Guidance on Individualized Treatment Rule Estimation in High Dimensions

Philippe Boileau[1], Ning Leng[2], Sandrine Dudoit[3]

[1]McGill University; [2]Genentech, Inc.; [3]University of California, Berkeley

Motivation

Precision medicine promises to tailor patients' treatments to optimize their outcomes. Patients' pre-treatment covariates, like age, sex-at-birth, and genetic profiles, influence therapies' efficacy, safety, and tolerability. These patient characteristics, known as **treatment effect modifiers, are used to derive individualized treatment rules (ITRs) that guide personalized treatment decisions**.

While numerous methods can successfully estimate ITRs in traditional asymptotic settings, **learning ITRs from high-dimensional data with more pre-treatment covariates than patients — a common occurrence in modern clinical data — is challenging**. ITR estimators developed for such data rely on simplifying assumptions. Recent advances in treatment effect modifier detection may help.

A comparison of ITR estimation procedures in high-dimensional settings has not been performed. Nor has an evaluation of these methods' sensitivities to assumption violations. As such, **selecting an appropriate ITR estimation strategy for high-dimensional settings is challenging for applied biomedical researchers**. Capacity for precision medicine is diminished as a result.

Primary Objectives

- **Provide guidance based on practical operating characteristics to applied scientists for ITR estimation in high dimensions.**
- **Determine whether treatment effect modifier detection procedures improve ITR estimation in high dimensions.**

Operating Characteristics

- **Rule Quality:** An ITR is "high-quality" when its expected outcome approaches that of the optimal rule. That is, the rule approximately optimizes mean outcome in the population.
- **Accurate Interpretability:** An ITR is accurately interpretable when it recovers treatment effect modifiers reliably in terms of the false discovery rate (FDR) and true positive rate (TPR).
- **Computational Efficiency:** An ITR is computationally efficient when it can be estimated quickly in serial with few computational resources.

Problem Formulation

Consider n i.i.d. observations $O = (W, A, Y)$ where W is a p-length random vector of pre-treatment covariates (and possible confounders) where $p \approx n$ or $p > n$, A is a binary treatment indicator, and Y is the continuous outcome.

We aim to estimate the ITR, defined as

$$I(\mathbb{E}[Y|W,1] - \mathbb{E}[Y|W,0] > 0) \ ,$$

where $\mathbb{E}[Y|W,1] - \mathbb{E}[Y|W,0]$ is the conditional average treatment effect (CATE) under standard identifiability conditions.

References

[1] Philippe Boileau, Ning Leng, Nima S Hejazi, Mark van der Laan, and Sandrine Dudoit. A nonparametric framework for treatment effect modifier discovery in high dimensions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(1):157–185, February 2025. ISSN 1369-7412. doi: 10.1093/jrsssb/qkae084.

[2] Robert Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58 (1):267–288, 1996. ISSN 0035-9246.

[3] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, August 2016. Association for Computing Machinery. ISBN 978-1-4503-4232-2. doi: 10.1145/2939672.2939785.

[4] Lu Tian, Ash A. Alizadeh, Andrew J. Gentles, and Robert Tibshirani. A Simple Method for Estimating Interactions Between a Treatment and a Large Number of Covariates. *Journal of the American Statistical Association*, 109(508):1517–1532, October 2014. ISSN 0162-1459. doi: 10.1080/01621459.2014.951443.

[5] Alexander R. Luedtke and Mark J. van der Laan. Super-Learning of an Optimal Dynamic Treatment Rule. *The International Journal of Biostatistics*, 12(1):305–332, May 2016. ISSN 1557-4679. doi: 10.1515/ijb-2015-0052.

[6] Mark J. van der Laan, Eric C. Polley, and Alan E. Hubbard. Super Learner. *Statistical Applications in Genetics and Molecular Biology*, 6 (1), September 2007. ISSN 1544-6115. doi: 10.2202/1544-6115.1309.

[7] Stefan Wager and Susan and Athey. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523):1228–1242, July 2018. ISSN 0162-1459. doi: 10.1080/01621459.2017.1319839.

Estimators

The CATE estimators in the table below were used to construct ITR estimators. Filtered versions of these CATE estimators relying on the treatment effect modifier variable importance parameter (TEM-VIP) methodology of Boileau et al. [1] were used to construct ITR estimators as well.

CATE Estimator	Details
Plug-In LASSO	A plug-in estimator using LASSO [2].
Plug-In XGBoost	A plug-in estimator using XGBoost [3].
Modified Covariates LASSO	A modified covariates estimator [4] using LASSO. The propensity score is estimated using logistic LASSO.
Modified Covariates XGBoost	A modified covariates estimator [4] using XGBoost. The propensity score is estimated using logistic LASSO.
Augmented Modified Covariates LASSO	An augmented modified covariates estimator [4] using LASSO. The propensity score is estimated using logistic LASSO.
Augmented Modified Covariates XGBoost	An augmented modified covariates estimator [4] using XGBoost. The propensity score is estimated using logistic LASSO.
AIPW-based LASSO	An AIPW-based estimator [5] using Super Learners [6] to estimate the expected conditional outcome and the propensity score. Differences in predicted pseudo-outcomes are modeled using LASSO.
AIPW-based Super Learner	An AIPW-based estimator [5] using Super Learners to estimate the expected conditional outcome and the propensity score. Differences in predicted pseudo-outcomes are modeled using a Super Learner.
Causal Random Forests	A causal random forest estimator [7] using cross-validation for hyperparameter selection.

