Non-parametric Estimation of Causal Effects on Spatially Clustered Survival Data

Durbadal Ghosh

April 15, 2025

Outline

Florida Cancer Registry Dataset

- **Data Structure:**
  - 76,106 breast cancer patients across 67 Florida counties
  - Spatial clustering at county level
  - Survival outcome: Time to death (right censored)
- **Research Context:**
  - Investigate causal effects of treatment delays on survival
  - Need to account for patient-level covariates and county-level effects
  - Challenge: Properly estimating treatment effects while accounting for spatial clustering
- **Methodological Need:**
  - Non-parametric approach to capture complex relationships
  - Framework that handles clustering, censoring, and causal inference simultaneously

Why Clustering is Important in Causal Context

- **Within-Cluster Homogeneity:**
  - Subjects in the same cluster share unobserved factors
  - Ignoring this can lead to confounding and biased treatment effect estimates
  - *Reference:* Imbens & Rubin (2015); Wooldridge (2010)
- **Accurate Uncertainty Quantification:**
  - Independence assumptions fail within clusters
  - Without adjustment, standard errors are underestimated
  - *Reference:* Abadie et al. (2017); Gelman & Hill (2007)
- **Heterogeneous Treatment Effects:**
  - Treatment effects often vary across clusters
  - Clustered analysis enables discovery of effect heterogeneity
  - *Reference:* Wager & Athey (2018); Hill (2011)

Why Spatial Clustering is Important in Causal Inference

- **Spatial Autocorrelation:**
  - Outcomes for spatially proximate observations are correlated
  - Ignoring spatial dependency leads to biased causal estimates
  - *Reference:* Anselin (1995); Tobler's First Law of Geography
- **Unmeasured Spatial Confounders:**
  - Geographic proximity implies shared environmental exposures
  - Spatial models capture unobserved area-level confounding
  - *Reference:* Reich et al. (2021); Schnell & Papadogeorgou (2019)
- **Geographic Variation in Treatment Effects:**
  - Treatment efficacy often varies geographically
  - Spatial models reveal location-specific effects
  - *Reference:* Linero (2020); Papadogeorgou et al. (2018)

Why BART in Causal Inference

- **Flexibility:**
  - Sum-of-trees model captures complex relationships without parametric assumptions
  - Automatically handles nonlinearities and interactions
  - *Reference:* Chipman et al. (2010); Hill (2011)
- **Uncertainty Quantification:**
  - Bayesian framework provides natural uncertainty intervals
  - Direct posterior inference on causal effects
  - *Reference:* Hahn et al. (2020); Dorie et al. (2019)
- **Empirical Performance:**
  - Consistently outperforms other methods in causal inference challenges
  - Robustness to model misspecification
  - *Reference:* Hill & Su (2013); Carnegie et al. (2019)
- **SoftBART Extension:**
  - Introduces smoothness to predictions
  - Particularly valuable for survival outcomes
  - *Reference:* Linero & Yang (2018)

Why Two-Stage Method

- **Addresses Confounding:**
  - First stage: Estimate propensity scores using BART
  - Second stage: Incorporate propensity scores in outcome model
  - *Reference:* Hill & Su (2013); Zigler et al. (2013)
- **Double Robustness:**
  - Consistent estimates if either propensity or outcome model is correct
  - Protects against misspecification of either model
  - *Reference:* Kennedy (2020); Kang & Schafer (2007)
- **Computational Efficiency:**
  - Avoids feedback between outcome and treatment models
  - Manageable with complex, high-dimensional data
  - *Reference:* Hahn et al. (2020); Zeldow et al. (2019)
- **Mitigates Prior Dogmatism:**
  - Reduces bias from regularization-induced confounding
  - Particularly important with high-dimensional covariates
  - *Reference:* Linero (2023); Hahn et al. (2018)

Our Contribution: SoftBART for Causal Inference with Spatially Clustered Survival Data

- **Novel Integration:**
  - First framework combining SoftBART with spatial random effects for causal inference in survival analysis
  - Handles right-censored survival data with spatial clustering
- **Methodological Advantages:**
  - Non-parametric modeling of complex treatment-covariate interactions
  - Spatial correlation through CAR priors for county-level random effects
  - Doubly robust estimation via two-stage procedure
  - Full posterior inference on heterogeneous treatment effects
- **Practical Impact:**
  - Guides location-specific treatment interventions
  - Quantifies uncertainty in estimated causal effects
  - Informs policy decisions at county and state levels

Important Features / Novelties in our method

- We implement the **Two-stage method:** (i) estimate PS $\hat{e}_{ij}$, (ii) plugin $\hat{e}_{ij}$ into the outcome model; Doubly robust
- Non-parametric(mBART) outcome modelling
- Frailty/ Random county effects
- Spatial Association among clusters.

