From Population Effects to Personalized Medicine:

Causal Inference with Overlap Weighting and Causal Trees

Group 2

W. K. H. Hasara – s16210
K. J. A. Jayawardana – s16035
O. K. Udayana – s16073

DS 4005: Causal Inference

Department of Statistics, University of Colombo

August 24, 2025

1. Abstract

This study examines the causal effect of antihypertensive medication on 10-year coronary heart disease (CHD) risk by applying advanced causal inference techniques to observational data. Although randomized controlled trials are the benchmark for causal inference, observational studies offer valuable insights when randomization is constrained. We utilised two complementary methods, overlap weighting (OW) and causal trees, to estimate both overall and subgroup-specific heterogeneous treatment effects (HTEs).

The overlap weighting method, based on propensity scores, constructs a pseudo-population focusing on patients with overlapping treatment probabilities, achieving covariate balance and enabling the estimation of the Average Treatment Effect in the Overlap population (ATO). Our findings showed a slight, non-significant increase in 10-year CHD risk with antihypertensive medication, as confidence intervals included the null.

Using causal trees, we detected variation in treatment effects among subgroups defined by baseline characteristics such as BMI, systolic blood pressure, heart rate, and cholesterol. Patients with higher BMI and BP seemed to benefit from medication, whereas leaner men may have experienced harm, demonstrating substantial heterogeneity in responses.

Overall, our results stress the need to look beyond average effects in medical decision-making. By combining weighting and machine learning–based causal inference, we deliver both robust population-level estimates and exploratory subgroup insights, supporting more personalized treatment approaches. Future work should confirm these findings in larger, external cohorts and consider alternative approaches like targeted maximum likelihood estimation (TMLE) or Bayesian causal forests to enhance robustness and clinical relevance.

2. Table of Contents

3. Introduction

3.1 Background and Context

Hypertension is one of the most common chronic conditions worldwide and a major risk factor for coronary heart disease (CHD). Antihypertensive medications are widely prescribed to reduce high blood pressure and lower cardiovascular risk. However, it remains an open question whether such medications provide equal benefits to all patients or if their effectiveness varies across different subgroups.

In medical research, causal inference plays a critical role in addressing such questions. Unlike traditional statistical associations, causal inference methods aim to estimate the effect of a treatment while adjusting for confounding factors. This allows us to go beyond "Does the treatment work?" and explore important follow-up questions such as:

- Does antihypertensive medication reduce the 10-year risk of CHD?
- For whom does it work best, and for whom might it be less effective?
- Are there subgroups of patients who may even be harmed by the treatment?

3.2 Objectives of the Study

This study aims to evaluate the causal effect of antihypertensive medication use at baseline on the 10-year risk of coronary heart disease. Specifically, we address two core objectives:

1. **Population-level effects**: Estimate the average treatment effect (ATE) of blood pressure medication using robust weighting techniques.
2. **Subgroup-level effects**: Explore heterogeneous treatment effects across patient subgroups using causal tree methods with doubly robust pseudo-outcomes.

By combining these approaches, we seek not only to provide a reliable estimate of the average causal effect but also to uncover potential subgroup-specific differences that can guide personalized treatment strategies.

4. Theory and Methodology

4.1 Overlap Weighting

Overlap weighting (OW) is a new propensity score-based technique designed to adjust for confounding in causal treatment effect estimation in observational studies (Zajichek & Grunkemeier, 2024). It serves as a robust alternative to inverse probability of treatment weighting (IPTW) to address issues related to extreme propensity scores (Zajichek & Grunkemeier, 2024; Li & Thomas, 2018).

The probability of a patient receiving the opposite treatment from what was observed is indicated by Overlap Weighting. For a treated unit (T=1), the weight is (1 - PS), and for a control unit (T=0), it is PS (Li & Thomas, 2018). This construction ensures that patients with propensity scores near 0.5 contribute most to the effect estimate, while those with scores close to 0 or 1 are down-weighted. This indicates uncertainty in treatment decisions (Zajichek & Grunkemeier, 2024). OW targets the "overlap population", comprising individuals with the most similar characteristics in treatment and control groups. The corresponding estimand is the Average Treatment Effect in the Overlap population (ATO), which is particularly relevant in clinical practice (Li & Thomas, 2018).

