Machine Learning for Estimating Individualized Treatment Effect from Real World Data for Use in Health Technology Assessment

Yingying Zhang[1] Noemi Kreif[1] Vijay Gc[2] Andrea Manca[1]

1. **Centre for Health Economics, University of York, York, UK 2. School of Human and Health Sciences, University of Huddersfield, Huddersfield, UK**

Contact email: yingying.zhang@york.ac.uk, andrea.manca@york.ac.uk

Introduction

ATE are at the heart of clinical and policy decision making, used to derive ICER and INB.

More nuanced decision-making accounting for heterogeneity in treatment effect may yield greater population health gains [1-3].

Clinicians and payers have focused more on considerations at the subgroup- and individual levels.

Patients and clinicians want to know what the outcomes of a treatment is for them, not for an average individual.

From ATE to ITE

The ITE for individual i with a vector of individual-specific predictors $X = x_i$ can be defined as:

$$ITE(x_i) = E[Y_i^{a=1}|X = x_i] - E[Y_i^{a=0}|X = x_i]$$

The $ATE(E[Y_i^{a=1}] - E[Y_i^{a=0}])$ is equal to the average of the $ITEs\left(E[Y_i^{a=1} - Y_i^{a=0}\right)$.

Identification Assumptions of ITE are the same as ATE, including **consistency**, **conditional exchangeability**, **positivity**, **no interference**.

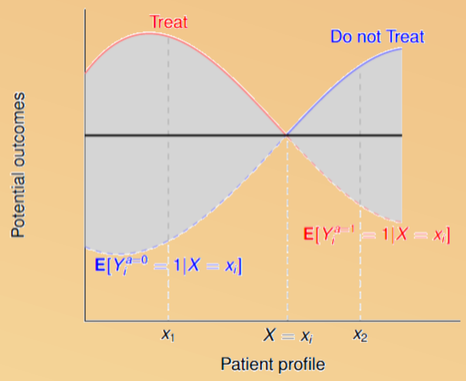


Figure 1: Optimal treatment strategy based on potential outcomes

Challenges in Estimating ITE

1. **What Data Is Required for ITE Estimation?**

ITE is essentially a highly conditional average treatment effect and can be realistically derived from large, well-designed, real-world studies.

2. **Why use ML to Estimate ITE?**

ML identify potential subgroups and select covariates (NICE real-world evidence framework June 2022). ML flexibly model complex interactions between treatment and high-dimensional individual characteristics. ML are not substitutes for content knowledge and clinicians' opinions.

3. **Outcomes**

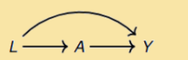
ML should focus on the potential outcomes instead of just the difference between them

4. **Uncertainty Quantification** makes ML more trustworthy and facilitate safer and more consistent treatment decisions.

5. **Parameters** focus on TTE outcome, baseline risk, related measures of treatment effect, HRQoL and costs.

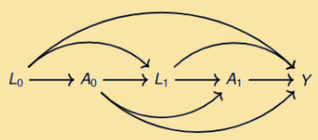
Risk of Bias in Causal Inference

- General to All Observational Studies
  1. Selection Bias
  2. Confounding
  3. Collider Bias
  4. Measurement Error



Baseline Confounding

Static setting



Time-varying confounding affected by prior treatment

Longitudinal setting

- Specific to Longitudinal Analysis
  1. Loss to Follow-Up
  2. Exposure Affected Time-varying Confounding
  3. Immortal Time Bias

Summarize ML Algorithms

We **extract data** based on:

- the available data (cross-sectional or longitudinal);
- the outcome of interest (continuous, binary or TTE);
- whether handle observed or unobserved confounders;
- whether quantify uncertainties of treatment effects or predicted outcomes;
- software implementation (R, Python or Stata).

ML Methods to Estimate ITE in Static Setting

Most ML methods:

- are designed for binary or continuous outcomes, require large samples;
- handle baseline confounding, assume no hidden confounding;
- not quantify uncertainty of both the predicted outcomes and treatment.

