Paper ML15

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

Alexis Bernard Dalam, Pfizer Inc., Makati City, Philippines

ABSTRACT

Small-sample trials in precision medicine often obscure treatment effect heterogeneity, making subgroup identification vital. This study explores integration of Elastic Net (EN) and Bayesian Additive Regression Trees (BART) within the Virtual Twins (VT) framework to improve individual treatment effect (ITE) estimation and subgroup discovery. EN, combining LASSO and Ridge penalties, offers robust variable selection in high-dimensional, low-sample settings, while BART provides flexible, nonparametric modeling with uncertainty quantification. We evaluate multiple modeling strategies: EN alone and sequentially/ensemble-combined with BART in Step1, followed by regression trees, conditional inference trees, or BART-based methods in Step2. Performance is assessed using simulation studies and real-world clinical application, across sample sizes ranging from 50 to 500, using metrics such as ITE mean squared error, subgroup classification accuracy and mean absolute difference. This evaluation provides practical guidance on modeling choices for VT, demonstrating that combining EN and BART enhances robustness and interpretability in small-sample subgroups.

INTRODUCTION

Subgroup identification has become a cornerstone of precision medicine, particularly in the context of randomized clinical trials (RCTs) where treatment effect heterogeneity (TEH) may be masked by limited statistical power or sample size. Traditional approaches to subgroup analysis often reliant on pre-specified hypotheses or post hoc interaction testing and have been criticized for their susceptibility to false discoveries and lack of reproducibility (Foster et al., 2011). As a result, there has been a growing interest in data-driven, model-based frameworks that can reliably uncover subgroups with enhanced treatment effects.

The Virtual Twins (VT) method, introduced by Foster et al. (2011), offers a principled two-step approach to subgroup identification. In Step 1, individual treatment effects (ITE) are estimated using predictive models; in Step 2, these estimates are used to identify subgroups via tree-based classification. VT has demonstrated superior performance over traditional logistic regression approaches, especially in scenarios where true subgroups exist but are obscured by noise or small sample sizes. However, the effectiveness of VT is highly contingent on the modeling choices made in each step.

Recent work by Deng et al. (2023) provides comprehensive guidance on VT modeling strategies, emphasizing the importance of flexible, high-performing models for ITE estimation. Their simulations show that ensemble methods like superlearner and nonparametric models such as Bayesian Additive Regression Trees (BART) outperform simpler linear models in capturing complex response surfaces. Moreover, the choice of subgroup identification method, whether regression trees, conditional inference trees, or BART-based classifiers, can significantly influence the interpretability and robustness of the results.

Complementing this, Zhang and Zhang (2022) propose a doubly robust classification framework for subgroup identification, integrating Elastic Net (EN) regularization with backward elimination to enhance variable selection. Their approach targets prescriptive variables, those that interact qualitatively with treatment and demonstrates improved performance in high-dimensional, low-sample settings. This is particularly relevant for rare disease trials, where the number of covariates often exceeds the sample size.

Together, these studies underscore the need for hybrid approaches that balance predictive accuracy, variable selection, and interpretability. By combining EN and BART within the VT framework, our study aims to leverage the strengths of both methods: EN's stability in variable selection and BART's flexibility in modeling nonlinear interactions. By leveraging both methods, our hybrid strategy aims to improve the reliability of subgroup identification in small-sample clinical trials, offering a practical and scalable solution for precision medicine.

OBJECTIVES

This study aims to:

1. Evaluate the performance of EN and BART, individually and in combination, for ITE estimation in the VT framework.
2. Compare subgroup identification strategies using regression trees and BART-based methods.
3. Assess model performance across simulated clinical datasets using:
    a. ITE mean squared error (MSE)
    b. Subgroup classification accuracy
    c. Mean Absolute Difference (MAD)

METHODS AND STATISTICAL DESIGN

The VT framework is a two-step modeling strategy designed to identify subgroups with differential treatment effects in randomized clinical trials (RCTs).