A significant advantage of OW is achieving exact balance in means of covariates between treatment and control groups when propensity scores are estimated via logistic regression, resulting in a standardized mean difference (SMD) of zero for confounders in the OW-adjusted sample (Li et al., 2016). Unlike IPTW, OW weights are bounded between 0 and 1, preventing extreme weights from disproportionately influencing the treatment effect and eliminating the need for arbitrary trimming. OW also minimises large-sample variance of the weighted average treatment effect, showing robustness in scenarios with highly separated PS distributions (Zajichek & Grunkemeier, 2024). Compared to IPTW, which estimates the average treatment effect for the entire population, OW focuses on the overlap population, effectively addressing biases and inefficiencies associated with extreme PS values. IPTW weights can be unbounded and skewed, while OW weights reduce the influence of outliers without discarding data (Li & Thomas, 2018).

4.2 Causal Trees

Causal trees analyze heterogeneous treatment effects by partitioning data to estimate outcomes in specific subgroups (Brouquet et al., 2025; Younas et al., 2022). Unlike predictive decision trees, they focus on causal inference, identifying which patients benefit most or least from interventions (Courthoud, 2023). Causal trees split data into nodes, compute average outcomes for treated and untreated groups, and derive treatment effects from these differences (Younas et al., 2022). They employ an auxiliary variable (Y*), which offers an unbiased but high-variance effect estimate, and use "honest" sample splitting to reduce bias (Courthoud, 2023). Splitting often minimizes within-leaf treatment effect variance (EMSE; Athey & Imbens, 2016).

Advantages are clear visualization, computational speed, and robust within-group inference, but estimates may lack generalizability, and sample splitting can diminish power (Courthoud, 2023; Younas et al., 2022). Ensembles like OCTE address limitations by combining multiple trees for improved accuracy (Younas et al., 2022). Applications include identifying subgroups with adverse responses to treatments or varied effects of marketing incentives (Brouquet et al., 2025; Courthoud, 2023).

5. Data and Implementation

The dataset used in this analysis comes from a cohort of **4,240 participants**, with information on **demographics, lifestyle factors, clinical measures, and cardiovascular outcomes**. The primary treatment variable was **antihypertensive medication use (BPMeds)**, coded as a binary indicator (1 = taking medication, 0 = not taking). The outcome of interest was the **10-year risk of coronary heart disease (TenYearCHD)**, also represented as a binary variable (1 = developed CHD within 10 years, 0 = did not).

5.1 Covariates

- **Demographics & socio-economic**: age, sex (male), education.
- **Lifestyle**: smoking status (currentSmoker), cigarettes per day (cigsPerDay).
- **Clinical history**: prevalent stroke, prevalent hypertension, diabetes.
- **Biomarkers & physiological measures**: total cholesterol (totChol), systolic and diastolic blood pressure (sysBP, diaBP), body mass index (BMI), heart rate, glucose.

5.2 Missing Data Handling

The dataset contained some missing values across variables:

Table 1: Missing Values Across Variables

Variable	Missing Values
Education	105
cigsPerDay	29
totChol	50
BMI	19
heartRate	1
BPMeds	53
glucose	388

- Missing **treatment (BPMeds)** were excluded to avoid ambiguity.
- For **continuous variables**, missing values were imputed with the **median** of the observed distribution.
- For **categorical variables** (education), missing values were imputed using the **mode**.
- All predictors (covariates) were organized into a design matrix **X**, with treatment **T** and outcome **Y** defined separately for causal analysis.

5.3 Final Analytical Sample

After cleaning, the dataset contained a slightly reduced sample (due to BPMeds filtering), with complete information across treatment, outcome, and imputed covariates. This preprocessing ensured that all individuals could be included in the causal inference analysis while minimizing bias from missingness.

