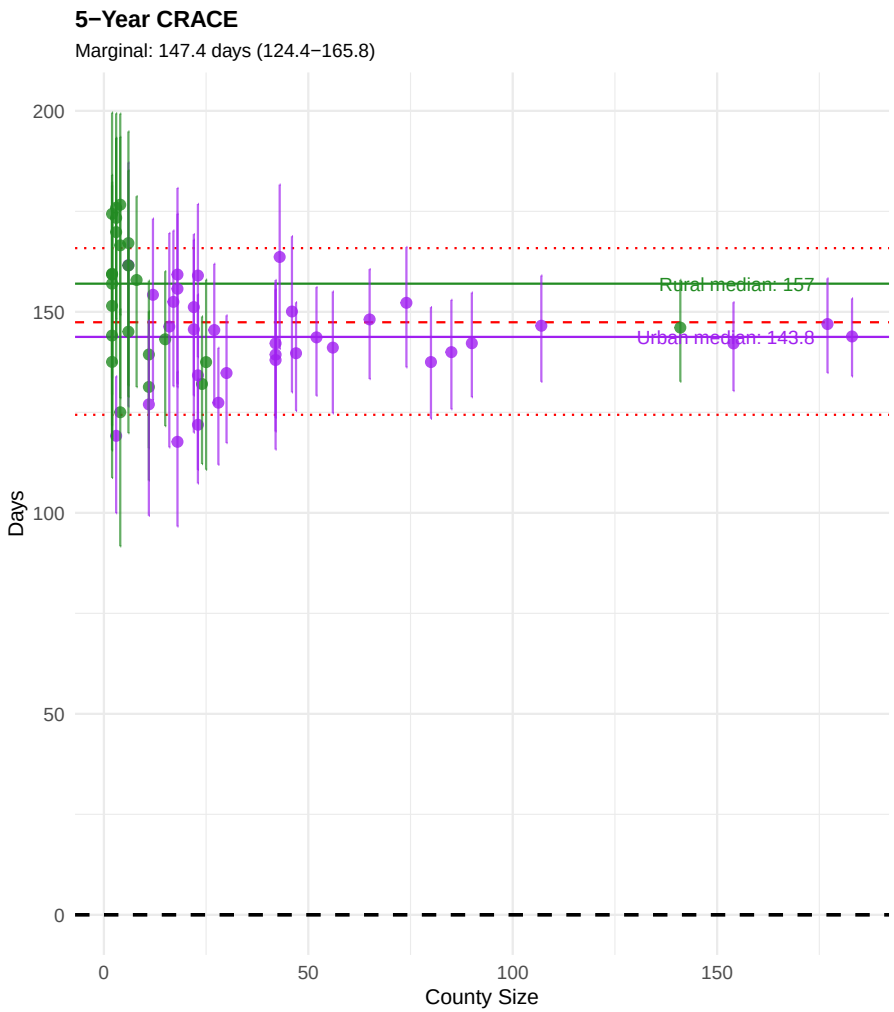
Simulation Results

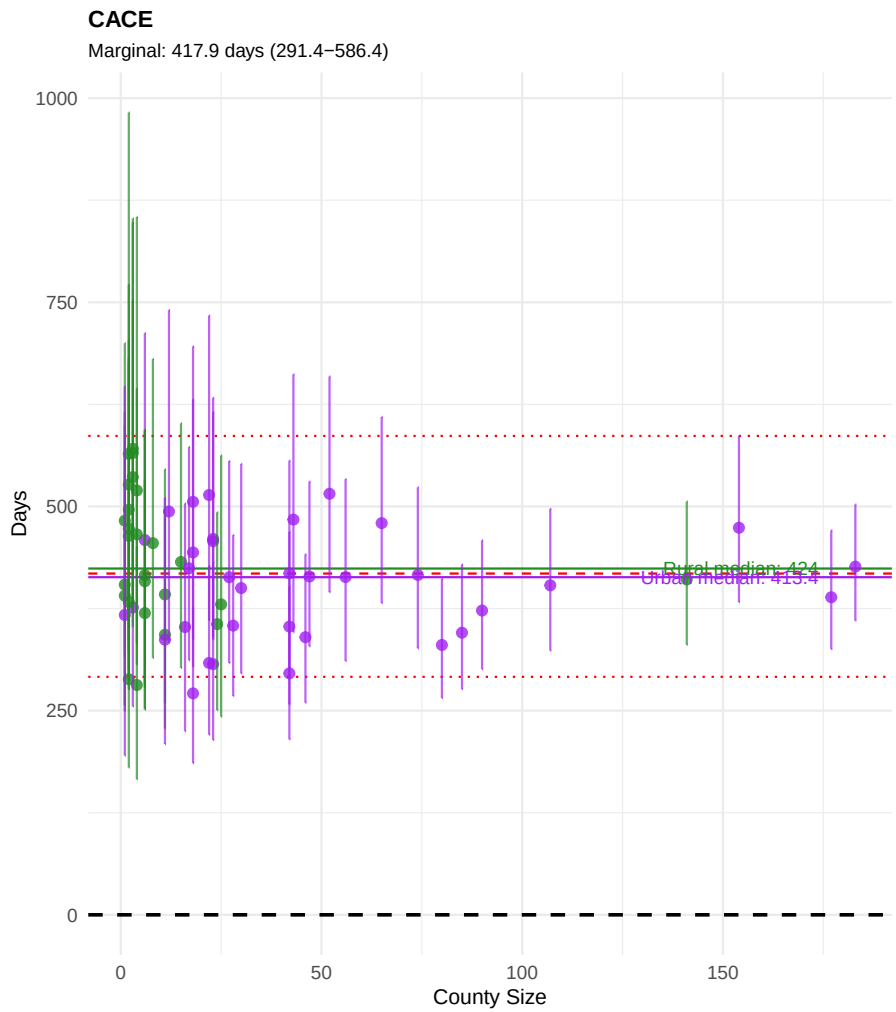
Table: Performance Comparison: Our Method vs Cox Frailty(Haugaard)

