
CAUSAL DISCOVERY REPORT ON HEART DISEASE

TECHNICAL REPORT



April 8, 2025

ABSTRACT

This report presents an analysis of heart disease risk factors through the examination of a comprehensive dataset encompassing demographic, physiological, and clinical variables. Utilizing the Fast Causal Inference (FCI) algorithm, we implemented a thorough causal discovery procedure, including data preprocessing, algorithm selection assisted by a large language model (LLM), and graph tuning with bootstrap methods. Our findings revealed complex causal relationships among variables such as age, chest pain type, and maximum heart rate. Notably, age directly influences coronary artery status and, consequently, heart disease outcomes. Additionally, the treatment effect of resting blood pressure on health outcomes was assessed, yielding an Average Treatment Effect (ATE) of approximately 0.059, albeit with considerable uncertainty. Our contribution lies in enhancing the understanding of causal interdependencies in cardiovascular health, highlighting the need for personalized treatment strategies based on individual characteristics such as age and sex, and paving the way for improved medical interventions in heart disease.

Keywords Causal Discovery, Large Language Model, FCI, heart disease

1 Introduction

The analysis of heart disease risk factors is a critical area of medical research, reflecting the interplay of various biological and environmental factors that contribute to cardiovascular health. This dataset encapsulates a range of variables, including demographic, physiological, and clinical factors that are known to influence the likelihood of heart disease. By exploring relationships between these variables, including age, sex, cholesterol levels, and exercise-induced symptoms, we aim to uncover potential causal connections that can shed light on the mechanisms of heart disease development. Through rigorous causal discovery and inference techniques, we will not only identify key predictors of heart disease but also enhance our understanding of the underlying risk factors impacting individuals' health, ultimately contributing to improved prevention and treatment strategies in clinical practice.

2 Background Knowledge

2.1 Detailed Explanation about the Variables

The dataset in question includes various factors associated with heart disease, each represented by specific variables. Key variables include **age**, which influences heart disease risk; **sex**, indicating differences in disease prevalence; **cp** (chest pain type), which reveals potential coronary artery disease severity; **trestbps** (resting blood pressure) and **chol** (cholesterol levels), both established risk factors; and **fbs** (fasting blood sugar), which points to diabetes risk. Additional variables such as **restecg** (resting ECG results), **thalach** (maximum heart rate), and **exang** (exercise-induced angina), play critical roles in assessing cardiac health, while the **ca** (number of vessels affected) and **thal** (thalassemia presence) gauge the severity of cardiovascular conditions. The target variable indicates the presence or absence of heart disease.

Understanding the underlying relationships among these variables necessitates considering broader domain knowledge. Familiarity with traditional risk factors—such as smoking, physical inactivity, and genetic predisposition—is crucial for a comprehensive analysis.

tions—provides vital context for analyzing causal links. Knowledge of heart disease pathophysiology, including concepts like ischemia and myocardial infarction, enables a deeper understanding of how these variables interact. Employing appropriate statistical techniques for causal inference and being aware of clinical correlations related to age, sex, and medical history further enhances the analysis. Additionally, familiarity with clinical guidelines for heart disease risk assessment is essential for interpreting the implications of findings in the dataset.

2.2 Possible Causal Relations found by LLM

The following are potential causal relationships suggested by the language model, which are visualized in Figure 1. Please note that only variables present in our dataset are included in the figure.

- **age → target:** As age increases, the risk of heart disease generally rises.
- **sex → target:** Males may have different risks and presentations of heart disease compared to females.
- **cp → target:** Different types of chest pain are linked to various levels of risk for heart disease.
- **trestbps → target:** Higher resting blood pressure can lead to greater risk of heart disease.
- **chol → target:** Elevated cholesterol levels can lead to increased risk of heart disease due to plaque buildup in arteries.
- **fbs → target:** High fasting blood sugar levels can indicate diabetes, a risk factor for heart disease.
- **restecg → target:** Abnormal resting ECG results may indicate underlying issues that are causal or closely linked to heart disease.
- **thalach → target:** Reduced maximum heart rate can suggest underlying cardiovascular problems that could lead to heart disease.
- **exang → target:** Presence of exercise-induced angina is directly linked to the likelihood of heart disease.
- **oldpeak → target:** Higher ST depression levels during exercise can indicate more significant heart disease.
- **slope → target:** Different slope values during peak exercise can be correlated with heart disease risk.
- **ca → target:** The number of affected vessels directly impacts heart disease severity and risk.
- **thal → target:** Abnormal thalassemia or condition impacts overall heart health and can be correlated with heart disease.