Step 1: Individual Treatment Effect (ITE) Estimation

In this step, predictive models are trained to estimate potential outcomes under both treatment and control conditions for each subject. The difference between these two predicted outcomes yields the individual treatment effect (ITE).

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

where

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

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

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

This step is crucial for quantifying how each subject responds to treatment relative to control.

Step 2: Subgroup Identification

Using the estimated ITEs, subjects are classified into subgroups based on their predicted treatment benefit. Classification is typically performed using tree-based methods (e.g., CART), which partition the covariate space to identify regions where treatment effects are significantly different. The threshold used for classification is the mean of the predicted ITEs, which serves as a proxy for the average treatment effect (ATE). Subjects with ITEs greater than or equal to the mean are labeled as Z = 1 (above-average responders), while others are labeled as Z = 0.

This framework allows for flexible modeling of treatment effects and facilitates the discovery of clinically meaningful subgroups.

MODELING STRATEGIES

To estimate ITEs, we explore four modeling configurations, each offering distinct advantages in terms of flexibility, interpretability, and robustness:

Elastic Net (EN) Regression

EN is a regularized linear regression technique that combines L1 (Lasso) and L2 (Ridge) penalties. It is particularly effective in high-dimensional settings where the number of covariates may exceed the sample size, a common scenario in rare disease trials. EN is used to estimate potential outcomes under treatment and control. EN hyperparameters (α and λ) are optimized via cross-validation and it performs automatic variable selection and handles multicollinearity. It assumes linear relationships and may underperform when interactions or nonlinearities are present.

Bayesian Additive Regression Trees (BART)

BART is a nonparametric Bayesian ensemble method that models the response surface as a sum of regression trees. It automatically captures nonlinear relationships and interactions without requiring explicit specification. BART is used to estimate potential outcomes for each subject under both treatment arms. It provides posterior distributions for predictions, allowing for uncertainty quantification. It is highly flexible and robust to model misspecification.

Hybrid Approaches

To leverage the complementary strengths of EN and BART, we implement two hybrid strategies:

- **Sequential Approach**
  This method first applies EN for variable selection, reducing the dimensionality of the covariate space. The selected variables are then used as inputs to BART for flexible outcome modeling. This approach aims to

improve computational efficiency and reduce overfitting while retaining BART's ability to model complex relationships.

- **Ensemble Approach**
  In this strategy, both EN and BART are independently used to estimate potential outcomes. The final ITEs are computed as the average of the predictions from both models. This ensemble method balances bias and variance, potentially improving generalization by smoothing out model-specific errors.

Subgroup Identification Strategy for Step 2

Subgroup identification is performed using Classification and Regression Trees (CART), a widely used tree-based method that partitions the covariate space based on splits that maximize treatment effect separation.

- Thresholding Rule:
- Subjects are classified based on whether their predicted ITE is greater than or equal to the mean ITE. This threshold serves as a proxy for the ATE.
  - Z = 1: ITE ≥ mean(ITE) → Above ATE
  - Z = 0: ITE < mean(ITE) → Below ATE

This binary classification enables the identification of subgroups with enhanced treatment effects, which can be further analyzed for clinical relevance and interpretability.

SIMULATION DESIGN

Simulated datasets are generated to mimic clinical trial scenarios with a sample size n = 500. Treatment effects are heterogeneous and designed to reflect realistic subgroup structures. Covariates include 6 continuous and 6 binary variables, with both informative and noise features. Here is the full mathematical formulation.

Let $i = 1, \ldots, n$ index subjects.

