Efficient Subgroup Analysis via Optimal Trees with Global Parameter Fusion

Zhongming Xie Joseph Giorgio Jingshen Wang

University of California, Berkeley

Scan for paper

Motivation: Heterogeneity in Alzheimer's Disease

Recent anti-amyloid trials (Lecanemab, Donanemab) show that lowering Amyloid-β (Aβ) plaques can yield cognitive benefits in Alzheimer's Disease (AD), but treatment effects are highly **heterogeneous** across subpopulations:

These findings demand statistical tools that can *uncover* which patients benefit from a treatment, supporting precision medicine and targeted future trials.

Problem 1: Suboptimal, Unstable Partitions

Tree-based methods (CART-style) split **greedily**, top-down: each split is chosen locally without anticipating downstream splits. The resulting tree may pick the *wrong* root variable and miss the true heterogeneity, especially in small samples.

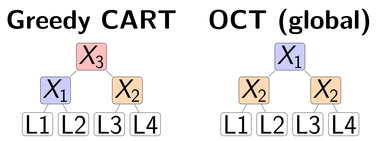

Figure 1. Greedy search picks X_3 at the root and misrepresents heterogeneity.

Goal: replace greedy search with a *global* search over hierarchical tree partitions.

Problem 2: Local estimation

AD outcomes depend on **rare** genetic risk factors (e.g., APOE-ε4, $\sim$5% prevalence). Subgroup-specific estimates are unstable when each leaf is fit in *isolation*. Reliable inference requires **borrowing strength** across leaves—but greedy trees can't enforce such global constraints.

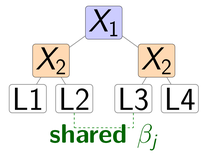

Figure 2. FOCT fuses coefficients across selected leaves (here L2 and L3) so rare-covariate effects can borrow strength.

Goal: a tree method that *selectively fuses* parameters across leaves to share information.

Our Contribution: Fused Optimal Causal Tree (FOCT)

A unified **mixed-integer optimization (MIO)** formulation that delivers:

- **Globally optimal** subgroup partitions.
- A **parameter-fusion** (L_0) constraint that lets related leaves share coefficients, stabilizing estimation in small samples.

Problem Setup

Observe i.i.d. $(\boldsymbol{X}_i, T_i, Y_i) \sim \mathbb{P}_0$ with covariates $\boldsymbol{X}_i \in \mathbb{R}^d$, binary treatment T_i, and outcome Y_i. There exist M^* unknown subpopulations $\mathcal{A}_1^*, \ldots, \mathcal{A}_{M^*}^*$ such that

$$Y_i = \sum_{m=1}^{M^*} \mathbf{1}_{(\boldsymbol{X}_i \in \mathcal{A}_m^*)} \big(\delta_m^* + \mu_m^* T_i + \boldsymbol{\alpha}_m^{*\top} \boldsymbol{X}_i + \boldsymbol{\beta}_m^{*\top} T_i \boldsymbol{X}_i\big) + \epsilon_i.$$

Let $\boldsymbol{Z}_i = [1, T_i, \boldsymbol{X}_i^\top, T_i \boldsymbol{X}_i^\top]^\top$ and $\boldsymbol{\gamma}_m^* = (\delta_m^*, \mu_m^*, \boldsymbol{\alpha}_m^{*\top}, \boldsymbol{\beta}_m^{*\top})^\top$.

Goal: Recover the partition $\Pi^* = \{\mathcal{A}_m^*\}_{m=1}^{M^*}$, the coefficients $\{\boldsymbol{\gamma}_m^*\}$, and the subgroup average treatment effect (SATE) $\tau(\boldsymbol{X}_i, \Pi^*) = \sum_m \mathbf{1}_{(\boldsymbol{X}_i \in \mathcal{A}_m^*)} (\mu_m^* + \boldsymbol{\beta}_m^{*\top} \bar{\boldsymbol{X}}_m)$.

Fused and Globally Optimal Causal Tree

We solve the *single* optimization problem

$$\min_{\boldsymbol{\gamma}_m, \Pi \in \mathcal{T}} \sum_{i=1}^{n} \Big\| Y_i - \sum_{m=1}^{M(\Pi)} \mathbf{1}_{(\boldsymbol{X}_i \in \mathcal{A}_m)} \boldsymbol{\gamma}_m^\top \boldsymbol{Z}_i \Big\|_2^2$$

subject to fusion: $\gamma_{l_1,j} = \gamma_{l_2,j}$ for some $l_1 \neq l_2$ and some j.

Two ingredients:

- The objective promotes **heterogeneous causal discovery** across leaves.
- Fusion enables **selective information sharing** across leaves and covariates.

MIO Formulation (Algorithm 1)

Optimization objective:

$$\min_{\Theta \in \boldsymbol{\Theta}} \frac{L_n\big(\mathcal{T}_L, \{\boldsymbol{\gamma}_{t'}\}_{t' \in \mathcal{T}_L}\big)}{\hat{L}(\mathcal{D}_n)} + \lambda \sum_{j \in [2d+2],} \sum_{t_1, t_2 \in \mathcal{T}_L,\, t_1 \neq t_2} r_{j t_1 t_2}$$

Constraints for the subgroup membership:

$$\Big(\sum_{t' \in \mathcal{T}_L} z_{it'} \boldsymbol{\gamma}_{t'}^\top \boldsymbol{Z}_i - \boldsymbol{\gamma}_{t_0}^\top \boldsymbol{Z}_i\Big) z_{it_0} = 0, \quad \forall i \in [n],\ t_0 \in \mathcal{T}_L$$
$$\sum_{t' \in \mathcal{T}_L} z_{it'} = 1, \quad z_{it'} \leq l_{t'}, \quad \sum_{i=1}^{n} z_{it'} \geq N_{\min}\, l_{t'}$$
$$z_{it'},\ l_{t'} \in \{0,1\}, \quad \forall i \in [n],\ t' \in \mathcal{T}_L$$