Notation and Definitions

- Study with K clusters; cluster i has n_i subjects; total $N = \sum_{i=1}^{K} n_i$.
- Binary Treatment: For each subject j in cluster i, $Z_{ij} \in \{0, 1\}$.
- For subject j in cluster i:
  - Individual covariates: $\mathbf{X}_{ij}$.
  - Failure time: T_{ij} (possibly right-censored at C_{ij}).
  - Observed outcome: $y_{ij} = \min(T_{ij}, C_{ij})$ with censoring indicator $\delta_{ij} = I(T_{ij} < C_{ij})$.
- Cluster-level covariates: $\mathbf{V}_i$.
- Counterfactual outcomes: $T_{ij}(1)$ and $T_{ij}(0)$ for treatment $Z_{ij} = 1$ and $Z_{ij} = 0$, respectively.

Potential outcomes

- $\{T_{ij}(1), T_{ij}(0)\}$: potential times
- $\{C_{ij}(1), C_{ij}(0)\}$: potential censoring times
- **Under Consistency, the relation between counterfactual and factual data:**

$$T_{ij} = Z_{ij}\, T_{ij}(1) + (1 - Z_{ij})\, T_{ij}(0)$$
$$C_{ij} = Z_{ij}\, C_{ij}(1) + (1 - Z_{ij})\, C_{ij}(0)$$

Causal Assumptions

(A1) SUTVA: For any two subjects j and j' in clusters i and i', with treatment assignments Z_{ij} and $Z_{i'j'}$:

$$T_{ij}(Z_{ij}, Z_{i'j'}) = T_{ij}(Z_{ij}, Z'_{i'j'})$$

(A2) Consistency:

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

(A3) Weak Unconfoundedness:

$$T_{ij}(z) \perp\!\!\!\perp Z_{ij} \mid \mathbf{X}_{ij}, \mathbf{V}_i \quad \text{for } z = 0, 1$$

(A4) Positivity: The propensity score

$$e(\mathbf{X}_{ij}, \mathbf{V}_i) = P(Z_{ij} = 1 \mid \mathbf{X}_{ij}, \mathbf{V}_i)$$

is bounded away from 0 and 1.

(A5) Covariate-dependent Censoring:

$$T_{ij}(z) \perp\!\!\!\perp C_{ij}(z) \mid \mathbf{X}_{ij}, \mathbf{V}_i, Z_{ij} \quad \text{for } z = 0, 1.$$

Estimands

- **Counterfactual survival functions for $z = 0, 1$**
  - $S^{(z)}(t|\mathbf{X}, \mathbf{V}) = P(T(z) \geq t|\mathbf{X}, \mathbf{V})$
  - $S^{(z)}(t) = P(T(z) \geq t) = \mathbb{E}_{\mathbf{X},\mathbf{V}}[S^{(z)}(t|\mathbf{X}, \mathbf{V})]$
- **Survival probability causal effect (SPCE) at t (Mao et al. 2018):**

$$\Delta^{SPCE}(t) := S^{(1)}(t) - S^{(0)}(t)$$

- **Average causal effect (ACE) :**

$$\Delta^{ACE} = \mathbb{E}[T(1)] - \mathbb{E}[T(0)]$$

- **Restricted average causal effect (RACE) :**

$$\Delta^{RACE}(t^*) = \mathbb{E}[\min(T(1), t^*)] - \mathbb{E}[\min(T(0), t^*)]$$

Estimands

- **Conditional survival probability causal effect (CSPCE) at t :**

$$\Delta^{CSPCE}(t, \mathbf{X}, \mathbf{V}) := S^{(1)}(t, \mathbf{X}, \mathbf{V}) - S^{(0)}(t, \mathbf{X}, \mathbf{V})$$

