

Elastic Net Meets BART: A Hybrid Virtual Twins Approach for Reliable Subgroup Identification

Alexis Bernard Dalam

Principal Biostatistician
Pfizer, Inc.

ML15: Machine Learning, AI & Large Language Models, Data Handling & Data Engineering
Hyderabad, India
21 February 2026

Overview

Objectives

Virtual Twins Framework

ITE Estimation

Subgroup Identification

Simulated Data and Case Study

Results

Objectives

- ❖ Evaluate the performance of **multiple predictive models** within the **Virtual Twins** framework.
 - ❖ Compare models based on their ability to estimate **Individual Treatment Effects (ITE)** accurately.
 - ❖ Assess how these models help identify **treatment effect heterogeneity** in a simulated data and synthetic Ulcerative Colitis trial.
 - ❖ Determine the effectiveness of the VT framework in detecting **meaningful subgroups** based on **Mean Absolute Difference (MAD)**
-

Strategy for Addressing the Research Objectives

- **Model Performance (Step 1):**
 - Summarize **ITE estimates** for each model using **descriptive statistics**.
 - Compare models using **Mean Squared Error (MSE)** to assess accuracy and consistency.
 - **Subgroup Identification (Step 2):**
 - Apply **Classification and Regression Trees (CART)** to identify subgroups with **differential treatment effects**.
 - Characterize subgroups based on the **MAD** of the 4 models' response rate
 - **Framework Demonstration:**
 - Demonstrate how combining **EN** and **BART** improves understanding of population heterogeneity and variability in treatment effects, and facilitates the detection of meaningful subgroups.
-

Virtual Twins Framework

A two-step modeling framework to identify subgroups with varying treatment effects.

- **Step 1: ITE Estimation**
 - Predict potential outcomes under **treatment** and **control** to **estimate ITE** using a model.
 - **Step 2: Subgroup Identification**
 - Classify subjects using tree-based methods based on estimated ITEs.
 - Subjects with estimated **ITE $\geq$ Drug ATE** are labeled as **High-Responders (HR)**.
-

Step 1: ITE Estimation

1

Elastic Net (EN) Regression

- Regularized linear regression model
- Combines L1 (**Lasso**) and L2 (**Ridge**) penalties
- Effective for high-dimensional, low-sample data

2

Bayesian Additive Regression Trees (BART)

- Nonparametric Bayesian ensemble method
- Captures nonlinear relationships and interactions
- Provides crucial uncertainty quantification

+ Hybrid Approaches

3

Sequential method:

- EN for variable selection, then **BART** for modeling

4

Ensemble method:

- Averaging predictions from both **EN** and **BART**

Step 1: ITE Estimation

- ✓ Estimate the **outcomes under treatment** using a model
- ✓ Estimate the **outcomes under control** using a model

➤ **Estimated ITE =**
Estimated Outcome (**Treatment**)
– Estimated Outcome (**Control**)

Formula:

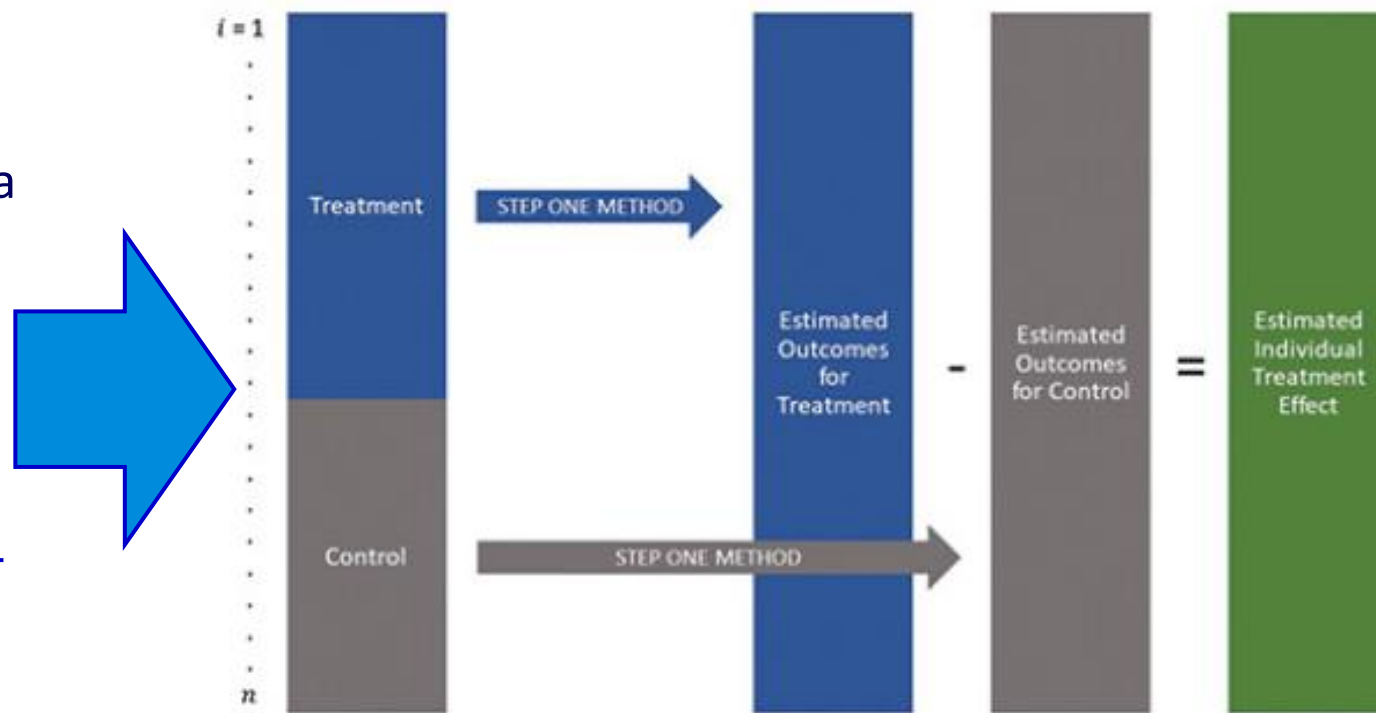
$$\hat{\tau}_i = \hat{Y}_i(1) - \hat{Y}_i(0)$$

where:

$\hat{\tau}_i$ = estimated ITE for individual i

$\hat{Y}_i(1)$ = predicted outcome for individual i if **treated**

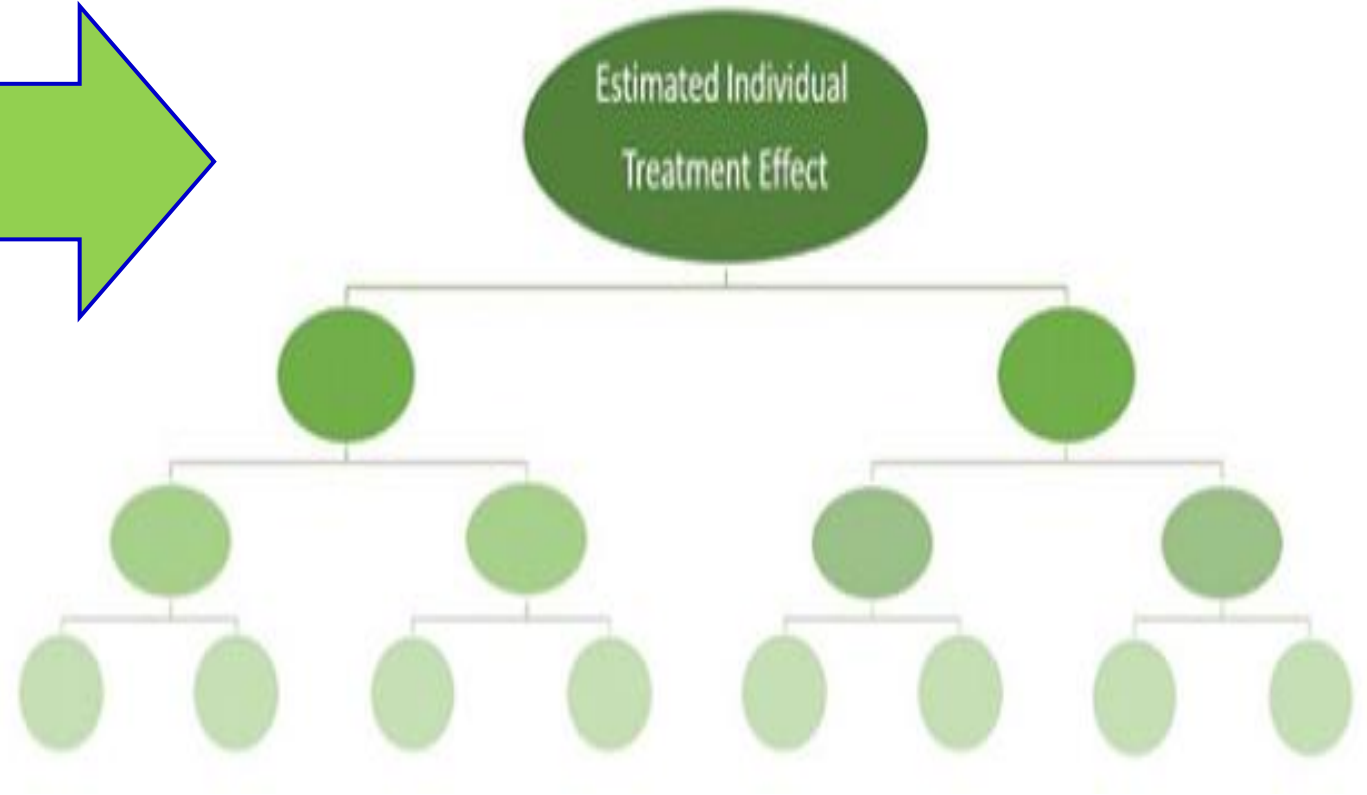
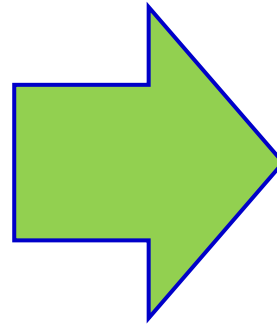
$\hat{Y}_i(0)$ = predicted outcome for individual i if **not treated**



Step 2: Subgroup Identification

Classification and Regression Trees (CART)

- A simple tree-based method that splits data into subgroups.
 - Splits are chosen to maximize differences in treatment effects between branches.
 - Helps identify meaningful subgroups based on MAD.
-



Simulated Data

- **Design:**

- N = 500; randomized 1:1 Drug vs Placebo (Treatment proportion ~ 0.50)
- Outcome : binary response generated from arm-specific probabilities

- **Covariates:**

- 12 **baseline features**: 6 continuous (standardized) + 6 binary
- **Signal** injected in **Cont1-Cont3** and **Bin1-Bin3** via logistic coefficients
Active betas (logit): **Cont** [0.5, 0.4, 0.3]; **Bin** [0.6, 0.5, 0.4]

- **Individual Treatment Effect (ITE):**

- ITE on probability scale $\sim \text{Normal}(\text{mean} = 20.45, \text{sd} = 7.39)$
- **Observed ITE summary** in this run: mean 21.18, sd 7.18

- **Observed Outcomes:**

- **Placebo-adjusted Response rate** 15.8
 - Response rate (**Drug**) 25.5; (**Placebo**) 6.02
-

Step 1: ITE Estimation Results (Simulated Data)

Simulated Data				
Statistics	Elastic Net	BART	Sequential	Ensemble
Mean (SD)	19.7 (9.88)	20.1 (11.2)	20.0 (12.6)	19.9 (9.87)
Median	19.9	17.5	18.0	18.8
Min, Max	-34.5, 50.3	-17.5, 63.9	-21.9, 68.5	-25.7, 54.1
Q1, Q3	14.6, 25.3	12.3, 26.7	11.3, 26.8	14.1, 25.2

Summary:

ATE (mean of predicted ITE's): ~ 20

Drug ATE (simulated data): 21.18

✓ Estimated ITE across the four models is ~ 20

Step 1: ITE Estimation Results (Simulated Data)

ITE Mean Square Error (MSE)

$$MSE_{ITE} = \frac{1}{n} \sum_{i=1}^n (\hat{\tau}_i - \tau_i)^2$$

where:

n = number of individual,

$\hat{\tau}_i$ = the *estimated ITE* for individual i , and

τ_i = the *true ITE* for individual i

Simulated Data:

