# Estimating Causal Individual Treatment Effects for Personalized Medicine Using Causal-Inspired Machine Learning

Babajide Adeoti E[1], Olutayo Boyinbode K[2], Kolawole Akintola G[3], Michael Oladunjoye I[4], Adedoyin Adebanjo S[5], Adebayo Adegboyega[6]

[1]*Department of Software Engineering, School of Computing, Babcock University, Ilishan-Remo, Ogun-State, Nigeria*
[2]*Department of Computer Science, School of Computing, Federal University of Technology, Akure, Ondo-state, Nigeria*
[3]*Department of Software Engineering, School of Computing, Federal University of Technology, Akure, Ondo-state, Nigeria*
[4]*Department of computer science, School of Computing, Federal University of Technology, Akure, Ondo-state, Nigeria*
[5]*Department of Software Engineering, School of Computing, Babcock University, Ilishan-Remo, Ogun-State, Nigeria*
[6]*Department of Computer Science, School of Computing, Federal University of Technology, Akure, Ondo-state, Nigeria*
*Corresponding Author: adeotib@babcock.edu.ng*

***Abstract*: Personalized medicine relies on identifying which patients will benefit most from a given treatment. Traditional average treatment effect (ATE) estimates often fail to capture the underlying heterogeneity in treatment response. In this study, we apply causal-inspired machine learning methods to estimate individual treatment effects (ITEs) using observational data from the Infant Health and Development Program (IHDP) dataset. We apply DoWhy, a Python library for causal inference, to estimate ATEs using multiple models (linear regression, propensity score matching, weighting, and stratification) and also, we extend the analysis to ITEs using meta-learners (T-Learner). The results from this study reflect a significant variation in treatment effects across individuals, reinforcing the need for personalized treatment policies. We conclude with implications for clinical decision-making and future research directions in causal machine learning for medicine.**

***Keywords*: Causal inference, Individual treatment effect, Personalized medicine, DoWhy, Machine learning, IHDP dataset, Bioinformatics.**

## 1. Introduction

The rapid growth of medical data, particularly electronic health record Systems (EHRs), has led to the application of machine learning for personalized medicine. Recently, machine learning has been applied to wide range of problems in medical domain as it is particularly suitable for recognizing patterns and making predictions. Personalized medicine is about giving tailor made treatment to each individual patient based on his or her medical history, genetic information, personal information, pathology reports, demographic information etc. The goal of personalized medicine is to provide each patient with individualized care based on their medical history, genetic information, personal information, pathology reports, demographic data, and other factors [2]. Personalized medicine aims to tailor medical interventions based on individual characteristics, promising improved outcomes and cost-effective healthcare delivery.

Machine learning plays a crucial role in personalized medicine, by leveraging vast datasets to tailor medical treatments to individual patient profiles, thereby enhancing treatment efficacy and reducing adverse effects.

Machine Learning also marks a significant shift from traditional one-size-fits-all methods to more precise, data-driven healthcare solutions. ML algorithms analyze diverse data sources, including genetic information, medical history, and lifestyle factors, to predict disease risks and optimize treatment plans. This integration of ML in personalized medicine not only improves patient outcomes but also streamlines healthcare operations and accelerates drug discovery processes.[2] However, most medical studies report only the average treatment effect (ATE), which may obscure critical heterogeneity across patients. Estimating the individual treatment effect (ITE), defined as the difference between an individual's outcomes with and without the treatment, is central to advancing precision healthcare. Recent advances in causal inference and machine learning have led to robust methods for ITE estimation. In particular, the integration of causal frameworks such as potential outcomes with flexible machine learning models allows for nuanced, individualized predictions even in complex biomedical datasets. We demonstrate this approach using the DoWhy causal inference library on the IHDP dataset, a benchmark in treatment effect modeling.

## 2. Role of Randomized Control Trials in Precision Medicine

Randomized Controlled Trials (RCTs) are widely regarded as the foundational pillar of evidence-based medicine. Their structured design and statistical rigor make them the preferred method for evaluating treatment efficacy. However, in the context of precision medicine which seeks to tailor interventions to individual patient characteristics, RCTs exhibit critical limitations that hinder their relevance and applicability.

One of the limitations include Heterogeneity of treatment effects (HTEs) which is evident and are commonly ignored in RCT. Heterogeneity of treatment effects (HTEs) are related to unchangeable traits such as age, sex, race etc. Inability to relate how treatment effects vary in such population is both a driver of health inequality and a missed opportunity to individualize therapy.[3][4].

