

# Comparing Modern Machine Learning Methods for Predicting Individual Treatment Effect

Pamela Solano<sup>1</sup> and Thomas Jaki<sup>1,2</sup>

<sup>1</sup>Regensburg University, <sup>2</sup>Cambridge University

July 24, 2024

- **not all participants will react equally to an intervention**

- **not all participants will react equally to an intervention**
- Characterizing this **heterogeneity in intervention effects** is key to improving patients outcomes.

# Objective

In order **to estimate Individual Treatment Effects**.

To compare state-of-the-art **Statistical models and Machine Learning algorithms**.

# The predicted individual treatment effects (PITE) framework

- **PITE is a method to estimate individual treatment effects.**

challenge

PITE is unobserved

# The predicted individual treatment effects (**PITE**) framework

- PITE consists of the difference between **experimental(E)** and **control (C) prediction** for each individual [Jaki et al., 2024].

$$\text{PITE}_i = f_E(\mathbf{X}_i) - f_C(\mathbf{X}_i), \quad f(\cdot) \text{ is a predictor}$$

?

Which method is the best to estimate PITE

# Can any approach for prediction be applied?

**Predictive models** should address

- population heterogeneity,

# Can any approach for prediction be applied?

**Predictive models** should address

- population heterogeneity,
- complex structures (linear/non-linear)



# Can any approach for prediction be applied?

**Predictive models** should address

- population heterogeneity,
- complex structures (linear/non-linear)
- and high-dimensionality [Lamont et al., 2016].

# Can any approach for prediction be applied?

## important!

- The best method to estimate the PITE is not **necessarily** the best fit model.
- An essential part is **to delineate this heterogeneity** based on the baseline covariates (features).

# Regressions approaches applied

	Heterogeneity	Linear	Non_Linear	Overfitting	Regularization	Tree_Regression
The lasso (lasso)						
The Bayesian Lasso (blasso)						
Supervised Principal Component Analysis (superpc)						
Spike and Slab Regression (spikeslab)						
Ridge (ridge)						
Random Forest (rf)						
Quantile Random Forest (qrf)						
Projection Pursuit Regression (ppr)						
Principal Component Analysis (pcr)						
Polynomial Kernel Regularized Least Squares (krlsPoly)						
Penalized Linear Regression (penalized)						
Partial Least Square (pls)						
Non-Negative Least Squares (nnls)						
lm with leapBackward (leapBackward)						
lm with Forward (leapForward)						
lm Stepwise (leapSeq)						
Linear Regression with Stepwise Selection (lmStepAIC)						
Linear Regression (lm)						
GLM with Stepwise Feature Selection (glmStepAIC)						
Generalized Linear Model (glm)						
Elasticnet (enet)						
Cubist (cubist)						
CART (rpart2)						
Boosted Tree-Boost number (blackboost)						
Boosted Linear Model (BstLm)						
Boosted Generalized Linear Model (glmboost)						
Bayesian Ridge Regression Model Averaged (blassoAveraged)						
Bayesian Ridge Regression (bridge)						
Bayesian Regularized Neural Networks (brnn)						
Bayesian Generalized Linear Model (bayesglm)						
Bayesian Additive Regression Trees (bartMachine)						

- The risk (Expected Prediction Squared Error)

$$\frac{1}{n} \sum_{i=1}^{i=n} (\text{tPITE}_i - \text{PITE}_i)^2 \times 100$$

- Sensitivity (Detect PITE direction)

$$\text{Same Direction}_i = \begin{cases} 1 & \text{if } \text{tPITE}_i \times \text{PITE}_i > 0 \\ -1 & \text{otherwise} \end{cases}$$

true  $\text{tPITE}_i$  and estimated  $\text{PITE}_i$ .

# Data generation mechanism

Sample size  $n = 40, 70, 100, 300, 400, \mathbf{500}, \mathbf{1000}, 1200, 1500$  with allocation ratio 1:1 ( $n_C = n_T = n/2$ ).

$$y = \mathbf{X}\beta + t\mathbf{Z}\gamma + \epsilon, \quad \epsilon \sim N(0, 1)$$

$t \in (0, 1) \Rightarrow$  benefit  $\mathbf{Z}\gamma$

Split the population 70% for training and 30% for validation.

- $\mathbf{Z}$  Normal, Linear, High correlations (**up to 0.5**)

# Data generation mechanism

Sample size  $n = 40, 70, 100, 300, 400, \mathbf{500}, \mathbf{1000}, 1200, 1500$  with allocation ratio 1:1 ( $n_C = n_T = n/2$ ).

$$y = \mathbf{X}\beta + t\mathbf{Z}\gamma + \epsilon, \quad \epsilon \sim N(0, 1)$$

$t \in (0, 1) \Rightarrow$  benefit  $\mathbf{Z}\gamma$

Split the population 70% for training and 30% for validation.