- The **MSE** is comparable across all models, ranging from 0.1 to 0.7



ITE MSE (Simulated data)					
PREDICTED ITE	TRUE ITE				Simulated Data
	Elastic Net	BART	Sequential	Ensemble	
Elastic Net →	0	0.6	0.7	0.1	1.5
BART →	0.6	0	0.3	0.1	1.7
Sequential →	0.7	0.3	0	0.3	2.0
Ensemble →	0.1	0.1	0.3	0	1.4

Step 2: Subgroup Selected

Simulated Data						
	Subgroups Identified*	Elastic Net HRR (ATE)	BART HRR (ATE)	Sequential HRR (ATE)	Ensembles HRR (ATE)	Population % MAD
1	CONT1 < 1.074 & CONT2 >= 0.2893 & BIN1 = 0 & CONT1 >= -0.6858	85% (25.1)	84% (28.3)	80% (28.6)	88% (26.7)	16.4 3%
2	CONT1 >= 0.1305 & BIN1 = 0 & BIN5 = 1	94% (29.2)	74% (28.9)	75% (29.0)	87% (29.1)	20.6 11%
3	CONT1 < 0.1305 & CONT2 < 0.2263 & CONT3 < 0.251	12% (14.7)	1% (11.5)	7% (11.4)	2% (13.1)	22.4 5%

HRR = High-Responder's Rate (Estimated ITE >= Drug ATE)

ATE = Average Treatment Effect (Mean of estimated ITE)

MAD = Mean Absolute Difference between models' HRR

*Only 3 of the at least 12 subgroups identified by the 4 models are presented.

Case Study Overview

- Evaluates efficacy of investigational therapies for **Ulcerative Colitis (UC)** using synthetic data from two completed trials.
 - Combined dataset: **~900 participants** (2 treatment arms (**Drug 1** and **Drug 2**) over control)
-

- **Primary Endpoint: Modified Clinical Remission** (binary), defined by FDA for UC indication.
 - **Study Results - Average Treatment Effects (ATE)**
 - **Drug 1 ATE** = 14.6; **Drug 2 ATE** = 20.6
-
- **Covariates Considered:**
 - **Demographics:** Age at UC Diagnosis, Sex, Region, Race
 - **Disease Severity:** Baseline Modified Mayo Score, Baseline Total Mayo Score, Baseline Stool Frequency, Baseline Rectal Bleeding, Baseline Endoscopy Score
 - **Prior Medications:** Prior Biologic Treatment, Prior Anti-TNF Failure, Steroid Use at Baseline, Prior Anti-IL12/23
 - **Progressive Disease:** Baseline high-sensitivity C-reactive protein (CRP), Baseline Fecal Calprotectin

Step 1: ITE Estimation Results

Drug 1				
Statistics	Elastic Net	BART	Sequential	Ensemble
Mean (SD)	15.2 (11.2)	15.0 (12.3)	15.4 (13.4)	15.1 (11.3)
Median	14.4	13.5	14.1	14.2
Q1, Q3	7.7, 21.8	5.8, 22.2	5.9, 23.4	6.8, 22.3

Drug 1:

ATE (mean of predicted ITE's): ~15

Drug 1 ATE (case study): 14.6

✓ Estimated ITE across the four models is ~15

Drug 2				
Statistics	Elastic Net	BART	Sequential	Ensemble
Mean (SD)	20.3 (7.6)	19.5 (12.8)	19.7 (13.5)	19.9 (8.5)
Median	22.1	18.4	19.3	20.8
Q1, Q3	16.8, 26.1	10.8 , 28.5	10.6 , 29.3	14.7, 25.5

Drug 2:

ATE (mean of predicted ITE's): ~20

Drug 2 ATE (case study): 20.6

✓ Estimated ITE across the four models is ~20

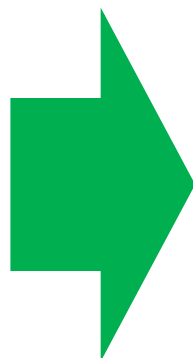
Step 1: ITE Estimation Results

Drug 1:

- The **MSE** is comparable across all models, ranging from 0.1 to 0.4.