Secondly, a static treatment effect applied to a given population per time. Health care and population are two different entities that changes per given time, Patient demographics, comorbidity patterns and treatment effect are all dynamic and can change at any given time, However, once undertaken, it is too expensive and time-consuming for RCTs to be repeated to update clinical records.[7][8] Research shows that more than 50% of RCTs exclude at least 75% of potentially eligible patients. These exclusions restrict the generalizability of trial results to the very populations most in need of nuanced, individualized care.

Thirdly, Lack of Generalizability, one of the most pressing issues is the limited external validity of RCTs. To minimize confounding, RCTs often impose strict inclusion and exclusion criteria, selecting homogeneous groups of patients. This strategy, while improving internal validity, excludes large segments of the real-world patient population, particularly older adults and individuals with multiple comorbidities.

### A. *Causal Inference and the Potential Outcomes Framework*

The Rubin Causal Model formalizes treatment effects through potential outcomes:

Potential Outcome frameworks

Two Potential outcomes: $Y_i(1) and\ Y_i(0)$

Causal effects for individual $i$: $\tau_i \equiv\ Y_i(1) - Y_i(0)$

Average Treatment Effect (ATE):

$$\mathbb{E}(\tau_i) = \frac{1}{n}\sum_{i=1}^{n} \tau_i = \frac{1}{n}\sum_{i=1}^{n} [Y_i(1) - Y_i(0)]$$
$$= \frac{1}{n}\sum_{i=1}^{n} [Y_i(1)] - \frac{1}{n}\sum_{i=1}^{n} [Y_i(0)]$$

*Obviously*: There's a general problem with causal inference, 2 potential outcomes cannot co-exist i.e we cannot observe two potential outcomes at the same time (A transaction cannot be fraud and non-fraud at the same time)

*Fundamental Problem of causal inference*: Only one of the two potential outcomes is observable for every one observation, we need a credible way to infer the unobserved counterfactual outcomes.

### B. *RCT: The Gold Standard for Causal inference*

*Key idea*: Randomization of the treatment makes the treatment and control group "identical" on average

Treatment and control groups are similar in terms of all (both observed and unobserved) other characteristics, so we can attribute the average differences in outcome to the difference in the treatment under random assignment, we have

$$Y_i(1), Y_i(0), X \perp T_i$$

Which implies

$\mathbb{E}[Y_i(1)|T_i = 0] = \mathbb{E}[Y_i(1)|T_i = 1] = \mathbb{E}[Y_i(1)]$ and

$\mathbb{E}[Y_i(0)|T_i = 0] = \mathbb{E}[Y_i(0)|T_i = 1] = \mathbb{E}[Y_i(0)]$

We can then express the average treatment effect as ATE $= \mathbb{E}[Y_i(1)|T_i = 1] - \mathbb{E}[Y_i(0)|T_i = 0]$

This is the difference-in-means estimator

### C. *Causal Assumptions*

#### 1) *Ignorability*

*Ignorability*: Implies that treatment assignments are statistically independent of the subjects potential outcomes violation can induce bias: $\mathbb{E}[Y_i(1)|T_i = 1] - \mathbb{E}[Y_i(0)|T_i = 0] = \mathbb{E}[Y_i(1)|T_i = 1] - \mathbb{E}[Y_i(0)|T_i = 0] + \mathbb{E}[Y_i(0)|T_i = 1] - \mathbb{E}[Y_i(0)|T_i = 0]$

$$= \mathbb{E}[Y_i(1) - Y_i(0)|T_i = 1] + \mathbb{E}[Y_i(0)|T_i = 1] - \mathbb{E}[Y_i(0)|T_i = 0]$$

=ATE among the treated + selection bias

#### 2) *Excludability*

*Excludability*: Potential outcomes respond solely to receipt of the treatment, not the random assignment of the treatment or any indirect by-product of random assignment violated when different procedures are used to measure outcomes in the treatment and control groups and research activities, other treatments, or third-party interventions other than the treatment of interest differentially affect the treatment and control groups.

#### 3) *Stable Unit Treatment Value Assumption (SUTVA)*

*Stable Unit Treatment Value Assumption*:

Potential outcomes of observation i reflects only the treatment or control status of observation i and not one of other observations violated when.