Table 1: Methods to Estimate ITE in Static Settings

Method	Handles confounding	Outcome variable	Quantifies uncertainty	Software	Reference
Bayesian Additive Regression Trees(BART), Bayesian Causal Forest	O	B, C	uncertainty intervals for predicted counterfactual outcomes and treatment effects	R: BART, bartCause, bcf	[33, 34, 40, 57]
Causal forest, Causal Multi-task Gaussian Processes(CMGP), Non-stationary Gaussian Processes(NSGP)	O, U	B, C	confidence intervals for potential outcomes and treatment effects	R: randomForestSRC, grf, BayesTree, causalForest	[36–38, 40, 42, 43, 58]
Virtual Twins Random Forests(VT), VT interaction, Counterfactual Random Forest (RF), counterfactual synthetic RF, Bivariate RF	O	B, C	No	aVirtualTwins, model4you	[35, 59, 60]
Balancing Neural Network(BNN)	O	B, C	No		[44]
Treatment-Agnostic Representation Network(TARNet)	O	B, C	No	Python: cfrnet	[45]
Local Similarity Preserved Individual Treatment Effect(SITE)	O	B, C	No	Python: SITE	[46]
Deep Counterfactual Networks with Propensity-Dropout(DCN-PD)	O	B, C	uncertainty in the estimated treatment effect	Python: DCN-PD	[41]
Multi-Task Deep Learning and K-Nearest Nighbours(MTDL-KNN)	O	B, C	No	Python: CNN	[47]
Generative Adversarial Nets for inference of Individualised Treatment Effects(GANITE)	O	B,C	only uncertainty in the counterfactual outcomes	Python: GAN-ITE	[48]
Person-Centered Treatment Effects Using a Local Instrumental Variables(PETIV)[11]	O, U	B, C	Yes	Stata: petiv	[15]
Counterfactual Survival Analysis(CSA)	O at baseline	TTE	uncertainty for the potential outcomes	Python only: Time_Series_Deconfounder	[12]
SurvITE	O at baseline	TTE	No	Python only: SurvITE	[11]
Cox Proportional Hasards Deep Neural Network(DeepSurv)	No	TTE	confidence intervals for predicted outcomes	Python only: DeepSurv	[61]
Non-Parametric Accelerated Failure Time(APT)	O at baseline	TTE	Yes	APTrees	[61, 62]
Non-Parametric Bayesian Additive Regression Trees within the framework of accelerated failure time mode(APT-BART-NP)	O at baseline	TTE	Yes	R and Python codes	[53]
Random Survival Forests(RSF)	O at baseline	TTE	uncertainty for treatment effects	No	[64, 65]
Causal Survival Forest (CSF)	O at baseline	TTE	uncertainty for treatment effects	grf	[66]
Deep Multi-task Gaussian Processes(DMGP)	O at baseline	TTE	No	Python only: DMGP	[42, 43, 67]

Legend: O: observed confounding; U: unobserved confounding; B: binary outcome; C: continuous outcome, TTE: time-to-event outcome.

ML Methods to Estimate ITE in Longitudinal Setting

In chronic conditions, treatments are sustained over time and we study a dynamic treatment regime.

Table 2: Methods to Estimate ITE in Longitudinal Settings

Method	Time-varying confounding	Baseline confounding	Outcome variable	Quantifies uncertainty	Software	Reference
Bayesian Non-parametric Method(BNP)	O	O	C	uncertainty for treatment effects	No	[30]
Bayesian Treatment Response Curves(BTRC)	No	No	C	No	No	[50]
Counterfactual Gaussian Process(CGP)	O	O	C	confidence intervals for predicted outcomes	No	[51]
Recurrent Marginal Structural Networks(RMSN)	O	O	B, C	No	Python only: RMSN	[52]
Counterfactual Recurrent Network(CRN)	O	O	B, C	No	Python only: CRN	[53]
Deep Sequential Weighting(DSW)	O, U	O, U	C	No	Python only: DSW	[56]
SyncTwin	O	O	C	No	synth_control	[54]
Time Series Deconfounder	O, U	O, U	B, C	No	Python only	[56]
Causal Dynamic Survival Model(CDS)	No	O	TTE	Yes	Python only: CDS	[68]

Legend: O: observed confounding; U: unobserved confounding; B: binary outcome; C: continuous outcome, TTE: time-to-event outcome.

ML Methods to Estimate ITE for TTE Outcomes

Survival model should account for potential bias from:

- non-randomised treatment assignment (confounding),
- informative censoring,
- event-induced covariate shift [17].

Modeling competing risks is another challenge.

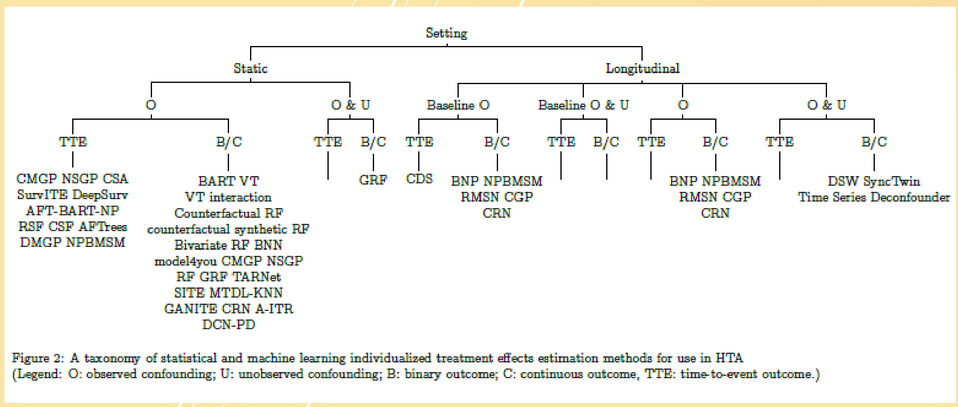


Figure 2: A taxonomy of statistical and machine learning individualized treatment effects estimation methods for use in HTA
(Legend: O: observed confounding; U: unobserved confounding; B: binary outcome; C: continuous outcome, TTE: time-to-event outcome.)