Treatment assignment:

$$T_i \sim Bernoulli(0.5)$$

Continuous covariates:

$$X_{ij}^{(c)} \sim N(\mu_j, 1), \quad \boldsymbol{\mu} = (1, \ldots, 6)$$

then standardized

$$Z_{ij} = \frac{X_{ij}^{(c)} - \bar{X}_j}{1}$$

Binary covariates:

$$X_{ij}^{(b)} \sim Bernoulli(p_j), \quad \boldsymbol{p} = (0.2, 0.3, 0.4, 0.5, 0.6, 0.7)$$

Baseline (placebo) linear predictor:

$$\eta_i = -2 + \sum_{j=1}^{6} Z_{ij}\beta_j^{(c)} + \sum_{j=1}^{6} X_{ij}^{(b)}\beta_j^{(b)}$$

$$\text{where } \boldsymbol{\beta}^{(c)} = (0.5, 0.4, 0.3, 0, 0, 0), \ \boldsymbol{\beta}^{(b)} = (0.6, 0.5, 0.4, 0, 0, 0)$$

Drug probability:

$$p_i^{(drug)} = \sigma(\eta_i) = \frac{1}{1 + e^{-\eta_i}}$$

Individual treatment effect:

$$ITE_i \sim N(0.2045, 0.0739^2)$$

Placebo probability:

$$p_i^{(placebo)} = p_i^{(drug)} - ITE_i$$

Observed probability:

$$p_i^{(observed)} = \begin{cases} p_i^{(drug)}, & T_i = 1 \\ p_i^{(placebo)}, & T_i = 0 \end{cases}$$

with bounding:

$$p_i^{(observed)} \leftarrow \min(\max(p_i, 0), 1)$$

Final observed binary response:

$$Y_i \sim Bernoulli\left(p_i^{(observed)}\right)$$

CASE STUDY

A synthetic data were generated based on selected completed clinical trials evaluating the efficacy of investigational therapies for ulcerative colitis (UC). The combined dataset includes approximately 900 participants, with two active treatment arms (Drug 1 and Drug 2) over placebo.

Estimands

Estimands are the proportion of participants who achieved the primary endpoint at the end of the induction period and the risk difference between treatment and placebo. The primary endpoint assessed is the modified clinical remission at the end of the induction period. The US Food and Drug Administration defined this binary endpoint (U.S. Food and Drug Administration, 2016) as an event meeting the following criteria:

- Endoscopic subscore of 0 or 1,
- Stool frequency subscore of 0 or 1, with a ≥1-point decrease from baseline,
- Rectal bleeding subscore of 0.

This endpoint reflects both symptomatic and endoscopic improvement, providing a clinically meaningful measure of treatment success.

Across two synthetic studies designed to reflect ulcerative colitis clinical trials, the placebo-adjusted response rates were 14.6% in Study 1 (Drug 1) and 20.6% in Study 2 (Drug 2). In both cases, the placebo response rate was roughly 10%, consistent with typical values reported in UC clinical research.

Covariates

The proposed framework applies a consistent set of baseline characteristics derived from both UC studies (Table 1). These include demographic factors, clinical severity measures, and prior treatment history, offering a thorough profile of participants at baseline. These covariates are then used within the framework to identify key variables, ensuring that the analysis focuses on factors most influential to clinical outcomes.

Table 1: List of baseline characteristics

	Variable	Description / Values	Reference Group
Demographics	Age at UC diagnosis	Range: 16 to 78	
	Sex	Female, Male	
	Geographic Region	Asia, North America, Europe, Others	Others
	Race	White, Asian, Others	Others
Disease Severity	Baseline Modified Mayo Score	Score: 3 to 10 (↑ disease severity)	
	Baseline Total Mayo Score	Score: 4 to 12 (↑ disease severity)	
	Baseline Stool Frequency	Score: 0 to 3 (↑ disease severity)	
	Baseline Rectal Bleeding	Score: 0 to 3 (↑ disease severity)	

	Baseline Endoscopy Score	Score: 0 to 3 (↑ disease severity)	
Prior Meds	Prior Biologic Treatment	Yes, No	No
	Prior Anti Integrin	Yes, No	No
	Prior Anti-TNF Failure	Yes, No	No
	Steroid Use at Baseline	Yes, No	No
	Prior Anti-IL12/23	Naive, Experienced	No
PD	Baseline high-sensitivity C-reactive protein	Range: 0.095 to 149.05	
	Baseline Fecal Calprotectin	Range: 3.8 to 46189.78	

EVALUATION METRICS

Performance is assessed using:

- **ITE Mean Squared Error** (MSE): Measures accuracy of treatment effect estimation.