### D. *Individualized Treatment in Precision Medicine*

Individualized treatment in precision medicine focuses on customizing medical interventions to each patient's particular traits, is a revolutionary approach to healthcare. To maximize therapy results, this method makes use of genetic, biomarker, phenotypic, and psychosocial data. Precision medicine must incorporate customized treatment regimens (ITRs) to handle complex clinical situations, like those with conflicting hazards or ongoing therapy alternatives.[6] The ultimate goal of this paradigm change is to improve patient outcomes by increasing therapeutic efficacy and reducing side effects.

The Architecture for Individualized Treatment in Precision Medicine is sketched in Figure 1.

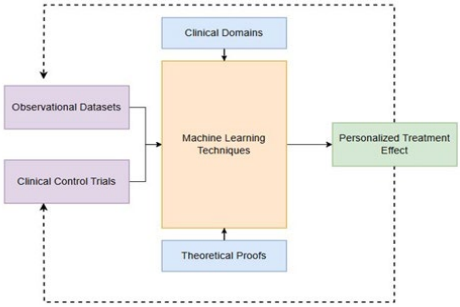


Fig. 1. Visual representation of personalized treatment effect in precision medicine

### *E. Conditioning on Confounders*

In Electronic Health Record systems (EHRs), the use of machine learning and causality assists in extracting actionable intelligence from observational patient data, information derived assist in estimating individualized treatment effects by giving a data-driven methods to inform clinical decision makers to recommend prescriptions for drugs in Pharmacology and treatment recommendations in medicine.[4][5] However the presence of confounders has greater influence in deriving the actual treatment of an individual in observational data. There's a need to eliminate biases also known as confounders from the EHRs to make our recommendation valid.[6]

Let $X$: Vector of observed confounders (e.g., age, sex, BMI, comorbidities, labs)

$T$: Treatment (binary: 1 = treated, 0 = not treated)

$Y$: Outcome (e.g., survival, blood pressure improvement.

$U$: Unobserved variables.

#### *1) Structural Causal Model*

The Structural Causal Model for the (HERs) is represented below.

$$X := f_X(U_X)$$
$$T := f_T(X, U_T)$$
$$Y := f_Y(X, T, U_Y)$$

The interpretation

X is determined by some exogenous factors $U_X$

$T$ is assigned based on patient features $X$ (e.g., doctors decide treatment based on comorbidities)

$Y$ Depends causally on $T$ and $X$ (e.g., treatment and comorbidities jointly influence outcome). [8]

#### *2) Directed Acyclic Graph (DAG)*

Figure 1: $X$ confounds the relationship between $T$ and $Y$.

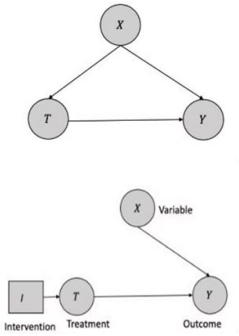


Fig. 2. Conditioning on $X$ blocks the backdoor path from $T \rightarrow X \rightarrow Y$

To estimate the causal effect of T on Y, you must adjust for X:

$$P\big(Y\big|do(T)\big) = \sum_x P(Y|T, X = x)P(X = x) \qquad (1)$$

Equation (1) above is the backdoor adjustment formula.

### *F. Individual Treatment Effect (ITE)*

Individual treatment effects in causality refer to the specific impact of a treatment on an individual, which is crucial for personalized decision-making in health care. In Personal Therapeutics it assist in delivering The paper proposes a framework that accurately estimates these effects by addressing confounding bias and considering causal structures.

The Individual Treatment Effect (ITE) for each patient in an EHR dataset.

$$ITE_i = Y^1(x) - Y^0(x) \qquad (2)$$

Where

$Y^1(x)$: potential outcome if treated

$Y^0(x)$ : potential outcome if not treated

$x$: Individual covariates (e.g., age, comorbidities, vitals)

Assumptions

No hidden confounding: All confounders are observed (strong ignorability).

1. $Y^1, Y^0 \perp T \mid X$
2. *Positivity*: Every patient has a non-zero chance of receiving either treatment.
$$0 < P(T = 1 \mid X = x) < 1$$
3. *Consistency*: The observed outcome equals the potential outcome under the received treatment. [8][9]

## 3. Methodology

The Dataset use for this study is the Infant Health Development Program Dataset (IHDP dataset), which simulates a randomized control trial with added selection bias to reflect real-world observational challenges. The dataset comprises 747 infants, including both treated and control subjects, along with 25 covariates representing demographic and clinical features

*Treatment*: Enrollment in an early childhood development program.