5.4 Implementation of Weighting and ATE Estimation

In the first method, we used **overlap weighting** to estimate the **Average Treatment Effect (ATE)** of blood pressure medication (BPMeds) on the 10-year risk of coronary heart disease (CHD).

5.5 Implementation of Heterogeneous Treatment Effect Analysis Using Causal Trees

To investigate the effect of **medication** on the **10-year risk of heart disease**, we fitted a **causal tree** using the **doubly robust (DR) pseudo-outcome**.

Because treatment assignment was **highly imbalanced** (≈127 treated vs ≈4000 controls) and the treatment/outcome relationships are likely **nonlinear with interactions**, we used a **doubly-robust (DR) approach** to estimate individual/leaf-level treatment effects and then **fit a causal tree on those DR pseudo-outcomes**. This approach is more reliable than splitting directly on the raw outcome ("normal" tree), which can be biased when overlap is poor or the outcome model is misspecified.

The causal tree allows us to identify **subgroups of patients where the treatment effect varies**, i.e., heterogeneous treatment effects (HTE).

6. Results and Discussion

6.1 Weighting and ATE Estimation

6.1.1 Propensity Score and Weights

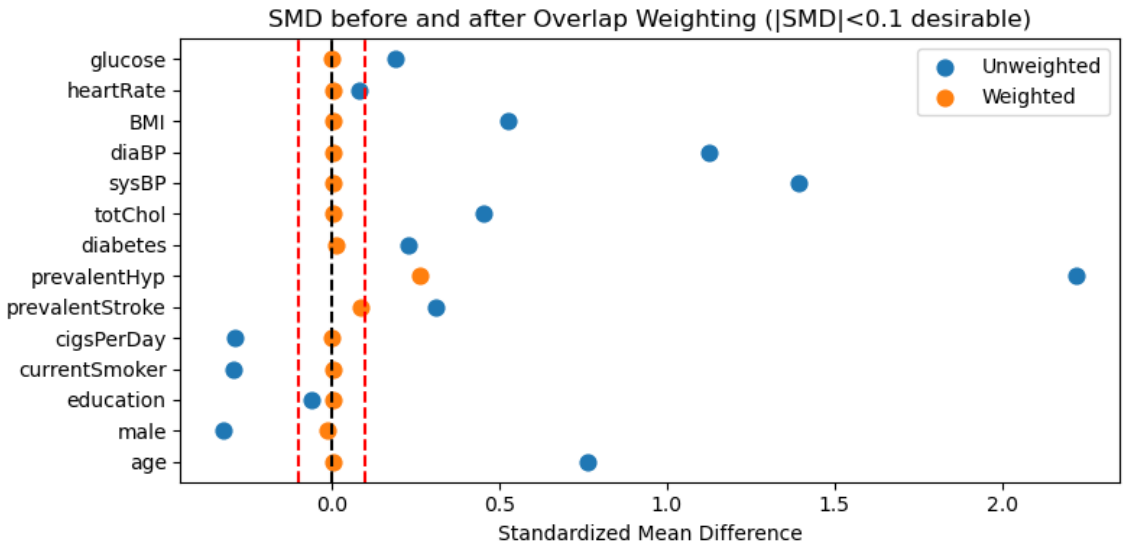
Propensity scores were estimated using logistic regression. The mean propensity score was **0.14** among treated individuals and **0.03** among controls, indicating treatment was relatively rare. After weighting, covariate balance was substantially improved, as shown by standardized mean differences (SMD). For example:

- **Unweighted SMD for systolic BP:** 1.39 → reduced to 0.004 after weighting.
- **Unweighted SMD for age:** 0.76 → reduced to 0.004 after weighting.
- All covariates achieved near-zero weighted SMD, indicating good balance.