- **Conditional average causal effect (CACE) :**

$$\Delta^{CACE}(\mathbf{X}, \mathbf{V}) = \mathbb{E}[T(1) \mid \mathbf{X}, \mathbf{V}] - \mathbb{E}[T(0) \mid \mathbf{X}, \mathbf{V}]$$

- **Conditional restricted average causal effect (CRACE) :**

$$\Delta^{CRACE}(t^*, \mathbf{X}, \mathbf{V}) = \mathbb{E}[\min(T(1), t^*) \mid \mathbf{X}, \mathbf{V}] - \mathbb{E}[\min(T(0), t^*) \mid \mathbf{X}, \mathbf{V}]$$

Model for Propensity Score

$$P[Z_{ij} = 1 \mid \mathbf{X}_{ij}, \mathbf{V_i}] = e_{ij}, \quad logit(e_{ij}) = g(\mathbf{X}_{ij}), \text{ where } g(\cdot) \sim BART$$

Two-stage implementation: (i) estimate PS $\hat{e}_{ij}$, (ii) plug in $\hat{e}_{ij}$ into the survival model. (Doubly Robust)

The AFT-BART Model with Spatial CAR Prior

- **Model:**

$$\log T_{ij} = f\Big(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i)\Big) + W_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2).$$

- **Spatial Random Effects:**

$$p(W \mid \tau^2, \rho) \propto \exp\Big\{-\frac{1}{2\tau^2} W^\top (D - \rho A) W\Big\},$$

where A is the spatial adjacency matrix, D is diagonal with $d_{ii} = \sum_{i'} A_{ii'}$, and ρ is the spatial parameter.

- **BART:**

$$f(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i)) = \sum_{h=1}^{H} g(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i); \mathcal{T}_h, \mathcal{M}_h).$$

Prior Specification

- Error variance: $\sigma^2 \sim \mathrm{IG}(a_\sigma, b_\sigma)$.
- Spatial variance: $\sigma^2_W \sim \mathrm{IG}(a_W, b_W)$.
- Spatial correlations: $\rho \sim \mathit{Uniform}(\frac{1}{\alpha_{(1)}}, \frac{1}{\alpha_{(K)}})$. $\alpha_{(1)}, \alpha_{(K)}$ are the minimum and maximum eigenvalues of A, respectively.
- Function $f(\cdot) \sim$ SBART

Observed Data Likelihood

- For each subject j in cluster i, we observe

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

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

$$\mu_{ij} = f(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i)) + W_i.$$

- Then the individual likelihood contribution is

$$L_{ij}^{\text{obs}}(\theta) = \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log y_{ij} - \mu_{ij})^2}{2\sigma^2} \right] \right\}^{\delta_{ij}} \left\{ 1 - \Phi\left(\frac{\log y_{ij} - \mu_{ij}}{\sigma} \right) \right\}^{1}$$

  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

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

Model for Propensity Score

$$P[Z_{ij} = 1 \mid \mathbf{X}_{ij}, \mathbf{V_i}] = e_{ij}, \quad logit(e_{ij}) = g(\mathbf{X}_{ij}), \text{ where } g(\cdot) \sim BART$$

Two-stage implementation: (i) estimate PS $\hat{e}_{ij}$, (ii) plug in $\hat{e}_{ij}$ into the survival model. (Doubly Robust)

The AFT-BART Model with Spatial CAR Prior

- **Model:**

$$\log T_{ij} = f\Big(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i)\Big) + W_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2).$$

- **Spatial Random Effects:**

$$p(W \mid \tau^2, \rho) \propto \exp\Big\{-\frac{1}{2\tau^2} W^\top (D - \rho A) W\Big\},$$

where A is the spatial adjacency matrix, D is diagonal with $d_{ii} = \sum_{i'} A_{ii'}$, and ρ is the spatial parameter.