Metric	**Estimand**	**Our Method**		**Cox Frailty**	
		M1	**M2**	**M1**	**M2**
Coverage (%)	ACE	70.0	70.0	0.0	0.0
	RACE	86.7	53.3	16.7	13.3
	SPCE	80.0	36.7	0.0	0.0
Abs. Bias	RACE	0.05	0.05	0.18	0.23
	SPCE	0.04	0.06	0.14	0.16
RMSE	RACE	0.12	0.10	0.21	0.26
	SPCE	0.07	0.07	0.14	0.16

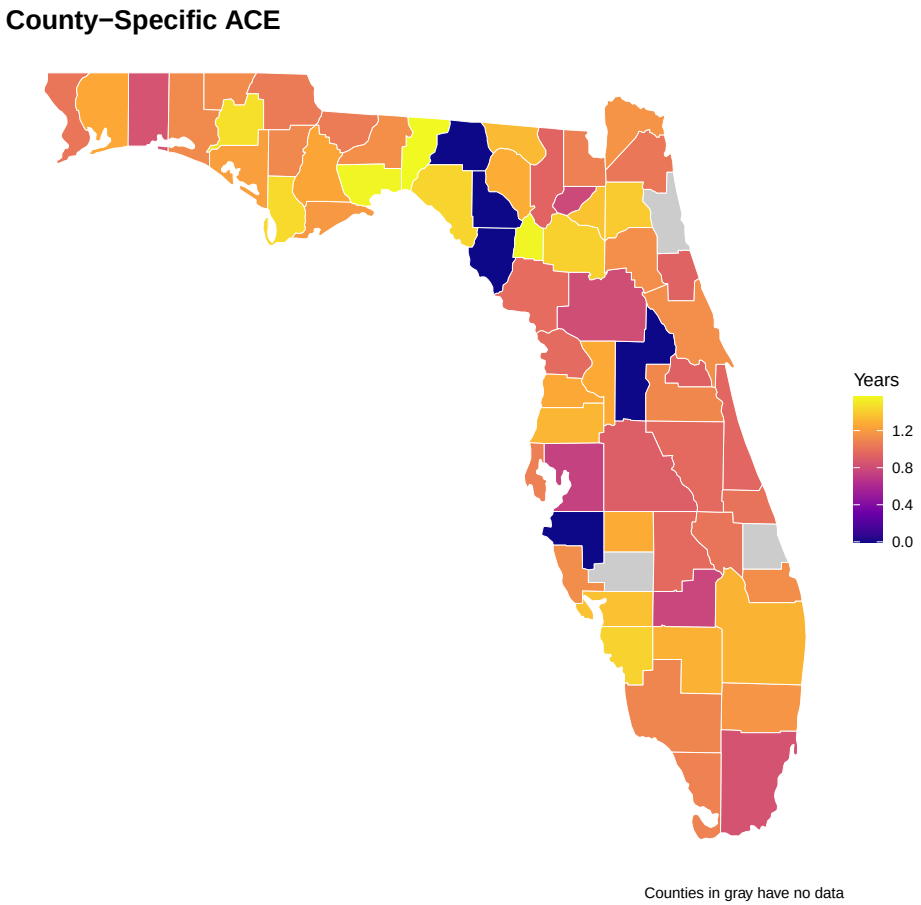
Application Dataset – Florida Cancer Registry