Table 2: Standardized Mean Differences Before and After Weighting

Covariate	Unweighted	Weighted
prevelentHyp	2.22	0.26
sysBP	1.39	<0.01
diaBP	1.12	<0.01
age	0.76	<0.01
BMI	0.52	<0.01
totChol	0.45	<0.01
male	-0.32	-0.01
prevalentstroke	0.31	0.09
currentSmoker	-0.29	0.01
cigsPerDay	-0.29	<0.01
diabetes	0.23	0.01
glucose	0.19	<0.01
heartRate	0.08	<0.01
education	-0.06	<0.01

Figure 1: Standardized Mean Differences Before and After Weighting



6.1.2 Risk Estimates

Using the weighted population, the following risk estimates were calculated:

- **Risk under treatment (Risk(1)):** 0.3263 (95% CI: 0.247–0.412)
- **Risk under no treatment (Risk(0)):** 0.2743 (95% CI: 0.246–0.303)
- **Average treatment effect(ATE):** 0.0520 (95% CI: –0.031, 0.140)
- **Risk Ratio (RR):** 1.189 (95% CI: 0.890, 1.537)

The weighted analysis suggests that **BPMeds use is associated with a slightly higher 10-year CHD risk** (RD ≈ +5.2 percentage points; RR ≈ 1.19). However, the **confidence intervals include the null** for both RD and RR, meaning the estimated effect is not statistically significant at conventional levels.

In addition to risk difference (RD) and risk ratio (RR) estimates, we also fitted a **Generalized Linear Model (GLM)** with a binomial family and logit link to further evaluate the association between BPMeds use and 10-year CHD risk. The GLM results are presented below:

Figure 2: Generalized Linear Model Regression Results

Generalized Linear Model Regression Results

Dep. Variable:	TenYearCHD	No. Observations:	4187
Model:	GLM	Df Residuals:	211.59
Model Family:	Binomial	Df Model:	1
Link Function:	Logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-130.18
Date:	Wed, 20 Aug 2025	Deviance:	260.36
Time:	14:12:55	Pearson chi2:	214.
No. Iterations:	4	Pseudo R-squ. (CS):	0.0001641
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.9729	0.217	-4.486	0.000	-1.398	-0.548
BPMeds	0.2478	0.299	0.828	0.408	-0.339	0.835

The estimated coefficient for BPMeds is **0.248 (p = 0.408)** with a 95% confidence interval spanning **-0.339 to 0.835**, which includes the null value. This confirms that there is **no statistically significant association** between BPMeds use and 10-year CHD risk at conventional levels of significance. These regression results are consistent with our earlier findings from RD and RR estimates, where the confidence intervals also included the null.

6.1.3 Sensitivity Analysis (E-values)

- Point estimate RR=1.189 → E-value = 1.664
- Lower CI (0.890) → E-value = 1.497

This implies that an unmeasured confounder associated with both treatment and outcome by an RR of at least **~1.5–1.7** could fully explain away the observed effect.