- **BART:**

$$f(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i)) = \sum_{h=1}^{H} g(Z_{ij}, \mathbf{X}_{ij}, \mathbf{V}_i, \hat{e}(\mathbf{X}_{ij}, \mathbf{V}_i); \mathcal{T}_h, \mathcal{M}_h).$$

Prior Specification

- Error variance: $\sigma^2 \sim \mathrm{IG}(a_\sigma, b_\sigma)$.
- Spatial variance: $\sigma^2_W \sim \mathrm{IG}(a_W, b_W)$.
- Spatial correlations: $\rho \sim \mathit{Uniform}(\frac{1}{\alpha_{(1)}}, \frac{1}{\alpha_{(K)}})$. $\alpha_{(1)}, \alpha_{(K)}$ are the minimum and maximum eigenvalues of A, respectively.
- Function $f(\cdot) \sim$ SBART

Simulation : Data Generation (Model Corretly Specified)

- **Subjects:** $n = 3000$ observations.
- **Covariates:** Two confounders:

$$X_1, X_2 \sim U(0, 1).$$

- **Clusters:** $K = 15$ clusters, with subjects randomly assigned.

Treatment Assignment

- **Propensity Score:**

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

  where $\text{expit}(x) = \frac{1}{1+e^{-x}}$.
- **Treatment:** $Z \sim \text{Bernoulli}(p)$.
- **Note:** Treatment assignment depends on the confounders.

Spatial Structure and CAR Random Effects

- **Adjacency:** A structure for $K = 15$ clusters: Based on 10 Florida counties in the Western region.
- **CAR Prior:**

$$p(\mathbf{W} \mid \sigma_W^2, \rho) \propto |Q(\rho)|^{1/2} \exp\Big\{-\frac{1}{2\sigma_W^2}\mathbf{W}^\top Q(\rho)\mathbf{W}\Big\},$$

where

$$Q(\rho) = D - \rho A, \quad D_{ii} = \sum_j A_{ij}.$$

- **Generation:** Cluster effects are drawn from a multivariate normal with covariance $\sigma_W^2 Q(\rho)^{-1}$.

Outcome Model and Censoring

- **True Outcome Model (log-scale):**

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

- **True Function:**

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

- **Censoring:**
  - Censoring times $C \sim \text{Exponential}(\lambda_c)$ (e.g., $\lambda_c = 0.05$).
  - Observed time: $y = \min(T, C)$ and indicator $\delta = I(T \leq C)$.

Simulation : Data Generation (Model Misspecified)

- **Subjects:** $n = 3000$ observations.
- **Covariates:** Two confounders:

$$X_1, X_2 \sim U(0, 1).$$

- **Clusters:** $K = 15$ clusters, with subjects randomly assigned.

Treatment Assignment

- **Propensity Score:**

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

  where $\text{expit}(x) = \frac{1}{1+e^{-x}}$.
- **Treatment:** $Z \sim \text{Bernoulli}(p)$.
- **Note:** Treatment assignment depends on the confounders.

Spatial Structure and CAR Random Effects

- **Adjacency:** A structure for $K = 15$ clusters: Based on 10 Florida counties in the Western region.
- **CAR Prior:**

$$p(\mathbf{W} \mid \sigma_W^2, \rho) \propto |Q(\rho)|^{1/2} \exp\Big\{-\frac{1}{2\sigma_W^2}\mathbf{W}^\top Q(\rho)\mathbf{W}\Big\},$$

where

$$Q(\rho) = D - \rho A, \quad D_{ii} = \sum_j A_{ij}.$$

- **Generation:** Cluster effects are drawn from a multivariate normal with covariance $\sigma_W^2 Q(\rho)^{-1}$.

Outcome Model and Censoring

- **True Outcome Model (log-scale):**

$$\log T = f(X, Z) + W \cdot \Big(1 + 0.5\, Z\, X_1\Big) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

- **True Function:**

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

- **Censoring:**
  - Censoring times $C \sim \text{Exponential}(\lambda_c)$ (e.g., $\lambda_c = 0.05$).
  - Observed time: $y = \min(T, C)$ and indicator $\delta = I(T \leq C)$.

Table: Performance Comparison: SBART vs Cox under Correct and Misspecified Models

Causal Effect	**Metric**	**SBART (Correct)**	**SBART (Misspec)**
ACE	Coverage (%)	70.000	70.000
RACE	Coverage (%)	86.667	53.333
SPCE	Coverage (%)	80.000	36.667
ACE	Absolute Bias	1.516	7.488
RACE	Absolute Bias	0.045	0.049
SPCE	Absolute Bias	0.041	0.055
ACE	RMSE	2.452	11.027
RACE	RMSE	0.121	0.096
SPCE	RMSE	0.066	0.071