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

where

$$n = \text{number of individual}$$

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

$$\tau_i = \text{the true ITE for individual } i$$

- **Subgroup Classification Accuracy**: Evaluates how well subgroups are identified by measuring the proportion of predictions that the model correctly assigns to each subgroup.
- **Mean Absolute Difference (MAD)**: Measures the consistency of subgroup predictions across the four models. A smaller value indicates greater agreement among the models.

RESULTS ON SIMULATED DATA

The results show that the estimated ITE across the four models (Elastic Net, BART, Sequential, and Ensemble) is approximately 20.0 (Table 2). This aligns with the simulated ITE of 21.18. The standard deviations are also comparable across all models.

The MSE values for Elastic Net, BART, and Sequential are similarly close (Table 3), indicating that all models perform well.

Out of the at least 12 subgroups identified by the four models, we present here three subgroups selected based on having the smallest to the highest MAD values while also having relatively large population sizes (Table 4).

Subgroup 1 has one of the lowest MADs. It includes the covariates Cont1, Cont2, and Bin1. The response rate is consistently high across models (approximately 80–88%) and ATE ranges from 25.1 to 26.7. This subgroup represents about 17% of participants (Table 4).

Subgroup 2 has one of the highest MADs. It includes the covariates Cont1, Bin1, and Bin5. The response rate varies from 74–94% across models, with an ATE of around 28.9–29.2. This subgroup represents approximately 21% of participants (Table 4).

Subgroup 3 is likely to identify non-responders. It includes the covariates Cont1, Cont2, and Cont3. The response rates are low (1–12%), and the ATE is small, ranging from 11.4 to 14.7. This subgroup accounts for about 22.4% of participants and provides valuable insight into patients who are less likely to benefit from the drug (Table 4).

Overall, these subgroups highlight meaningful heterogeneity in treatment response.

Table 2: ITE Estimation Results (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

Table 3: ITE Mean Squared Error (Simulated Data)

	TRUE ITE				
PREDICTED ITE	Elastic Net	BART	Sequential	Ensemble	Observed Data
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

Table 4: Subgroup Selected (Simulated Data)

Subgroups Identified*	**Elastic Net** HRR (ATE)	**BART** HRR (ATE)	**Sequential** HRR (ATE)	**Ensembles** HRR (ATE)	**Population** %	**MAD**
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%
CONT1 >= 0.1305 & BIN1 = 0 & BIN5 = 1	94% (29.2)	74% (28.9)	75% (29.0)	87% (29.1)	20.6	11%
CONT1 < 0.1305 & CONT2 < 0.2263 & CONT3 < 0.251	12% (14.7)	1% (11.5)	7% (11.4)	2% (13.1)	22.4	5%

*Only 3 of the at least 12 subgroups identified by the 4 models are presented here. These subgroups were selected based on their MAD values and population size. HHR = High-responder's Rate (Estimated ITE >= Drug ATE). ATE = Average Treatment Effect (Mean of estimated ITE). MAD = Mean Absolute Difference between models' HRR.

RESULTS ON CASE STUDY

For Drug 1, the estimated ITE across the four models (Elastic Net, BART, Sequential, and Ensemble) is approximately 15% (Table 5), which aligns closely with the case study ATE of 14.6. For Drug 2, the estimated ITE is around 20.0, comparable to the case study ATE of 20.6. These results indicate that all four modeling approaches yield stable and consistent estimates on average.

For MSE, Drug 1 shows comparable values across all models, ranging from 0.1 to 0.4, suggesting high consistency and accuracy (Table 6). For Drug 2, however, there is a slight difference: Elastic Net exhibits a somewhat higher MSE relative to BART and Sequential when compared to the true ITEs. Overall, these findings confirm that the models perform well, exhibiting only minor variations.

