Bayesian Supervised Causal Clustering

Luwei Wang [1] Nazir Lone [2] Sohan Seth [1]

[1]School of Informatics, University of Edinburgh
[2]Usher Institute, University of Edinburgh

Motivation

Problem: How to find patient subgroups that are similar in *both* their covariate profiles *and* their treatment responses?

- **Unsupervised clustering** (e.g., GMM) ignores treatment effects → clusters heterogeneous in treatment response.
- **Effect modeling** (e.g., CF, BART) estimates individual effects but lacks interpretable subgroup structure.
- **Causal clustering** clusters on potential outcomes but ignores covariate similarity.
- **Our Solution:** BSCC jointly clusters individuals based on **covariate similarity** and **treatment effect heterogeneity**.

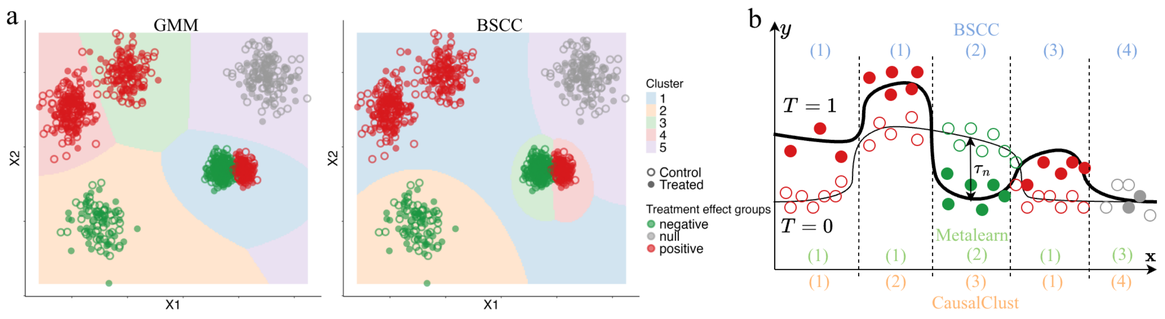

Figure 1. BSCC jointly models covariates and treatment responses.

Comparison of Approaches

Category	Supervision	Probabilistic	Covariates	TreatEffect	Example
Unsupervised	×	✓	✓	×	GMM
Supervised (outcome)	✓	✓	✓	×	SGMM
Subgroup analysis	✓	×	✓	✓	IT, MOB
Effect modeling	✓	×	×	✓	CF, BART
BSCC (ours)	✓	✓	✓	✓	

Method: Generative Framework

Given observed data $\mathcal{D} = \{\mathbf{x}_n, y_n^{\text{obs}}, a_n\}_{n=1}^N$ with K clusters:

$$\begin{aligned} z_n &\sim \text{Cat}(\boldsymbol{\pi}) \\ \mathbf{x}_n &\sim f(\boldsymbol{\theta}_{z_n}) \\ \mu_n^0 &= \mu^0(\mathbf{x}_n; \boldsymbol{\phi}) \\ \tau_n &= \beta_{z_n} \\ y_n^{\text{obs}} &\sim \mathcal{N}(\mu_n^0 + a_n\tau_n,\ \sigma_{a_n}^2) \end{aligned}$$

Key Components:

- **Covariates:** Mixed-type (Gaussian + Bernoulli) with cluster-specific soft feature selection $\boldsymbol{\gamma}_k$.
- **Control outcome:** GP prior $\mu^0(\mathbf{x}) \sim \text{GP}(\mathbf{0}, \mathbf{K}_0)$ with ARD kernel.
- **Treatment effect:** Constant per cluster $\tau_k = \beta_k$ for interpretability.
- **Binary outcome:** $y_n^{\text{obs}} \sim \text{Bern}(\text{logit}^{-1}(\mu_n^0 + a_n\tau_n))$, τ_n as log odds ratio.
- **Inference:** ADVI in RStan with parallel random restarts.