Conclusions and Discussions

1. Most ML for ITE estimation can handle **confounding at baseline** but **not time-varying or hidden confounding**.
2. ML accounting for time-varying confounding are developed mostly for use with **continuous or binary outcomes**.
3. Most ML methods do **not quantify uncertainty** of treatment effects estimates or predicted outcomes, especially in longitudinal settings.
4. Modeling **assumptions** should be properly assessed before making causal conclusions.
5. No ML can estimate **ITE for TTE outcomes AND account for time-varying confounders**.

Reference

[1] Douglas Coyle, Martin J Buxton, and Bernie J O'Brien. Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. Health Economics, 12(5):421–427, 2003.
[2] Mark Sculpher. Subgroups and heterogeneity in cost-effectiveness analysis. Pharmacoeconomics, 26(9): 799–806, 2008.
[3] Manuel A Espinoza, Andrea Manca, Karl Claxton, and Mark J Sculpher. The value of heterogeneity for cost-effectiveness subgroup analysis: conceptual framework and application. Medical Decision Making, 34(8): 951–964, 2014.
[4] Hugh A Chipman, Edward I George, and Robert E McCulloch. Bart: Bayesian additive regression trees. The Annals of Applied Statistics, 4(1):266–298, 2010.
[5] Jennifer L Hill. Bayesian nonparametric modeling for causal inference. Journal of Computational and Graphical Statistics, 20(1):217–240, 2011.
[6] Rodney Sparapani, Charles Spanbauer, and Robert McCulloch. Nonparametric machine learning and efficient computation with bayesian additive regression trees: the bart r package. Journal of Statistical Software, 97: 1–66, 2021.
[7] Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences, 113(27):7353–7360, 2016.
[8] Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523):1228–1242, 2018.
[9] Ahmed M Alaa and Mihaela van der Schaar. Bayesian inference of individualized treatment effects using multi-task gaussian processes. arXiv preprint arXiv:1704.02801, 2017.
[10] Ahmed Alaa and Mihaela Schaar. Limits of estimating heterogeneous treatment effects: Guidelines for practical algorithm design. In International Conference on Machine Learning, pages 129–138. PMLR, 2018.
[11] P Richard Hahn, Jared S Murray, and Carlos M Carvalho. Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion). Bayesian Analysis, 15 (3):965–1056, 2020.
[12] Heidi Seibold, Achim Zeileis, and Torsten Hothorn. Individual treatment effect prediction for amyotrophic lateral sclerosis patients. Statistical Methods in Medical Research,27(10):3104–3125, 2018.
[13] Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. The Annals of Statistics, 47(2): 1148–1178, 2019.
[14] Yanbo Xu, Yanxun Xu, and Suchi Saria. A bayesian nonparametric approach for estimating individualized treatment-response curves. In Machine Learning for Healthcare Conference, pages 282–300. PMLR, 2016.
[15] Hossein Soleimani, Adarsh Subbaswamy, and Suchi Saria. Treatment-response models for counterfactual reasoning with continuous-time, continuous-valued interventions. arXiv preprint arXiv:1704.02038, 2017.
[16] I Bica, A Alaa, and M Van Der Schaar. Time series deconfounder: Estimating treatment effects over time in the presence of hidden confounders. In International Conference on Machine Learning, pages 884–895. PMLR, 2020.
[17] Alicia Curth, Changhee Lee, and Mihaela van der Schaar. Survite: Learning heterogeneous treatment effects from time-to-event data. In Thirty-Fifth Conference on Neural Information Processing Systems, 2021.
[18] Paidamoyo Chapfuwa, Serge Assaad, Shuxi Zeng, Michael J Pencina, Lawrence Carin, and Ricardo Henao. Enabling counterfactual survival analysis with balanced representations. In Proceedings of the Conference on Health, Inference, and Learning, pages 133–145, 2021.
[19] Liangyuan Hu, Jiayi Ji, and Fan Li. Estimating heterogeneous survival treatment effect in observational data using machine learning. Statistics in Medicine, 2021.
[20] Jie Zhu and Blanca Gallego. Cds–causal inference with deep survival model and time-varying covariates. arXiv preprint arXiv:2101.10643, 2021.
[21] Nathan Kallus, Aahlad Manas Puli, and Uri Shalit. Removing hidden confounding by experimental grounding. Advances in neural information processing systems, 31, 2018.
[22] Susan Athey and Stefan Wager. Policy learning with observational data. Econometrica, 89(1):133–161, 2021.

Next Generation Health Technology Assessment

Centre For Health Economics