- $\mathbf{Z}$  Normal, Linear, High correlations (**up to 0.5**)
- $\mathbf{Z}$  Normal, Linear, Low correlations (**up to 0.3**)

# Data generation mechanism

Sample size  $n = 40, 70, 100, 300, 400, \mathbf{500}, \mathbf{1000}, 1200, 1500$  with allocation ratio 1:1 ( $n_C = n_T = n/2$ ).

$$y = \mathbf{X}\beta + t\mathbf{Z}\gamma + \epsilon, \quad \epsilon \sim N(0, 1)$$

$t \in (0, 1) \Rightarrow$  benefit  $\mathbf{Z}\gamma$

Split the population 70% for training and 30% for validation.

- $\mathbf{Z}$  Normal, Linear, High correlations (**up to 0.5**)
- $\mathbf{Z}$  Normal, Linear, Low correlations (**up to 0.3**)
- $\mathbf{Z} \sim U[0.1, 0.5]$ , Non-linear. Benefit:  $Z_1\gamma_1/(Z_2 + \gamma_2)$

# Data generation mechanism

Sample size  $n = 40, 70, 100, 300, 400, \mathbf{500}, \mathbf{1000}, 1200, 1500$  with allocation ratio 1:1 ( $n_C = n_T = n/2$ ).

$$y = \mathbf{X}\beta + t\mathbf{Z}\gamma + \epsilon, \quad \epsilon \sim N(0, 1)$$

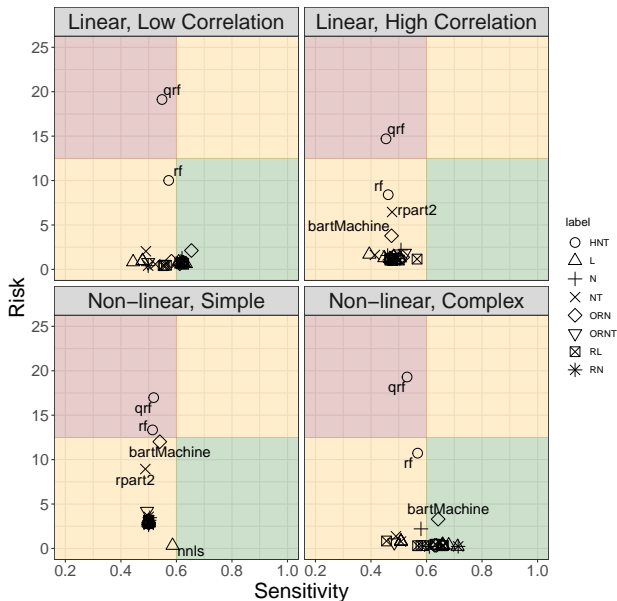
$t \in (0, 1) \Rightarrow$  benefit  $\mathbf{Z}\gamma$

Split the population 70% for training and 30% for validation.

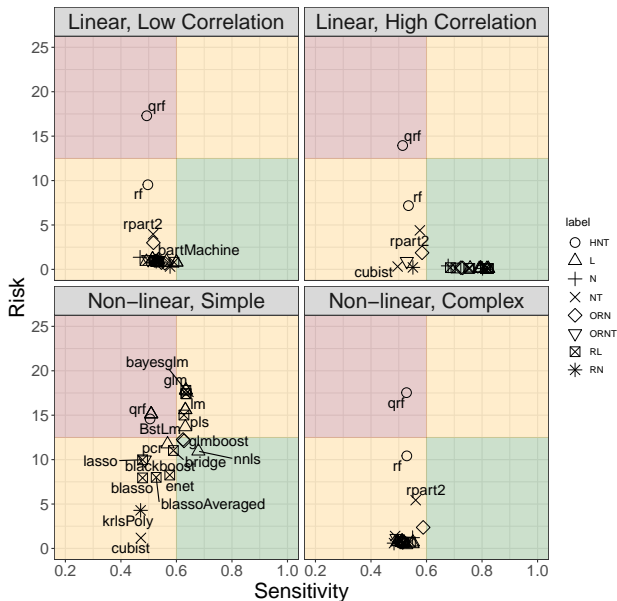
- $\mathbf{Z}$  Normal, Linear, High correlations (**up to 0.5**)
- $\mathbf{Z}$  Normal, Linear, Low correlations (**up to 0.3**)
- $\mathbf{Z} \sim U[0.1, 0.5]$ , Non-linear. Benefit:  $Z_1\gamma_1/(Z_2 + \gamma_2)$
- $\mathbf{Z} \sim U[0.1, 0.5]$ , Non-linear. Benefit:  $\frac{\log(Z_1)\gamma_0}{Z_1\gamma_2 - \gamma_3\sqrt{Z_2+2}}$