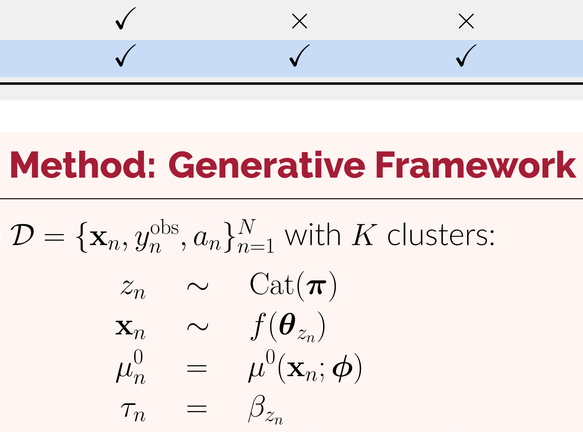

Figure 2. Plate diagram of the BSCC generative model.

Simulation Results

Setup: N=1200, D=12 covariates, 5 clusters with $\boldsymbol{\tau} = (0.5, 5, -5, 0, 0)$.

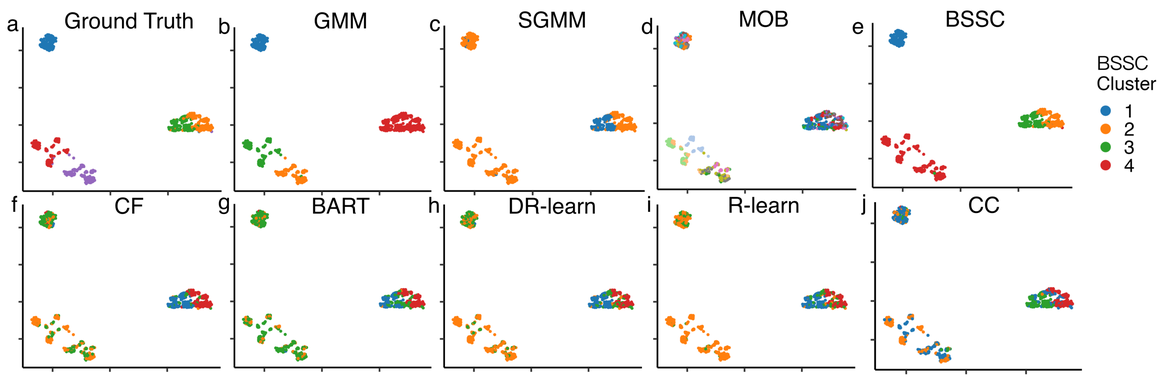

Figure 3. Simulation results: estimated treatment effects vs. ground truth.

Method	ARI	SATE Range	PEHE
GMM	**0.768**	$[-0.40, 1.32]$	3.13
SGMM	0.625	$[-5.54, 0.67]$	2.65
MOB	0.437	$[-4.73, 2.38]$	2.58
BSCC (ours)	**0.721**	$[-5.13, 3.99]$	**1.45**
CF	0.544	$[-4.71, 3.67]$	1.98
BART	0.381	$[-4.46, 4.61]$	1.86
DR-LEARNER	0.548	$[-4.56, 4.25]$	2.14
R-LEARNER	0.508	$[-4.78, 3.57]$	1.72
CC	0.524	$[-3.72, 3.76]$	1.86

Key findings:

- BSCC achieves the **lowest PEHE** (1.45) among all methods.
- SATE range closely matches ground truth $[-5, 5]$.
- GMM has high ARI but fails to capture treatment heterogeneity.
- Robust to treatment imbalance (proportion 0.2 vs 0.5).

Real Application: IST-3 Stroke Trial

Applied to Third International Stroke Trial: N=2737 patients, 12 covariates, binary outcome (6-month mortality).