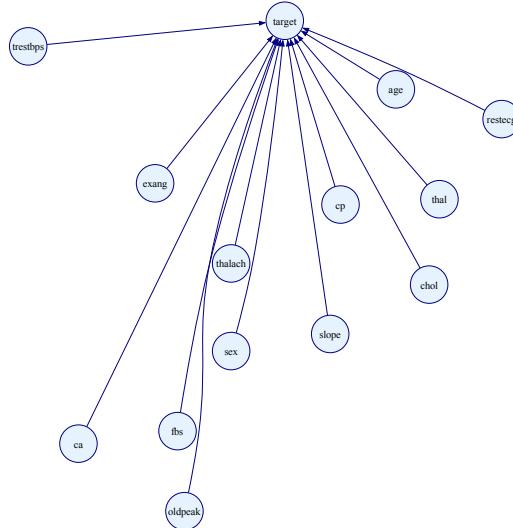


Figure 1: A Causal Graph Suggested by LLM.

3 Dataset Descriptions and EDA

The following provides a preview of our original dataset. If the dataset contains more than 10 columns, a random subset of 10 columns is displayed for illustrative purposes.

Table 1: Dataset Preview.

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
52	1	0	125	212	0	1	168	0	1.000000	2	2	3	0
53	1	0	140	203	1	0	155	1	3.100000	0	0	3	0
70	1	0	145	174	0	1	125	1	2.600000	0	0	3	0
61	1	0	148	203	0	1	161	0	0.000000	2	1	3	0
62	0	0	138	294	1	1	106	0	1.900000	1	3	2	0

3.1 Data Properties

We employed several statistical methods to identify data properties, including:

Basic Data Characteristics

The shape of the data, variable types, and the presence of missing values were assessed directly from the DataFrame. In contrast, properties such as time-series structure and heterogeneity were inferred with LLM based on user queries and DataFrame.

Linearity Testing

We conducted the Ramsey's RESET test to assess linearity between each pair of variables. When the total number of possible variable pairs was fewer than 100, all pairs were tested. If the number exceeded 100, a random subset of 100 pairs was selected for testing to ensure computational feasibility. To account for multiple testing, we employed the Benjamini and Yekutieli procedure, which is robust when dealing with dependent or correlated data. The linearity assumption was considered satisfied only if all tested pairs exhibited linearity; otherwise, it was considered violated.

Normality of Residuals

The assumption of Gaussian (normally distributed) noise was assessed using the Shapiro-Wilk test. The testing approach depended on the outcome of the linearity evaluation. If linearity was satisfied, we fitted ordinary least squares (OLS) models for each variable pair and extracted the residuals for testing. If linearity was not satisfied, we used a flexible non-parametric method—locally weighted scatterplot smoothing (LOWESS)—to model the relationships and obtain residuals. The Benjamini and Yekutieli correction was again applied to control for false discovery under multiple testing.

Properties of the dataset we analyzed are listed below.

Table 2: Data Properties.

Shape ($n \times d$)	Data Type	Missing Value	Linearity	Gaussian Errors	Time-Series	Heterogeneity
(1025, 14)	Mixture	False	False	False	False	False

3.2 Distribution Analysis

The following figure presents distributions of various variables. The orange dashed line indicates the mean, while the black solid line denotes the median. Variables are categorized into three types based on their distributional characteristics.

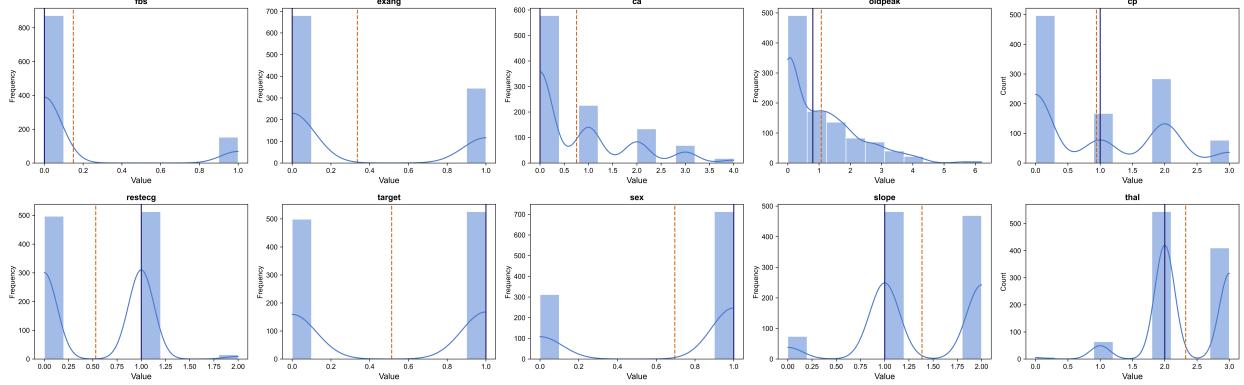


Figure 2: Distribution Plots of Variables.