Drug 2:

- Elastic Net exhibits a somewhat higher MSE between BART and Sequential as a true ITE relative to other models.



Drug 1 ITE MSE				
PREDICTED ITE	TRUE ITE			
	Elastic Net	BART	Sequential	Ensemble
Elastic Net →	0	0.4	0.4	0.1
BART →	0.4	0	0.3	0.1
Sequential →	0.4	0.3	0	0.2
Ensemble →	0.1	0.1	0.2	0

Drug 2 ITE MSE				
PREDICTED ITE	TRUE ITE			
	Elastic Net	BART	Sequential	Ensemble
Elastic Net →	0	1.6	1.6	0.4
BART →	1.6	0	0.2	0.4
Sequential →	1.6	0.2	0	0.5
Ensemble →	0.4	0.4	0.5	0

Step 2: Subgroup Selected

Drug 1						
	Subgroups Identified*	Elastic Net HRR (ATE)	BART HRR (ATE)	Sequential HRR (ATE)	Ensembles HRR (ATE)	Population % MAD
1	Prior Anti-TNF Failure = No & Baseline Total Mayo Score < 9.03 & Regions (EU, NA) & Steroid Use at Baseline = No	88% (24.1)	88% (24.9)	87% (25.8)	89% (24.5)	29.88 2%
2	Prior Anti-TNF Failure = No & Baseline Endoscopy Score = 2 & Baseline Fecal Calprotectin < 6040.3	89% (24.6)	76% (23.9)	74% (22.4)	84% (24.3)	32.97 11%
3	Prior Anti-TNF Failure = No & Baseline Total Mayo Score >= 9.03 & Sex = Male & Baseline CRP >= 1.198	15% (10.4)	6% (8.7)	14% (8.4)	6% (9.5)	15.88 6%

HRR = High-Responder's Rate (Estimated ITE >= Drug ATE)

ATE = Average Treatment Effect (Mean of estimated ITE)

MAD = Mean Absolute Difference between models' HRR

EU = Eastern Europe, NA = North America, CRP = C-Reactive Protein

*Only 3 of the at least 12 subgroups identified by the 4 models are presented.

Step 2: Subgroup Selected

Drug 2						
	Subgroups Identified*	Elastic Net HRR (ATE)	BART HRR (ATE)	Sequential HRR (ATE)	Ensembles HRR (ATE)	Population % MAD
1	Prior Biologic Treatment = No & Extent of Diseases (Proctitis, Left-sided Colitis) & Baseline Fecal Calprotectin < 7890.82 & Steroid Use at Baseline = No & Race = White	88% (23.9)	95% (33.9)	96% (34.8)	96% (28.9)	21.44 3%
2	Steroid Use at Baseline = No & Baseline Fecal Calprotectin < 10169.33 & Baseline Endoscopy Score = 3	98% (26.7)	53% (22.8)	55% (23.5)	78% (24.8)	33.22 17%
3	Steroid Use at Baseline = Yes & Baseline CRP < 41.996 & Race = White	2% (11.4)	24% (12.1)	22% (10.5)	11% (11.8)	27.63 17%

HRR = High-Responder's Rate (Estimated ITE $\geq$ Drug ATE)

ATE = Average Treatment Effect (Mean of estimated ITE)

MAD = Mean Absolute Difference between models' HRR

*Only 3 of the at least 12 subgroups identified by the 4 models are presented.

Conclusion and Future Work

- ❑ All four models yielded consistent and comparable ATE estimates for both Drug 1 and Drug 2. Similar observation for the simulated data.
 - ❑ The differences between the ITE estimates are minimal, as indicated by consistent MSE values.
 - ❑ Some subgroups exhibit consistent high-response rates across models.
 - ❑ Other subgroups show varying HRR but maintain comparable average treatment effects (ATE).
 - ❑ Some high-response subgroups represent at least 20% of the population, indicating substantial relevance.
-

➤ Future Work:

- Evaluate subgroups based on clinically meaningful criteria.
- Apply an alternative tree-based method for Step 2.

Thank You

Contact Information

Alexis Bernard Dalam, MSc

Email: AlexisBernard.Dalam@pfizer.com

LinkedIn: www.linkedin.com/in/alexisbernarddalam