For subgroup selection, we identified at least 20 for both Drug 1 and Drug 2. In this paper, we presented 3 subgroups for Drug 1 and 3 subgroups for Drug 2 chosen based on the smallest to highest MADs while also having relatively large population sizes (Table 7).

Drug 1 Subgroups

Subgroup 1 has one of the lowest MADs. It includes patients with no prior Anti-TNF failure, a baseline Total Mayo Score below 9.03, regions limited to Eastern Europe and North America, and no steroid use at baseline. The response rate is high (about 88% across models) and the ATE is around 25.0. This subgroup represents nearly 30% of the participants (Table 7).

Subgroup 2 has a moderately high MAD of around 11%. It includes patients with no prior Anti-TNF failure, a baseline endoscopy score of 2, and fecal calprotectin levels below 6,040.3 micrograms per gram. Response rates range from

74% to 89%. Although the High-Responder Rate (HRR) varies widely across models, the ATE remains consistent, between 22.0 and 24.0. This subgroup accounts for approximately 33% of participants (Table 7).

Subgroup 3 likely identifies non-responders. Response rates are as low as 6%, and the ATE is only 8.0 to 10.0. Although the subgroup is relatively small (around 16% of participants), it offers valuable insight into patients who are less likely to benefit from Drug 1 (Table 7).

Drug 2 Subgroups

Subgroup 1 shows high response rates (between 88% and 96% across models). The ATE ranges from 23 to 35 (Table 7), indicating substantial variability across modeling approaches.

Subgroup 2 has a higher MAD of around 17%. Response rates vary widely, from 53% to 98%, reflecting notable inconsistency across models. The ATE also spans a broad range, from 22 to 27 (Table 7).

Subgroup 3 likely identifies non-responders. Response rates range from 2% to 24%, demonstrating considerable variability. Despite this, the ATE remains relatively stable (between 10 and 12). This subgroup represents about 28% of participants (Table 7).

Overall, these subgroups highlight meaningful heterogeneity in treatment response across both Drug 1 and Drug 2.

Table 5: ITE Estimation Results (Case Study Data)

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 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

Table 6: ITE Mean Squared Error (Case Study Data)

Drug 1				
	TRUE ITE			
PREDICTED 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				
	TRUE ITE			
PREDICTED 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

Table 7: Subgroup Selected (Case Study Data)

Drug 1						
Subgroups Identified*	**Elastic Net** HRR (ATE)	**BART** HRR (ATE)	**Sequential** HRR (ATE)	**Ensembles** HRR (ATE)	**Population** %	**MAD**
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%
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%
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%

Drug 2						
Subgroups Identified*	**Elastic Net** HRR (ATE)	**BART** HRR (ATE)	**Sequential** HRR (ATE)	**Ensembles** HRR (ATE)	**Population** %	**MAD**
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%
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%
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%

***Only 6 of the at least 20 subgroups identified by the 4 models are presented here. These subgroups were selected based on their MAD values and population size. HHR = 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.

DISCUSSION

Across four modeling strategies (Elastic Net, BART, Sequential EN→BART, and EN–BART Ensemble), the results demonstrate strong consistency and robustness in estimating both average and individual treatment effects. All models produced highly comparable ATE estimates for Drug 1 and Drug 2 in both case study clinical trial and simulated datasets, indicating that the overall treatment effect signal is stable and not sensitive to modeling choice.

Differences in ITE estimates were minimal, as evidenced by closely aligned MSE values across models, confirming that each method effectively captures patient-level treatment variation. This reinforces the value of integrating Elastic Net's variable selection strengths with BART's nonlinear modeling flexibility.

In identifying subgroups, the models consistently detected high-response subgroups, some representing 20% or more of the population, underscoring their potential clinical significance. While certain subgroups showed variation in high-response rates between models, their average treatment effects remained similar, suggesting stable underlying treatment-effect heterogeneity.