Numerical Variables

- Slight left skewed distributed variables: cp, restecg, target, sex
- Slight right skewed distributed variables: fbs, exang, ca, oldpeak, slope, thal
- Symmetric distributed variables: None

3.3 Correlation Analysis

- **Strongly Correlated Variables (≥ 0.9):** None
- **Moderately Correlated Variables (0.1 – 0.9):** slope - oldpeak
- **Weakly Correlated Variables (≤ 0.1):** ca - fbs, ca - exang, oldpeak - exang, oldpeak - ca, cp - exang, cp - ca, cp - oldpeak, restecg - fbs, target - exang, target - ca, target - oldpeak, target - cp, target - restecg, sex - exang, sex - ca, sex - target, slope - exang, slope - cp, slope - target, thal - exang, thal - ca, thal - oldpeak, thal - cp, thal - target, thal - sex, etc.

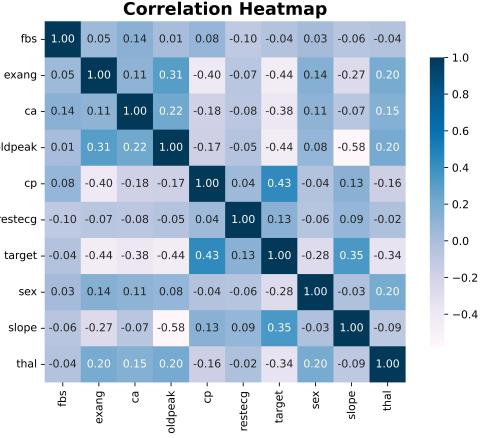


Figure 3: Correlation Heatmap of Variables.

4 Causal Discovery Procedure

In this section, we provide a detailed description of the causal discovery process implemented by Causal Copilot. We also provide the chosen algorithms and hyperparameters, along with the justifications for these selections.

4.1 Data Preprocessing

In this initial step, we preprocessed the data and examined its statistical characteristics. This process involved data cleaning, handling missing values, and performing exploratory data analysis to examine variable distributions and inter-variable relationships.

4.2 Algorithm Recommendation assisted with LLM

Following preprocessing, we employed a large language model (LLM) to assist in selecting appropriate algorithms for causal discovery based on the statistical characteristics of the dataset and relevant background knowledge. The top three chosen algorithms, listed in order of suitability, are as follows:

- **PC:**

- **Description:** The Peter-Clark Algorithm (PC) is a constraint-based method that is flexible in handling both linear and non-linear relationships. It is efficient for medium-to-large-scale datasets and provides a CPDAG output.
- **Justification:** The PC algorithm is suitable for this dataset as it is flexible in handling both linear and non-linear relationships, which aligns with the dataset's non-linear characteristics. It is efficient for medium-scale datasets and provides a CPDAG output, which is acceptable for the user. Its strong empirical performance makes it a reliable choice.

- **FCI:**

- **Description:** The Fast Causal Inference (FCI) algorithm is a constraint-based method that is robust in handling latent variables and flexible with noise types. It is efficient for medium-to-large-scale datasets and provides a PAG output.
- **Justification:** The FCI algorithm is robust in handling latent variables and flexible with noise types, which is beneficial given the dataset's non-Gaussian noise. It provides a PAG output, which is acceptable for the user, and its robust empirical performance makes it a reliable choice for this dataset.

Considering data properties, algorithm capability and user's instruction, the final algorithm we choose is FCI.

4.3 Hyperparameter Values Proposal assisted with LLM

Once the algorithms were selected, the LLM aided in proposing hyperparameters for the chosen algorithm, which are specified below:

- **Significance Level:**

- **Value:** 0.05
- **Explanation:** Using a significance level of 0.05 is a standard choice for moderate sample sizes, ensuring a balance between detecting true causal relationships and avoiding false positives.