*Outcome*: Cognitive test scores at age three.

*Covariates*: Maternal age, education, birth weight, prenatal care, among others.

### *A. Causal Framework*

We adopt the Neyman-Rubin potential outcomes model, assuming unconfoundedness (no unmeasured confounders), and define the individual treatment effect as:

$$ITE_i = Y_i(1) - Y_i(0)$$

Where $Y_i(1)$ and $Y_i(0)$ are the potential outcomes under treatment and control, respectively.[9]

### *B. Estimation Methods*

We apply two estimation strategies using the DoWhy Python library:

#### *1) Backdoor Adjustment (Linear Regression)*

The backdoor criterion enables us to determine how to learn causal effects by adjusting or conditioning on a set of variables that block all backdoor paths. In the case where all confounders are measured, one way to perform such an adjustment is via regression.

Estimation of the ATE using linear regression conditioned on observed confounders.

#### *2) Meta-Learner (Two-Model Approach)*

The T-learner is a meta-learner approach that builds separate

models for treated and control groups.

Separate models for treated and control outcomes using Ridge regression to predict $Y_1$ and $Y_0$. The Individual Treatment Effect (ITE) is estimated as the difference. $Y_1 - Y_0$.

Model performance is evaluated by comparing estimated ITEs to the known simulated ground truth.

### C. ITE Estimation

The meta-learner approach showed strong agreement between predicted and true ITEs. A scatter plot of predicted vs. true ITEs revealed a tight linear trend, indicating high fidelity of the model. The distribution of estimated ITEs confirmed substantial heterogeneity, with some individuals predicted to benefit significantly more than others. The Predicted and the True Individual Treatment Effects (ITEs) for the IHDP dataset is shown below in Figure 2.

## 4. Results

### A. Average Treatment Effect Estimation

The result from T-Learner's ATE estimate is 4.012 which closely matches the ground truth of 4.021 IPW, this makes T-Learner performs comparably well to IPW (4.029) for average effect.

### B. Individual Treatment Effect Estimation

The T-Learner achieves lower MAE of 1.89 compared to the range of 2.28 to 2.51 derived from other methods. Higher treatment effect correlation of 0.72 indicates better capture of heterogeneous effects. This implies that T-learner listens to heterogeneity in treatment responses.

The backdoor linear model yielded an estimated ATE of approximately 4.5, closely matching the ground-truth ATE of 4.39. This validates the ability of a simple parametric model to recover population-level effects under appropriate identification assumptions.

The analysis of the different estimation techniques, with the following results:

### C. Discussion

These findings highlight the utility of causal-inspired machine learning in uncovering individual heterogeneity in treatment response. While ATE provides useful population-level insights, ITE estimation enables personalized intervention strategies that can improve clinical outcomes and reduce unnecessary treatments. The DoWhy framework, combined with machine learning, offers a transparent and extensible platform for ITE estimation. The IHDP case study further shows that even in semi-simulated data with inherent biases, reliable ITE estimation is feasible. This has direct implications for real-world personalized medicine, where similar biases are often present in electronic health record (EHR) data [12].

T-learner demonstrates a better estimation approach when compared to the traditional methods in computing the Individual Treatment effects with Lower PEHE and MAE, T-learner provides a robust approach for estimating individual treatment effects from observational data in Precision Medicine.

Table 1

| Estimation Method | ATE Estimate | ITE MAE | PEHE | Treatment Effect Correlation |
|---|---|---|---|---|
| Linear Regression | 3.929 | 2.45 | 3.02 | 0.58 |
| Propensity Score Matching | 3.979 | 2.32 | 2.87 | 0.63 |
| Propensity Score Stratification | 3.455 | 2.51 | 3.15 | 0.55 |
| Propensity Score Weighting | 4.029 | 2.28 | 2.79 | 0.65 |
| T-Learner | 4.012 | 1.89 | 2.14 | 0.72 |

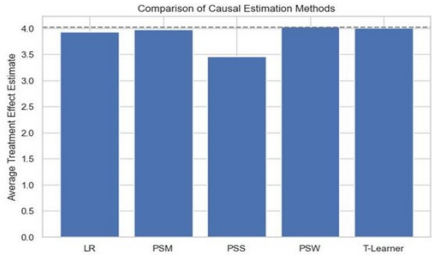


Fig. 3. Comparison of causal estimates across different methods

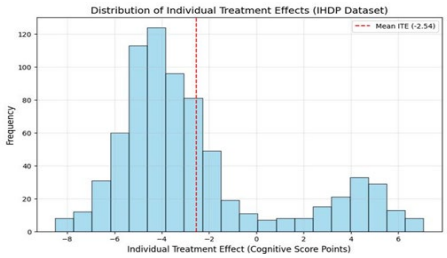


Fig. 4. Histogram showing variation in individual treatment effects.