Overall, the findings show that the hybrid EN–BART framework improves reliability in subgroup identification while maintaining consistent treatment effect estimates, helping uncover meaningful patient differences that can inform precision medicine strategies.

LIMITATIONS

While the hybrid VT framework demonstrated promising results in estimating ITEs and identifying dominant subgroups, several limitations remain:

- Limited Scope of Subgroup Classifiers:
  This study focused on Classification and Regression Trees (CART) for subgroup identification in Step 2. More advanced classifiers such as Conditional Inference Trees and BART-based classification models were not evaluated. These methods may offer improved interpretability and statistical rigor, especially in complex subgroup structures.

- Unassessed Variable Selection Quality:
  Although EN was used for variable selection in some configurations, the quality of selected variables and the rate of false discoveries were not formally assessed. Evaluating these metrics is essential to ensure the reliability and reproducibility of identified subgroups, particularly in high-dimensional settings.

- Clinically meaningful Subgroups:
  Subgroups identified where not assessed based on clinically meaningful criteria.

- Sample Size and Subgroup Complexity:
  Model performance varied depending on sample size and subgroup structure. While dominant subgroups were consistently identified, it remains necessary to assess whether smaller or more complex subgroups exhibit inconsistent classification accuracy across models.

CONCLUSION

This study shows that the hybrid EN–BART framework provides reliable and consistent estimates of both average and individual treatment effects, while also effectively identifying meaningful patient subgroups. The strong agreement across multiple modeling approaches reinforces the robustness of the results and highlights the potential of combining Elastic Net and BART for precision medicine applications.

However, several limitations should be noted. The subgroup classification relied on CART alone, the quality of variable selection was not formally evaluated, clinically meaningful subgroup criteria were not assessed, and model performance may vary for smaller or more complex subgroups.

FUTURE WORK

Future work should explore more advanced classifiers, assess the rigor of variable selection, and further evaluate subgroup interpretability and clinical relevance. It should focus on:

- Extending subgroup classifiers to include more advanced methods such as Conditional Inference Trees and BART-based classification models, which may improve interpretability and statistical rigor in subgroup identification.
- Evaluating subgroups identified based on clinically meaningful criteria.
- Exploring alternative simulation designs to test model robustness under diverse clinical scenarios. Suggested simulation types include:
  - Time-to-event (survival) outcomes, to assess performance in oncology or chronic disease trials
  - Longitudinal data simulations, where repeated measurements over time may influence treatment effects.
  - Multi-arm trial simulations, to evaluate subgroup identification across multiple treatment comparisons.
  - Missing data scenarios, to test model resilience under incomplete covariate or outcome information.
  - Non-randomized (observational) data simulations, to explore applicability in real-world evidence settings.

REFERENCES

1. Deng, C., Wolf, J. M., Vock, D. M., Carroll, D. M., Boatman, J. A., Hatsukami, D. K., Leng, N., & Koopmeiners, J. S. (2023). Practical guidance on modeling choices for the virtual twins method. Journal of Biopharmaceutical Statistics, 33(5), 653–676. https://doi.org/10.1080/10543406.2023.2170404
2. Foster, J. C., Taylor, J. M. G., & Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. Statistics in Medicine, 30(24), 2867–2880. https://doi.org/10.1002/sim.4322
3. ZHANG B, & ZHANG M. (2022). SUBGROUP IDENTIFICATION AND VARIABLE SELECTION FOR TREATMENT DECISION MAKING. The Annals of Applied Statistics Vol 16. No. 1. 40-59. https://doi.org/10.1214/21-AOAS1468

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Alexis Bernard A. Dalam
Company: Pfizer, Inc.
Address: 8 Rockwell Building, Rockwell Center, Makati City, Philippines
Work Phone: (+63) 9177166539
Email: AlexisBernard.Dalam@pfizer.com
Website: www.linkedin.com/in/alexisbernarddalam