- **Independence Test Method:**

- **Value:** rcit
- **Explanation:** User specified

- **Maximum Depth for Skeleton Search:**

- **Value:** 4
- **Explanation:** A depth of 4 provides a good balance between accuracy and computational efficiency for a graph with 14 variables, allowing for a comprehensive search of causal structures.

4.4 Graph Tuning with Bootstrap and LLM Suggestion

In the final step, we performed graph tuning with suggestions provided by the Bootstrap and LLM.

We first applied the Bootstrap method to estimate the confidence level associated with each edge in the initial graph. Specifically:

- If an edge not present in the initial graph exhibited a Bootstrap confidence greater than 90%, we added it to the graph.
- Conversely, if an existing edge had a confidence lower than 10%, we removed it.
- For edges with moderate confidence (between 10% and 90%), we consulted the LLM to assess their validity and directionality, drawing on its extensive background knowledge.

The LLM contributed by:

- Reintroducing plausible edges that may have been overlooked by statistical methods;
- Removing or redirecting edges that appeared statistically valid but were conceptually implausible.

To improve the robustness of LLM-generated suggestions, we employed a voting mechanism. Importantly, LLM recommendations were not allowed to override high-confidence decisions made by the Bootstrap procedure. By integrating insights from both of Bootsrap and LLM to refine the causal graph, we can achieve improvements in graph's accuracy and robustness.

5 Causal Graph Estimation Summary

5.1 Causal Graph Discovered by the Algorithm

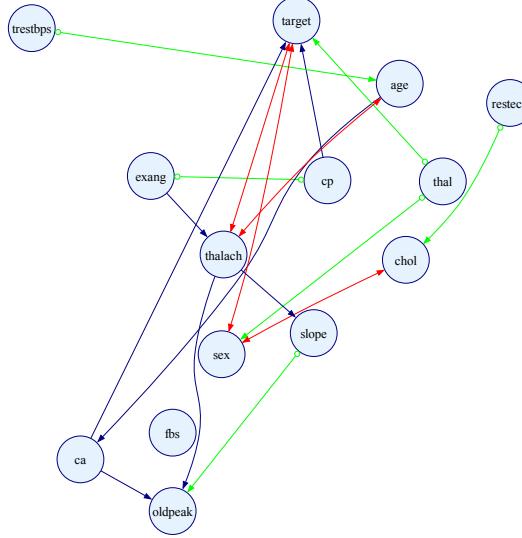


Figure 4: Causal Graph Discovered by the Algorithm. Solid lines represent causal edges identified by the algorithm, while dashed lines indicate strong correlations without inferred causality.

The above is the original causal graph produced by our algorithm.

The causal relationships among the variables exhibit a complex interplay influenced by direct causation, hidden confounding factors, and structural dependencies.

- The variable age directly affects the coronary artery status (ca), which in turn has implications on the target variable, indicative of heart disease outcomes. This demonstrates the role of age as a fundamental factor in cardiovascular health.
- The chest pain type (cp) exhibits a causal relationship with the target variable, reflecting its influence on heart disease diagnosis.
- The thalach (maximum heart rate achieved) serves as a critical mediating variable, influenced by exang (exercise induced angina), and subsequently affects oldpeak (depression induced by exercise relative to rest) and slope (of the peak exercise ST segment). These relationships suggest that physical exertion, as determined by thalach, is linked to indicators of cardiac stress and function during exercise.
- There is an absence of direct descendant relationships among certain variables, which indicates that trestbps (resting blood pressure) and thal (thalassemia) are not influenced by factors like age and sex, respectively. This might imply that these measurements capture distinct aspects of cardiovascular health that operate independently of these demographic characteristics.
- Additionally, some relationships are impacted by hidden confounders, particularly those involving age with thalach and sex with several variables including chol (cholesterol level) and target disease status, demonstrating that underlying factors can obscure direct causal pathways.

Overall, these interactions stress the importance of considering both direct and indirect connections within cardiovascular health assessments, as well as accounting for confounding influences that could affect interpretations of the data.