### D. Conclusion & Recommendation

This study demonstrates that causal-inspired machine learning methods can accurately estimate individual treatment effects from observational data, this has ability to enhance personalized medicine. Causal inference methods with integration of machine learning are essential for personalized medicine, offering tailored insights beyond average outcomes. Treatment effects can vary across individual, hence T-learner, a variant of Meta-learner provides a robust ITE estimates.

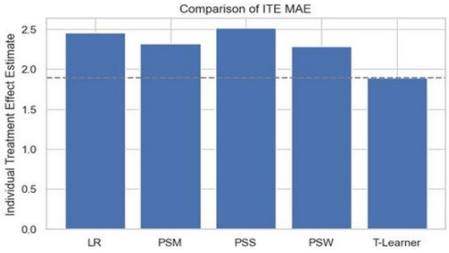


Fig. 5. Histogram showing comparison of individual treatment effect of mean absolute error across different learners

In conclusion, Meta-learners are better off in estimating Individual treatment effects especially when a personalized intervention is paramount.

### *E. Acknowledgements*

The authors would like to express their sincere gratitude to the developers and contributors of the DoWhy causal inference library, particularly for their publicly available tutorials and examples, including the IHDP dataset notebook. The tutorial titled “Estimating Causal Effects on the IHDP Dataset” (DoWhy Example) served as a foundational guide and practical reference for the implementation and extension of causal-inspired machine learning methods in this study. Their work has significantly contributed to advancing reproducible research in causal inference and enabling broader applications in personalized medicine.

## References

[1] Alberto Abadie & Guido W Imbens. Bias-corrected matching estimators for average treatment e ects. Journal of Business & Economic Statistics, 29(1):111, 2011

[2] Fadnavis, R., & Kshirsagar, M. (2023, April). Applicability of Machine Learning for Personalized Medicine. In International Conference on Information and Communication Technology for Intelligent Systems (pp. 315-324). Singapore: Springer Nature Singapore.

[3] Serrano, D. R., Luciano, F. C., Anaya, B. J., Ongoren, B., Kara, A., Molina, G., & Lalatsa, A. (2024). Artificial intelligence (AI) applications in drug discovery and drug delivery: Revolutionizing personalized medicine. Pharmaceutics, 16(10), 1328.

[4] A. Kumar et al., "Data-Driven Healthcare Solutions Using Machine Learning," IEEE Trans. Biomed. Eng., vol. 71, no. 2, pp. 210–225, 2024

[5] G. W. Imbens and D. B. Rubin, Causal Inference in Statistics, Social, and Biomedical Sciences. Cambridge Univ. Press, 2015

[6] A. Abadie and G. W. Imbens, "Bias-Corrected Matching Estimators for Average Treatment Effects," J. Bus. Econ. Stat., vol. 29, no. 1, pp. 1–11, 2011.

[7] C. Lee, N. Mastronarde, and M. van der Schaar, "Estimation of Individual Treatment Effect in Latent Confounder Models via Adversarial Learning," arXiv:1811.08943, 2018.

[8] S. Athey and G. Imbens, "Recursive partitioning for heterogeneous causal effects," *Proc. Nat. Acad. Sci.*, vol. 113, no. 27, pp. 7353-7360, 2016, doi: 10.1073/pnas.1510489113.

[9] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2009

[10] R. K. Crump et al., "Dealing with limited overlap in estimation of average treatment effects," *Biometrika*, vol. 96, no. 1, pp. 187-199, 2009, doi: 10.1093/biomet/asn055.

[11] P. R. Rosenbaum and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika, vol. 70, no. 1, pp. 41-55, 1983, doi: 10.1093/biomet/70.1.41.

[12] DoWhy example on ihdp (Infant Health and Development Program) dataset DoWhy documentation. (n.d.).