6.2 Heterogeneous Treatment Effect Analysis Using Causal Trees

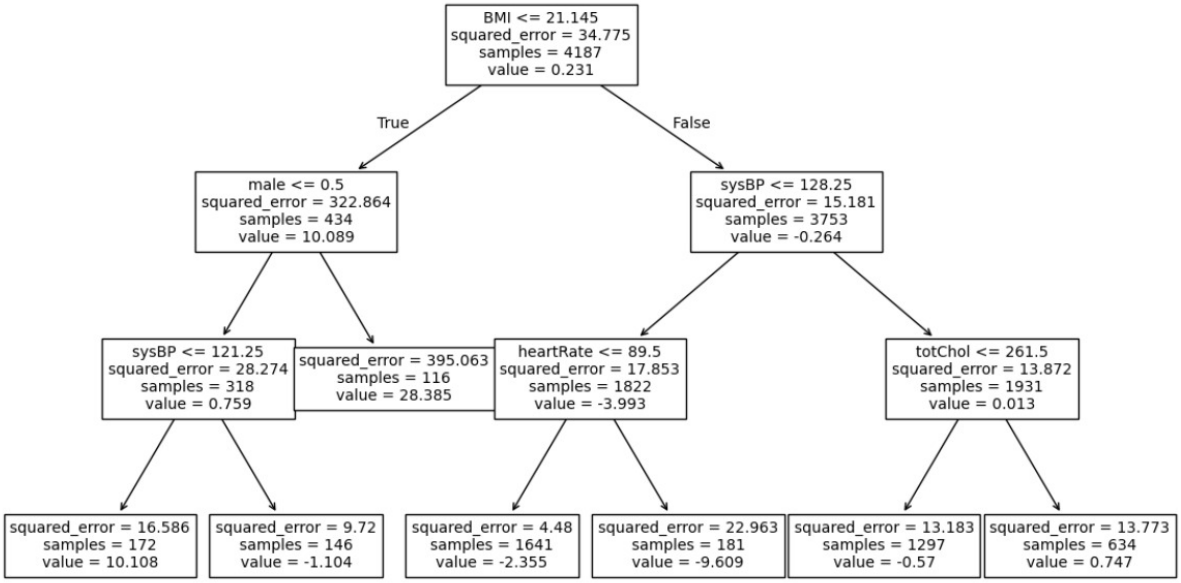
6.2.1 Causal Tree Structure

The tree split the population based on key covariates, including **BMI, gender, systolic blood pressure (sysBP), heart rate, and total cholesterol (totChol)**.

Key splits include:

- **BMI ≤ 21.145**: further split by gender and systolic BP.
- **BMI > 21.145**: further split by systolic BP, heart rate, and total cholesterol.

Figure 3: Causal Tree Structure



This hierarchical structure identifies subgroups of patients with similar baseline characteristics and **similar responses to the treatment**.

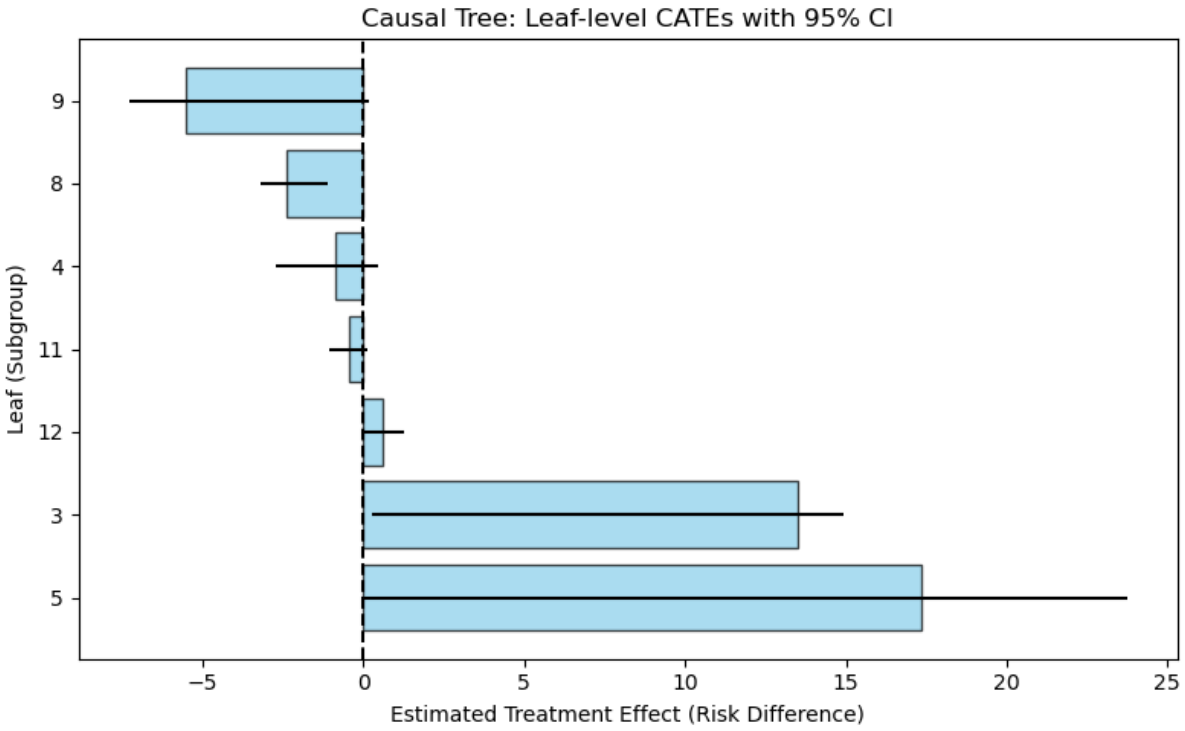
6.2.2 Leaf-level Treatment Effects

At the terminal nodes (leaves), we computed **tau_hat**, the estimated treatment effect (difference in 10-year heart disease risk between treated and untreated), along with bootstrap.