5.2 Causal Graph after Revision with Bootstrap and LLM

5.2.1 Bootstrap Probability

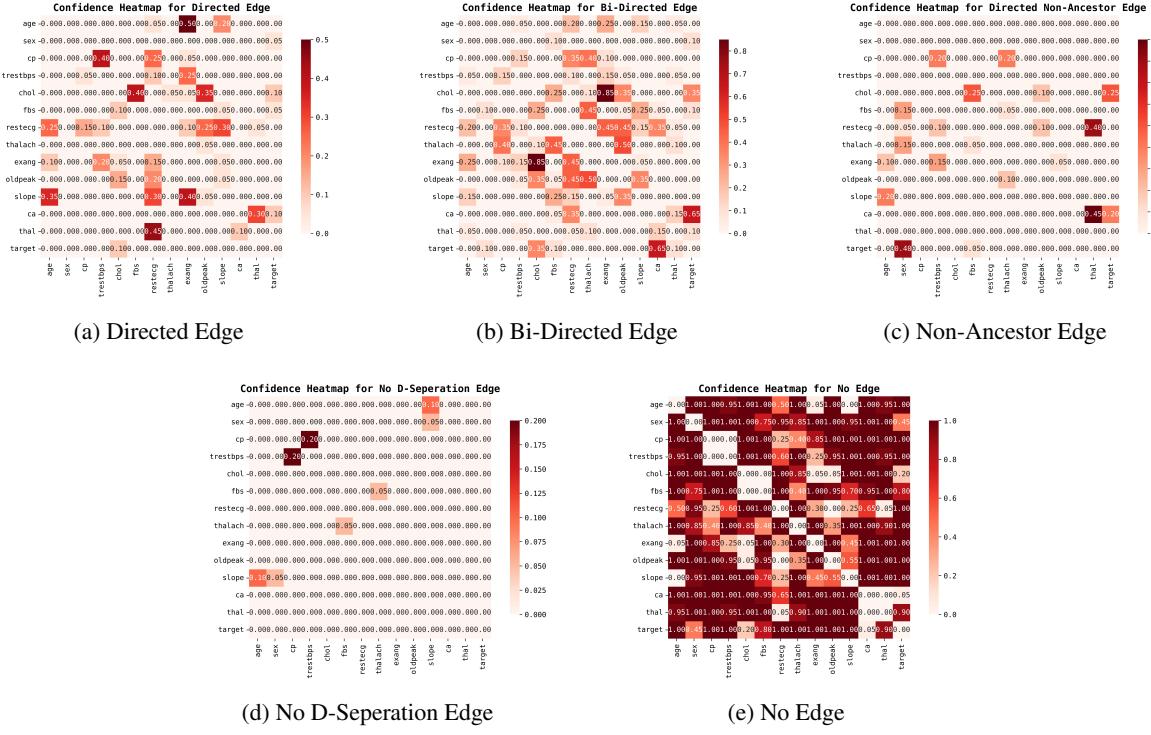


Figure 5: Confidence Heatmap of Different Edges

The above heatmaps show the confidence probability we have on different kinds of edges, including directed edge (\rightarrow), edge with hidden confounders (\leftrightarrow), edge of non-ancestor (o \rightarrow), egde of no D-Seperation set (o-o), No Edge, The heatmap of uncertain-edges, half-uncertain-edges is not shown because probabilities of all edges are 0.

To evaluate the confidence associated with each edge in the causal graph, we employed a bootstrapping procedure to estimate the probability of existence for each edge. From a statistical perspective, we categorize these probabilities into three levels:

- **High Confidence Edges:** none.
- **Moderate Confidence Edges:** exang causes thalach, thalach causes oldpeak, thalach causes slope, age causes ca, cp causes target, ca causes target, age has a hidden confounder with thalach, sex has a hidden confounder with chol, sex has a hidden confounder with target, thalach has a hidden confounder with target, cp has no D-separation set with exang.
- **Low Confidence Edges:** ca causes oldpeak, age causes ca, trestbps is not a descendant of age, thal is not a descendant of sex, restecg is not a descendant of chol, slope is not a descendant of oldpeak, thal is not a descendant of target.

5.2.2 LLM Pruning

By using the method mentioned in the Section 4.4, we provide a revise graph pruned with Bootstrap and LLM suggestion. Pruning results are as follows.

Bootstrap doesn't force or forbid any edges.

The following relationships are forbidden by LLM:

- **sex → thal:** There is no established evidence suggesting that sex directly influences thalassemia status, as thalassemia is primarily determined by genetic factors rather than demographic characteristics like sex.