Simulated Data-Generating Processes

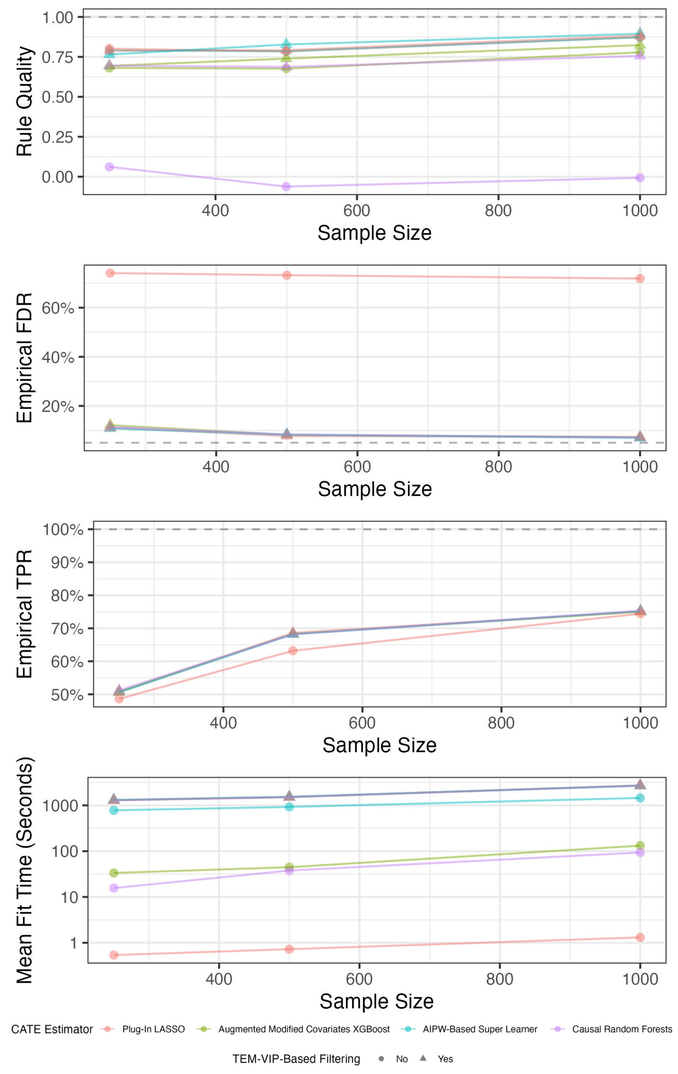
ITR estimators were benchmarked in 16 data-generating processes with continuous outcomes and binary treatment assignments **reflecting a diversity of randomized and observational studies**. Realizations of random vector O with $p = 500$ were generated according to the following data-generating process template:

$$\begin{aligned} W &\sim N(0, \Sigma) \\ A|W &\sim \text{Bernoulli}(\pi(W)) \\ Y|W, A &\sim N(\mu(W, A), 1) \ . \end{aligned}$$

Here, Σ is some 500×500 covariance matrix, $\pi(W) = \mathbb{P}[A = 1|W]$, and $\mu(W, A) = \mathbb{E}[Y|W, A]$. Data-generating processes are defined using combinations of the following factors:

$$\begin{aligned} \Sigma_1 &= I_{500\times 500} \\ \Sigma_2 &= \text{Block diagonal} \\ &\times \\ \pi_1(W) &= \frac{1}{2} \\ \pi_2(W) &= \text{logit}^{-1}\left(\frac{W_1 + W_2 + W_3 + W_4}{5}\right) \\ &\times \\ \mu_1(A, W) &= A + \gamma^\top W + (\delta^{(10)})^\top W A \\ \mu_2(A, W) &= A + \gamma^\top W + (\delta^{(50)})^\top W A \\ \mu_3(A, W) &= \gamma^\top W + 2 \arctan\left\{(\delta^{(10)})^\top W A\right\} \\ \mu_4(A, W) &= \gamma^\top W + 2 \arctan\left\{(\delta^{(50)})^\top W A\right\} \end{aligned}$$

Results Snapshot



Practical Guidance

No estimator uniformly dominates others across all operating characteristics; tradeoffs must be made. Since rule quality is generally of primary concern, practitioners must choose between accurate interpretability and computationally efficiency:

- **High-quality ITRs that are accurately interpretable:** The **TEM-VIP-filtered plug-in LASSO and AIPW-based estimators** generally produce high-quality ITR estimates while accurately recovering treatment effect modifiers. These estimators are computationally intensive, however. The computational burden might be lessened by parallelizing the estimation procedure.
- **High-quality ITRs that require few computational resources:** The **plug-in LASSO estimator** produces among the most high-quality rules in our simulation studies, providing empirical evidence that it is robust to model misspecification while being exceptionally computationally efficient. This estimator's built-in feature selection capabilities should not be used for treatment effect modifier discovery, however.

Funding

P.B. gratefully acknowledges the support of the Fonds de recherche du Québec - Nature et technologies and the Natural Sciences and Engineering Research Council of Canada.

Link to paper