- **Outcome:** Time to death (right-censored)
- **Exposure:** Treatment delay ($>$ 90 days vs no delay)
- **Covariates:**
  - Age: Continous
  - Biopsy Delay : Yes/ No
  - Tumor grade: 1/ 2/ 3
  - HR status: Positive/ Negative
  - Stage: I/ II/ III
  - Race: AA/ WA
- We perform stratified analysis by stage of cancer and race



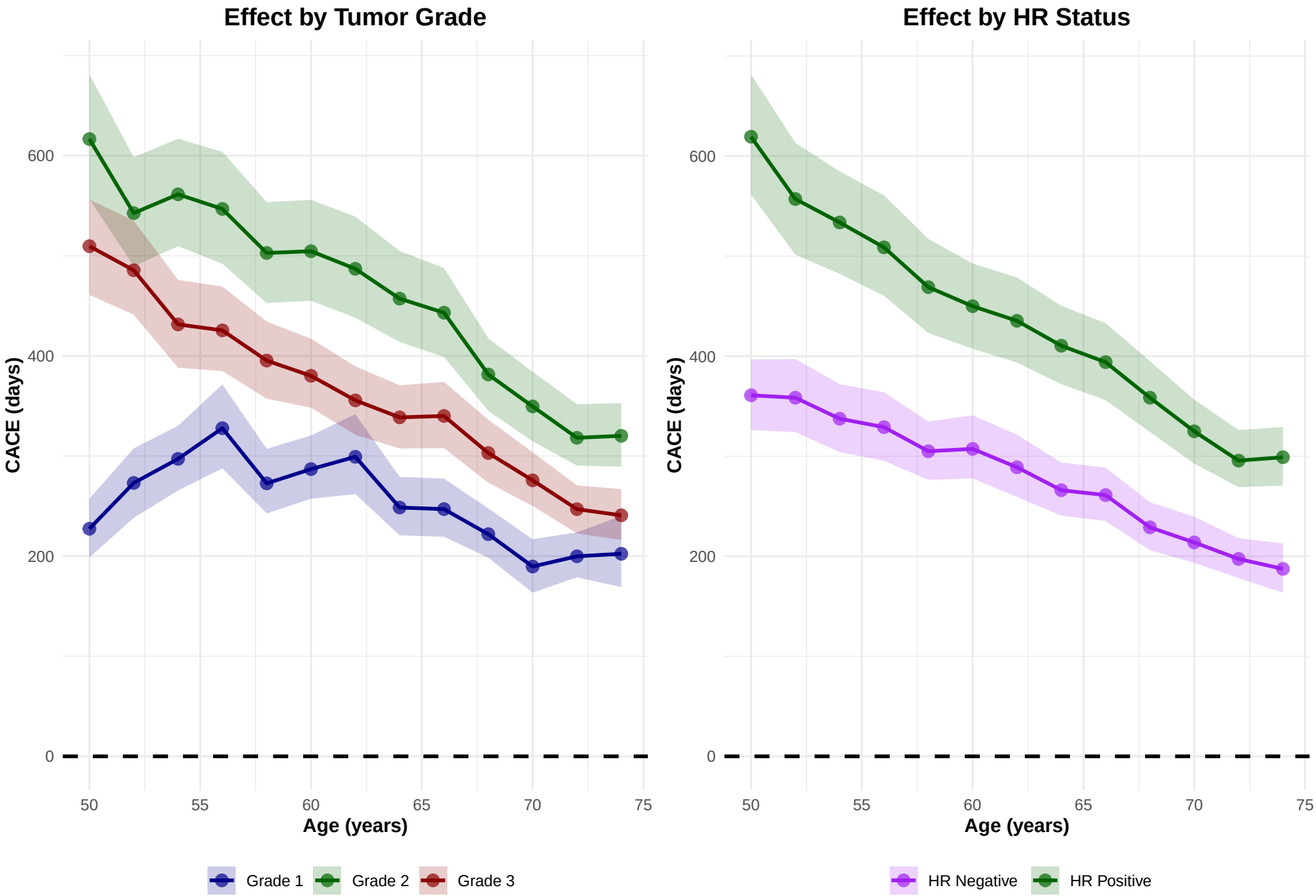


County Type Rural Urban



Geographic Distribution

- Northern regions: highest effects ($>$1.0 year)
- Central Florida: moderate effects ($\sim$0.8 years)
- Scattered counties: minimal effects ($<$0.3 years)
- Spatial clustering suggests regional healthcare factors


Effect by Tumor Grade
CACE (days)
600
400
200
0
50
55
60
65
70
75
Age (years)
Grade 1
Grade 2
Grade 3
Effect by HR Status
CACE (days)
600
400
200
0
50
55
60
65
70
75
Age (years)
HR Negative
HR Positive

Age Effects (50-75 years)