The following are directions confirmed by the LLM:

- **oldpeak → slope:** Oldpeak, a measure of ST depression induced by exercise relative to rest, directly influences the slope of the ST segment on an electrocardiogram, which reflects heart function during stress and can indicate ischemia;
- **oldpeak → target:** Oldpeak can serve as a significant indicator of heart disease severity, directly influencing the likelihood of a positive diagnosis for the target variable, which represents the presence of heart disease.
- **ca → target:** The number of coronary artery blockages (ca) directly impacts the severity of heart disease, which is represented by the target variable, indicating whether a patient has heart disease or not;
- **thalach → slope:** An increase in thalach (maximum heart rate achieved) can indicate greater exercise capacity and, consequently, influence the slope of the ST segment during exercise testing, as a steeper slope may reflect better cardiovascular fitness;
- **thalach → target:** Thalach is directly influenced by physical activity and cardiovascular health, making it a key factor in determining the target variable of heart disease presence, as higher maximum heart rates typically suggest better heart health;
- **thalach → oldpeak:** Thalach contributes to the relationship with oldpeak since exercise capacity (reflected in thalach) can impact the level of ST segment depression (oldpeak) observed during stress testing, indicating ischemic responses;
- **thalach → exang:** Intervention in thalach, by increasing exercise, could lead to the onset of angina (exang) symptoms if the heart does not receive adequate oxygen during increased heart rate activities, establishing a causal link from thalach to exang;
- **age → trestbps:** Age is associated with an increase in blood pressure as one ages, due to physiological changes and the accumulation of risk factors that affect cardiovascular health;
- **thal → target:** Thalassemia and other thal-related cardiovascular conditions can lead to an increased risk of heart disease, thereby influencing the likelihood of having a positive target diagnosis for such diseases in medical assessments;
- **cp → target:** Chest pain (cp) is a symptom of underlying heart conditions, which directly influences the likelihood of a target event, such as a heart disease diagnosis;
- **cp → exang:** Chest pain (cp) can lead to increased levels of exertional angina (exang) as it reflects the heart's response to stress or physical activity, indicating that patients may experience angina during exertion related to their chest pain symptoms.
- **slope → target:** The slope of the segment on a stress test can indicate the presence and severity of heart disease, as a downward slope typically suggests ischemia, influencing the likelihood of a specific target outcome related to cardiac events;

This structured approach ensures a comprehensive and methodical analysis of the causal relationships within the dataset.

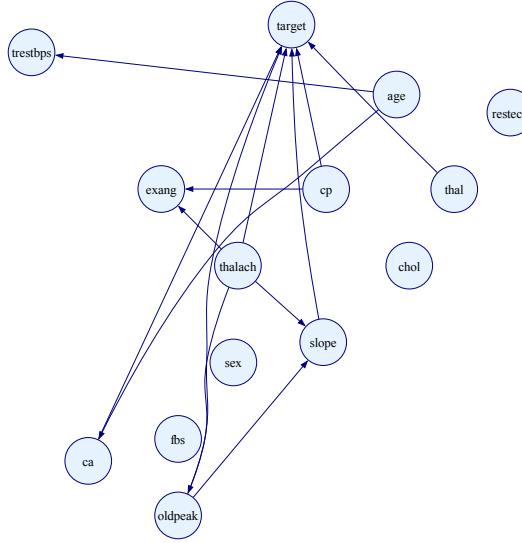


Figure 6: Revised Graph by LLM

5.3 Graph Reliability Analysis

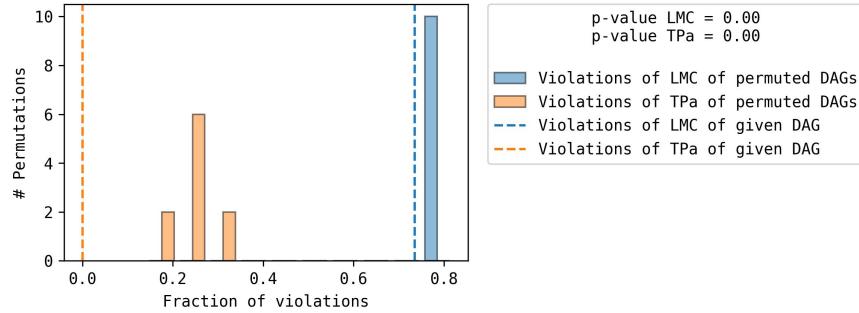


Figure 7: Refutation Graph

The results of the graph refutation test indicate that the given Directed Acyclic Graph (DAG) demonstrates strong informativeness, as evidenced by the absence of any permutations falling within its Markov equivalence class (0/10 with a p-value of 0.00). However, it also shows significant violations of the Local Markov Conditions (LMCs), with 114 out of 155 LMCs being violated. Despite this, the DAG outperforms 100% of the permuted graphs, further supporting its validity with a p-value of 0.00. Given this context and the specified significance level of 0.05, we conclude that the DAG remains reliable and valid, as we do not reject it based on these compelling results. The findings suggest that the DAG's structure captures essential causal relationships that are not solely attributable to random variation in the data.