# Highlight based on risk and sensitivity $n = 500$

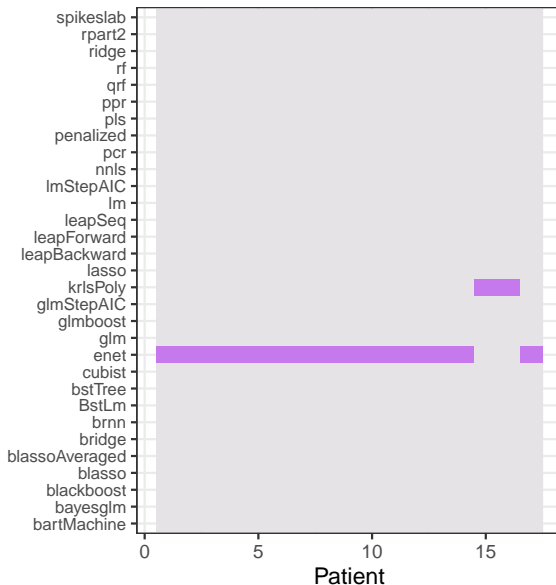


# Highlight based on risk and sensitivity $n = 1000$



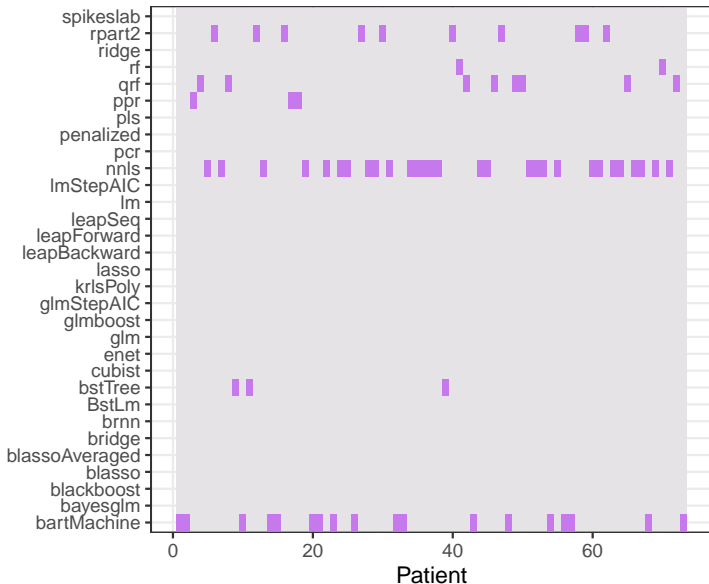
# Difficult cases Non-lin Complex $n = 500$

only **one model** captures the PITE direction



## Difficult cases Non-lin Complex $n = 1000$

only **one model** captures the PITE direction



- This work evidences potential approaches for different contexts.

# Conclusions

- This work evidences potential approaches for different contexts.
- An analyst should know what characteristics their dataset presents.

# Conclusions

- This work evidences potential approaches for different contexts.
- An analyst should know what characteristics their dataset presents.
- For each situation sensitivity varies little for different methods.

- This work evidences potential approaches for different contexts.
- An analyst should know what characteristics their dataset presents.
- For each situation sensitivity varies little for different methods.
- Risk is more variable.



# Conclusions

- This work evidences potential approaches for different contexts.
- An analyst should know what characteristics their dataset presents.
- For each situation sensitivity varies little for different methods.
- Risk is more variable.
- Some methods benefit more from smaller sample size.

- This work evidences potential approaches for different contexts.
- An analyst should know what characteristics their dataset presents.
- For each situation sensitivity varies little for different methods.
- Risk is more variable.
- Some methods benefit more from smaller sample size.
- Leave complex methods for large sample size.

- Thomas Jaki, Chi Chang, Alena Kuhlemeier, and M. Lee Van Horn.  
Predicting individual treatment effects: Challenges and opportunities for machine learning and artificial intelligence. *KI - Künstliche Intelligenz*, January 2024. ISSN 1610-1987. doi: 10.1007/s13218-023-00827-4. URL <http://dx.doi.org/10.1007/s13218-023-00827-4>.
- Andrea Lamont, Michael D Lyons, Thomas Jaki, Elizabeth Stuart, Daniel J Feaster, Kukatharmini Tharmaratnam, Daniel Oberski, Hemant Ishwaran, Dawn K Wilson, and M Lee Van Horn. Identification of predicted individual treatment effects in randomized clinical trials. *Statistical Methods in Medical Research*, 27(1):142–157, March 2016. doi: 10.1177/0962280215623981. URL <https://doi.org/10.1177/0962280215623981>.

# Comparing Modern Machine Learning Methods for Predicting Individual Treatment Effect

Pamela Solano<sup>1</sup> and Thomas Jaki<sup>1,2</sup>

<sup>1</sup>Regensburg University, <sup>2</sup>Cambridge University

July 24, 2024