Figure 4: Leaf-level Treatment Effects

```
--- Leaf-level Honest Estimates (tau_hat with bootstrap 95% CI) ---
 leaf    n      n_eff   tau_hat     ci_lo     ci_hi
    5  116   3.449563 17.369410 -0.081759 23.753646
    3  172   1.123565 13.496641  0.249393 14.949569
   12  634  83.836187  0.606747 -0.016861  1.275028
   11 1297 105.474874 -0.459265 -1.049838  0.138670
    4  146   5.640430 -0.872862 -2.713023  0.451409
    8 1641  10.888151 -2.378820 -3.192333 -1.134029
    9  181   3.175713 -5.542758 -7.296424  0.188924
```

Figure 5: Causal Tree Leaf-Level CATEs with 95% Confidence Interval



- **Positive tau_hat** indicates the medication is associated with an **increase in 10-year heart disease risk**.
- **Negative tau_hat** indicates the medication **reduces the risk**, which is desirable.

Policy Recommendations per Leaf

Based on the leaf-level estimates, a **simple treatment policy** was proposed:

- Subgroups with **negative tau_hat** benefit from the medication and should be treated.
- Subgroups with **positive tau_hat** are potentially harmed by the medication, so treatment is **not recommended**.

Figure 6: Policy Recommendations per Leaf

Policy summary (per-leaf):

leaf	n	tau_hat	ci_lo	ci_hi	recommend_treat_if_tau_negative
5	116	17.369410	-0.081759	23.753646	False
3	172	13.496641	0.249393	14.949569	False
12	634	0.606747	-0.016861	1.275028	False
11	1297	-0.459265	-1.049838	0.138670	True
4	146	-0.872862	-2.713023	0.451409	True
8	1641	-2.378820	-3.192333	-1.134029	True
9	181	-5.542758	-7.296424	0.188924	True

The analysis reveals that the effects of the medication are not uniform across all patients, highlighting the presence of heterogeneous treatment effects. Specifically, patients with a body mass index (BMI) greater than 21.145, combined with elevated blood pressure or heart rate, tend to experience a reduction in their 10-year risk of heart disease and are thus likely to benefit from the medication. In contrast, the medication may be detrimental for patients characterised by a low BMI (≤21.145) who are also male, as they show an increased risk when treated. These findings underscore the clinical importance of moving away from a 'one-size-fits-all' approach to medication, instead supporting the need for personalised treatment decisions. The causal tree analysis provides clear guidance for clinicians to identify which patient subgroups are likely to benefit from treatment and which might be at risk, ultimately enabling more tailored and effective healthcare recommendations.

Together, the two methods provide a fuller picture of the treatment effect:

- **Overlap Weighting (Method 1)**: Provides a stable and interpretable **population-level estimate** of the ATE, which in this case suggests no statistically significant effect of BPMeds on average.
- **Causal Tree (Method 2)**: Highlights **heterogeneity of treatment effects**, showing that the impact of BPMeds depends strongly on patient characteristics.

From a clinical standpoint, these results imply that **a one-size-fits-all treatment strategy may not be optimal**. While BPMeds may not significantly change CHD risk on average, certain subgroups (e.g., higher BMI, higher BP patients) may benefit, while others (e.g., lean males) may experience adverse effects.

7. Limitations

Several limitations should be acknowledged. First, the dataset may contain **unmeasured confounding variables** that were not accounted for, which could bias the causal estimates. Second, missing data and potential measurement errors in the covariates could reduce reliability. Third, the causal tree method relies heavily on sample size and overlap; in subgroups with sparse data or poor overlap, treatment effect estimates were unstable. Finally, our analysis focused only on observational data, so residual confounding cannot be completely ruled out.