5.4 Result Graph Comparision

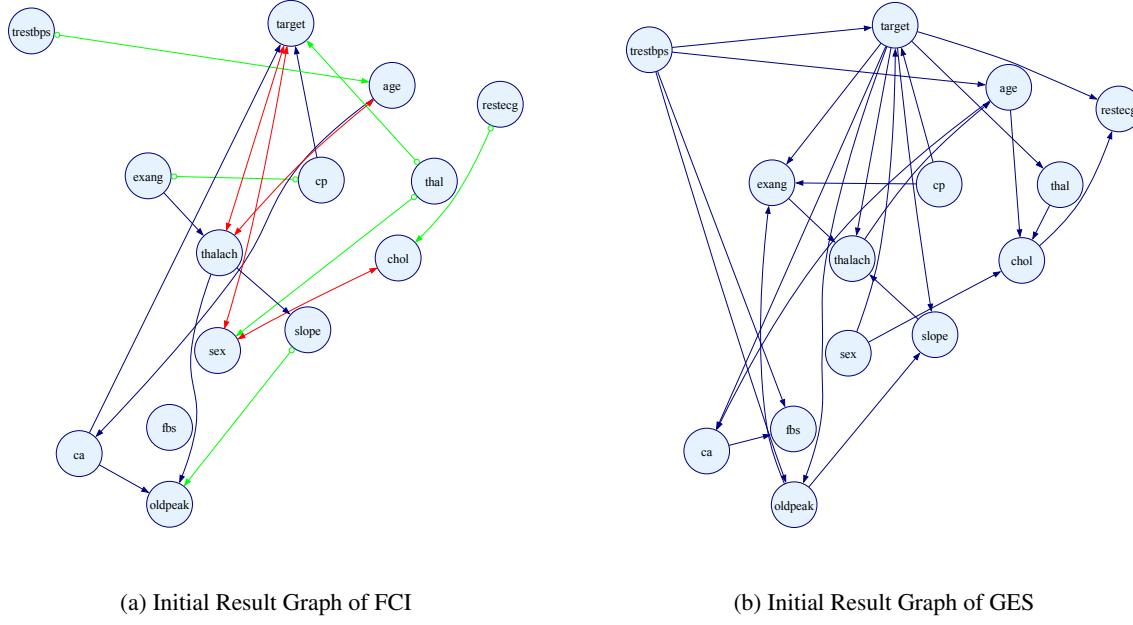


Figure 8: Result Graph Comparision of Different Algorithms

Comparing the causal discovery results from the FCI and GES algorithms reveals several unique edges, as well as commonalities between the two approaches.

For the edges identified in the FCI algorithm, notable directed relationships include “exang causes thalach,” “thalach causes oldpeak,” “ca causes oldpeak,” and “cp causes target.” In contrast, the GES algorithm identifies a different set of directed edges, such as “trestbps causes age,” “thalach causes age,” and “sex causes chol.” The two algorithms do not share many directed edges, which suggests their approaches may capture different aspects of the underlying causal relationships in the data.

Both algorithms have certain directed edges that are absent in the other results. For instance, FCI indicates edges linking “age to ca” and “cp to target,” while GES captures relationships such as “target to restecg” and “oldpeak to exang.” The differences suggest that different assumptions or methodologies about causation influence the resulting graphs.

When focusing on the common edges, one noteworthy shared relationship is “ca causes target,” which suggests a consistent inference about the influence of ‘ca’ on ‘target’ across both algorithms. Despite variations in the overall structure of the causal graphs, identifying this shared edge lends credibility to the causal influence specified.

Regarding the reliability of edges, those consistently recognized across multiple algorithms and those supported by domain knowledge or theoretical justification generally hold more weight. In this case, the edge “ca causes target” exemplifies a reliable causal relationship since it appears in both causal graphs. In contrast, edges that are unique to a single algorithm may require further validation or could reflect artifacts of the specific assumptions made by that method, such as potential hidden confounders addressed differently in GES compared to FCI. Thus, edges substantiated by multiple algorithms and corroborated by theoretical or empirical backing are the most trustworthy in indicating causal relationships.