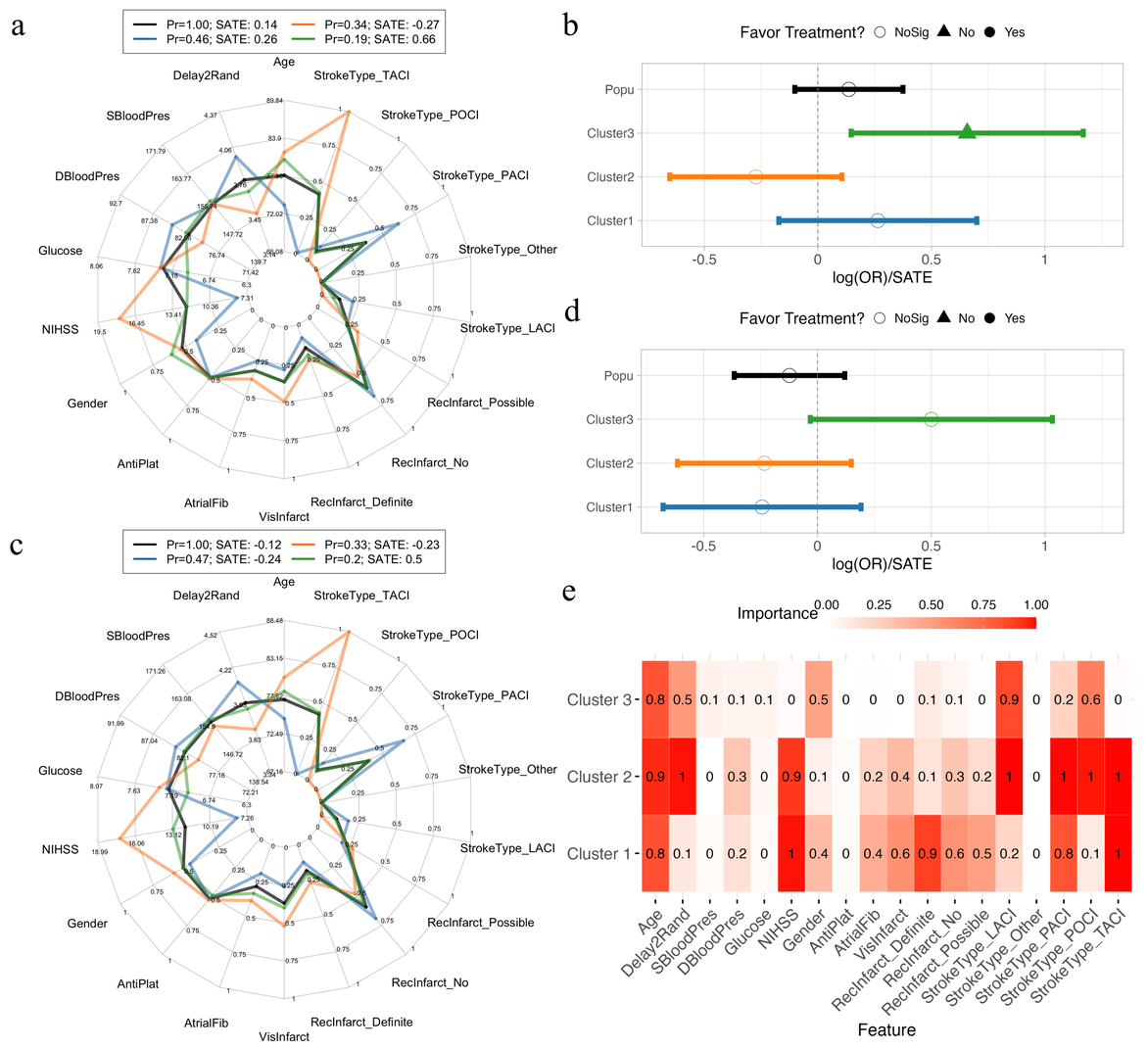

Figure 4. IST-3 results: cluster-specific treatment effects and covariate profiles.

Three clinically meaningful clusters:

- **Cluster 1:** Younger, low NIHSS, milder strokes → 13.3% mortality.
- **Cluster 2:** Older, high NIHSS, severe TACI → 47.6% mortality, treatment beneficial.
- **Cluster 3:** Moderate severity, delayed presentation → 22.8% mortality.

Conclusions & Future Work

Advantages: Joint clustering on covariates and treatment effects **(Interpretability)**; mixed-type covariates with feature selection **(Flexibility)**; ADVI **(Computational efficiency)**; validated on real clinical trial data.

Future Directions: Extension to observational data with propensity score modeling; semi-supervised clustering; multiple treatment arms and time-to-event outcomes.

Website: demi-wlw.github.io

Contact: luwei.wang@ed.ac.uk