Constraints for Tree structure:

$$\boldsymbol{a}_m^\top \boldsymbol{X}_i \geq b_m - (1 - z_{it'}), \quad m \in \mathcal{R}(t')$$
$$\boldsymbol{a}_m^\top \boldsymbol{X}_i < b_m + (1 - z_{it'}), \quad m \in \mathcal{L}(t')$$
$$\textstyle\sum_{k=1}^{d} a_{t'k} = d_{t'}, \quad 0 \leq b_{t'} \leq d_{t'}, \quad d_{t'} \leq d_{p(t')}$$
$$a_{t'k},\ d_{t'} \in \{0,1\}, \quad \forall k \in [d],\ t' \in \mathcal{T}_B$$

Fusion constraints for information sharing:

$$(\gamma_{t_1 j} - \gamma_{t_2 j})\,(1 - r_{j t_1 t_2}) = 0, \quad \forall t_1 \neq t_2 \in \mathcal{T}_L$$
$$r_{j t_1 t_2} \in \{0,1\}, \quad \forall j \in [2d+2]$$

Solved with **Gurobi**; λ tuned by **BIC**.

Theoretical Guarantees

Let f^* be a piecewise-linear function induced by a depth-K tree with coefficient ℓ_1-norm bounded by B, and let $\hat{f}(\mathcal{T}_K^{\mathrm{OT}})$ be the FOCT estimator.

Theorem 4.2 (Optimal Tree)

$$\mathbb{E}_{\mathcal{D}_n} \Big\| \hat{f}(\mathcal{T}_K^{\mathrm{OT}}) - f^* \Big\|_2^2 \leq C_1 \frac{2^K \log^2(N)\big((p+1)\log(N) + \log(p)\big)}{N}.$$

Theorem 4.3 (Greedy CART)

$$\mathbb{E}_{\mathcal{D}_n} \Big\| \hat{f}(\mathcal{T}_K^{\mathrm{CART}}) - f^* \Big\|_2^2 \leq \frac{C_0\, h}{K + 2h + 1} + (\text{same rate term}).$$

FOCT achieves near-minimax convergence at *finite* depth K; CART carries an extra term $C_0 h/(K{+}2h{+}1)$ that vanishes only as $K \to \infty$, exacerbating overfitting.

Simulation Setup

Data-generating process with $M^* = 4$ true subgroups defined by signs of X_1, X_2:

$$Y = \sum_{m=1}^{4} \mathbf{1}_{(\boldsymbol{X} \in \mathcal{A}_m^*)} \big(\delta_m + \mu_m T + \boldsymbol{\alpha}_m^\top \boldsymbol{X} + \boldsymbol{\beta}_m^\top T \boldsymbol{X}\big) + \epsilon.$$

A distractor X_3 is correlated with the true split variable X_2 by $\rho \in \{0.7, 0.8\}$. Tree depth $= 2$, $N_{\min} = 1$, $n_{\text{train}} = 200$, $n_{\text{test}} = 2000$, 100 Monte Carlo runs.

Methods compared

- **CT-CART:** greedy causal tree.
- **OCT:** optimal causal tree via MIO, $\lambda = 0$ (no fusion).
- **FOCT:** optimal causal tree with fusion, λ tuned by BIC.

Subgroup Identification Accuracy

Number of correctly recovered tree structures (out of 100):

Method	$\rho = 0.7$	$\rho = 0.8$
CT-CART	51 / 100	39 / 100
OCT	78 / 100	54 / 100
FOCT	**81 / 100**	**67 / 100**

Table 1. FOCT recovers the true partition more often than OCT and CT-CART; the gap widens with stronger spurious correlation ($\rho = 0.8$).

On the held-out test set, FOCT also yields the lowest MSE for SATE (subgroup average treatment effect) estimation and the lowest out-of-sample risk; the advantage over OCT grows when overfitting pressure increases (small n, large ρ, or extra noise covariates—see Appendix C).

Case Study: HABS-HD (Black Participants)

Data. 237 Black participants from HABS-HD; covariates include Sex, APOE2, APOE4, tau in entorhinal cortex (Tau_1) and neocortex (Tau_2), Age, Education. Treatment: A$\beta \geq 10$ vs. A$\beta < 10$. Outcome: MMSE.

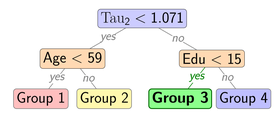

Figure 3. FOCT subgroups: split on Tau_2, then on Age (low-tau branch) and Education (high-tau branch). **Group 3** (high tau, low education) is highlighted as the only subgroup with significant treatment benefit.

Key clinical finding

Only **Group 3** (low education, high neocortical tau) shows a statistically significant benefit from Aβ lowering on MMSE. CT-CART and OCT, lacking fusion, yield wider confidence intervals and report *implausible* APOE4/APOE2 interactions that contradict the AD literature.

Takeaways

- **Global optimization** + **selective fusion** = better subgroup analysis.
- FOCT improves recovery over greedy and non-fused optimal trees, especially under small samples.
- Theory matches empirics: optimal trees avoid the greedy approximation term and are consistent at finite depth.
- On HABS-HD, FOCT yields a clinically interpretable subgroup and avoids spurious APOE interactions.