7.1 Future Research Directions

Future work could extend this analysis in several ways:

- Incorporating **additional causal inference methods** such as targeted maximum likelihood estimation (TMLE) or Bayesian causal forests for robustness checks.
- Exploring **longitudinal models** to capture changes in medication use and CHD risk over time, rather than a single 10-year outcome.
- Including **external validation datasets** to assess the generalizability of findings.

8. Conclusion

This report examined the causal effect of antihypertensive medication on long-term cardiovascular outcomes through a dual-method approach that combined overlap weighting and causal tree analysis. The overlap weighting method successfully balanced confounding variables, offering a robust population-level estimate of the treatment effect. Our findings suggested that, on average, antihypertensive medication did not significantly alter the 10-year risk of CHD, with effect estimates close to the null. While this aligns with some prior observational research, it also raises critical questions about the limits of population-level inference in guiding clinical decisions.

The causal tree analysis provided a more nuanced view, revealing that treatment effects were not homogeneous across all patients. Instead, the impact of antihypertensive medication varied by demographic, lifestyle, and physiological characteristics. Specifically, higher-risk patients with elevated BMI, blood pressure, or heart rate tended to benefit, while certain low-BMI male subgroups appeared to be adversely affected. This heterogeneity illustrates the limitations of a "one-size-fits-all" approach in cardiovascular prevention and points toward the need for individualized treatment strategies.

From a methodological standpoint, this study demonstrates the value of combining weighting-based approaches with machine learning–driven heterogeneity detection. Overlap weighting ensured reliable baseline adjustment and reduced bias from extreme propensity scores, while causal trees uncovered clinically meaningful subgroups that would remain hidden under average effect estimates. Nevertheless, challenges remain, including small treatment groups, limited overlap in some subpopulations, and the ever-present risk of unmeasured confounding.

In conclusion, our analysis highlights both the promise and complexity of causal inference in medical research. While antihypertensive medication may not provide uniform benefits across the population, careful subgroup analysis suggests that targeted treatment could yield meaningful improvements in patient outcomes. The results underscore the importance of integrating causal inference methods into precision medicine frameworks, where treatment recommendations are informed not only by average effects but also by patient-specific characteristics. Future work should expand on these findings using larger datasets, richer covariate information, and advanced causal machine learning methods, thereby strengthening the evidence base for personalized hypertension management.

9. References

Brouquet, A., Schmidt, A., & Herquelot, E. (2025). P47 - Using causal trees to explore the heterogeneity of effects. Deleted Journal, 73, 203078. https://doi.org/10.1016/j.jeph.2025.203078

Darondeau, P., & Degano, P. (n.d.). *HAL Id: inria-00075607 https://inria.hal.science/inria-00075607v1 Causal trees*. https://inria.hal.science/inria-00075607v1

Li, F., Lock, K., Alan, M., & Zaslavsky, M. (n.d.). *Balancing Covariates via Propensity Score Weighting*.

Li, F., & Thomas, L. E. (2019). Addressing Extreme Propensity Scores via the Overlap Weights. *American Journal of Epidemiology*, *188*(1), 250–257. https://doi.org/10.1093/aje/kwy201

We Want to Hear from You! (n.d.). https://towardsdatascience.com/understanding-causal-trees-920177462149/

Younas, N., Ali, A., Hina, H., Hamraz, M., Khan, Z., & Aldahmani, S. (2022). Optimal Causal Decision Trees Ensemble for Improved Prediction and Causal Inference. *IEEE Access*, *10*, 13000–13011. https://doi.org/10.1109/ACCESS.2022.3146406

Zajichek, A. M., & Grunkemeier, G. L. (2024). Propensity scores used as overlap weights provide exact covariate balance. In *European Journal of Cardio-thoracic Surgery* (Vol. 66, Issue 3). European Association for Cardio-Thoracic Surgery. https://doi.org/10.1093/ejcts/ezae318

10. Appendices

Link to the dataset: https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset

GitHub Repository: https://github.com/kusaraudayana/causal_inference_medicaldata_group2