- Consistent decline in benefit with increasing age

Tumor Characteristics

- Grade 2: highest benefit (620→320 days)
- Grade 3: moderate benefit (500→250 days)
- Grade 1: lowest benefit (230→200 days)
- HR+ status: substantially higher benefit (620→300 days)
- HR- status: lower benefit (360→180 days)
- HR status more impactful than tumor grade

Clinical Implications

- Highest priority: young, HR+, Grade 2 patients (>600 days benefit)
- Lowest priority but still beneficial: older, HR-, Grade 1 patients (∼180 days)
- Rural settings: ensure timely access despite geographic barriers
- Benefit persists beyond 5 years for most patients

Insights

County-Level Effects

- Average survival benefit: 417.9 days (95% CI: 291.4–586.4)
- 5-year restricted benefit (CRACE): 147.4 days (95% CI: 124.4–165.8)
- All counties show positive effects regardless of size
- Rural counties: slightly higher benefit (median 424 vs 413.4 days)
- No correlation between county size and effect magnitude

Consistency of Benefits

- CRACE estimates more precise than CACE (narrower CIs)
- Universal positive effect: all counties above zero
- 5-year CRACE shows lasting impact of early treatment

Conclusions

- **Methodological Contributions:**
  - First framework integrating causal inference with spatial survival data
  - SoftBART with CAR priors for complex treatment-outcome relationships
  - Doubly robust estimation for clustered survival outcomes
- **Key Findings:**
  - 418-day survival benefit from avoiding treatment delays
  - Universal positive effects across all counties
  - Strongest benefits: younger, HR+, Grade 2 patients
- **Clinical Implications:**
  - Prioritize timely treatment, especially in high-benefit groups
  - Address regional healthcare disparities
- **Future Work:**
  - Handle missingness in county data
  - Extend to other healthcare settings and outcomes

Thank You

Questions?

- For each subject j in cluster i, we observe

$$y_{ij} = \min(T_{ij}, C_{ij}), \quad \delta_{ij} = \mathbb{1}(T_{ij} < C_{ij}).$$

- Define the model mean (on the log-scale) as

$$\mu_{ij} = f(z_{ij}, \mathbf{x}_{ij}, \mathbf{v}_i, \hat{e}(\mathbf{x}_{ij}, \mathbf{v}_i)) + W_i.$$

$L_{ij}^{\text{obs}}(\theta) = \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log y_{ij} - \mu_{ij})^2}{2\sigma^2} \right] \right\}^{\delta_{ij}} \left\{ 1 - \Phi\left(\frac{\log y_{ij} - \mu_{ij}}{\sigma} \right) \right\}^{1-\delta_{ij}}$, where $\Phi(\cdot)$ is the standard normal CDF.

The full observed-data likelihood is

$$L_{\text{obs}}(\theta) = \prod_{i=1}^{K} \prod_{j=1}^{n_i} L_{ij}^{\text{obs}}(\theta).$$

Data Augmentation

- Define the latent log survival time $\tilde{y}_{ij}$ as

$$\tilde{y}_{ij} = \begin{cases} \text{TruncNormal}\left(\mu_{ij}, \sigma^2; \log y_{ij}\right), & \text{if } \delta_{ij} = 0, \\ \log y_{ij}, & \text{if } \delta_{ij} = 1. \end{cases}$$

Here, TruncNormal$(\mu, \sigma^2; a)$ denotes a $N(\mu, \sigma^2)$ distribution truncated to the interval (a, ∞). The imputed values are used in the complete-data likelihood.

Complete Data Likelihood

- Introduce the latent (complete) log survival times:

$$\tilde{y}_{ij} = \begin{cases} \log y_{ij}, & \delta_{ij} = 1, \\ \text{draw from } N(\mu_{ij}, \sigma^2) \text{ truncated to } [\log y_{ij}, \infty), & \delta_{ij} = 0. \end{cases}$$

- With μ_{ij} defined as before, the complete-data likelihood is

$$L_{\text{complete}}(\theta) = \prod_{i=1}^{K} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\tilde{y}_{ij} - \mu_{ij})^2}{2\sigma^2} \right].$$

Algorithm 1: A Single Iteration

1. **Update Spatial Random Effects & Variance:** Update W, τ^2, and ρ from their full conditionals based on the CAR prior.
2. **Impute Censored Data:** For subjects ij, sample the latent log survival time as

$$\tilde{y}_{ij} = \begin{cases} \text{TruncNormal}\Big(\mu_{ij}, \sigma^2; \log y_{ij}\Big), & \text{if } \Delta_{ij} = 0, \\ \log y_{ij}, & \text{if } \Delta_{ij} = 1. \end{cases}$$

3. **Update BART:** With responses $\tilde{y}_{ij} - W_i$ and covariates $(z_{ij}, \mathbf{x}_{ij})$, update the BART parameters $\{\mathcal{T}_h, \mathcal{M}_h\}$ and the error variance σ^2 via Bayesian backfitting.