5.5 Conclusion

In this report, we analyzed a comprehensive dataset encompassing various variables linked to heart disease risk factors, such as age, sex, cholesterol levels, and exercise-induced symptoms. Utilizing a robust causal discovery methodology, we employed algorithms including the Peter-Clark (PC) and Fast Causal Inference (FCI) to derive causal relationships from the data, while integrating insights from a large language model (LLM) to optimize our approach through graph

tuning and bootstrap reliability assessments. Our findings revealed intricate causal connections, such as the influence of age and maximum heart rate on coronary artery status and heart disease diagnosis, highlighting significant direct and indirect pathways within cardiovascular health.

Our contribution lies in the identification of nuanced causal relationships impacting heart disease, which not only enhance understanding of the disease's mechanisms but also offer insights into potential intervention strategies. This report emphasizes the importance of personalized treatment approaches, as evidenced by the heterogeneous treatment effects we observed based on demographic factors such as age and sex. By addressing confounding variables and employing advanced techniques, our work aims to inform clinical practice and pave the way for improved prevention strategies in cardiovascular health management. Future research directions will focus on expanding the dataset and employing refined causal inference techniques to further validate and enhance these findings.

6 Causal Inference Results

6.1 Proposal Overview

In this report, we aim to investigate the causal effect of resting blood pressure (`trestbps`) on the binary target variable, which signifies the presence or absence of a specific health condition. The selection of this causal inference task stems from the critical need to understand the relationship between physiological metrics and health outcomes, particularly in a time when cardiovascular health is a leading concern in public health. By estimating the treatment effect of `trestbps` on the target, we can gain valuable insights into whether higher or lower blood pressure readings are associated with increased risk of adverse health conditions, potentially informing prevention strategies and therapeutic interventions.

Utilizing causal inference methodologies will enable us to draw more reliable conclusions than traditional correlational analyses. By controlling for confounding variables and employing appropriate statistical techniques, such as propensity score matching or instrumental variable analysis, we can more accurately estimate the causal impact of `trestbps` on the target variable. This understanding is vital for clinicians and public health policymakers, as it directly influences treatment guidelines and health risk assessments related to blood pressure management. Ultimately, addressing this causal inference query will enhance our comprehension of the physiological underpinnings of health outcomes and guide future research in the domain of cardiovascular health.

6.2 Treatment Effect Estimation Results

6.2.1 Basic Information Overview

- **Treatment Variable:** `trestbps`.
- **Outcome Variable:** `target`.
- **Confounders:** `thalach,age,exang,oldpeak,chol,restecg`.
- **Heterogeneous Variables:** `age,sex`.

6.2.2 Estimation Method & Justification

Chosen Method: `drl`

Justification

Matching methods become inefficient in high-dimensional settings. ML-based methods are better at handling many confounders while reducing bias. DRL combines propensity score modeling (IPW) and outcome regression, making it robust even if one model is misspecified. It's more stable with small sample sizes than DML.

6.2.3 Estimation Results

Average Treatment Effect (ATE) and Average Treatment Effect on the Treated (ATT)

- The analysis shows an Average Treatment Effect (ATE) of approximately 0.059, indicating a positive but small estimated effect of the treatment on the target outcome.
- The confidence interval for the ATE ranges from approximately -0.298 to 0.416, suggesting that the effect could be near zero or even negative, reflecting uncertainty in the estimate.
- The Average Treatment Effect on the Treated (ATT) is estimated at about 0.023, which also signifies a marginal positive influence; however, the confidence interval spans from roughly -0.171 to 0.217, reinforcing the limited certainty regarding the treatment's impact on those who received it.

- Overall, the results point to a potential positive effect of the treatment on the target variable but underscore significant uncertainty, as indicated by the wide confidence intervals for both ATE and ATT. Further investigation or additional data may be needed to draw more definitive conclusions.

Heterogeneous Treatment Effect (HTE)

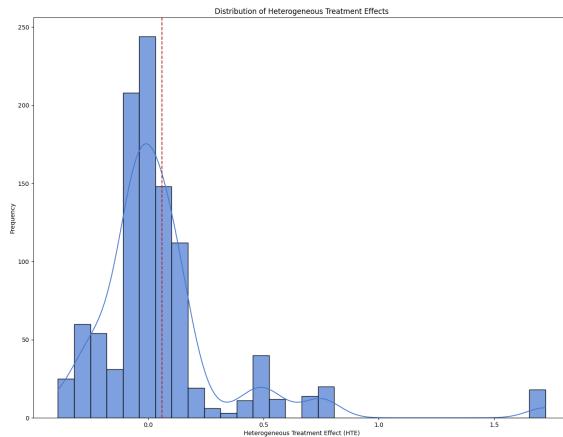


Figure 9: Distribution of HTE

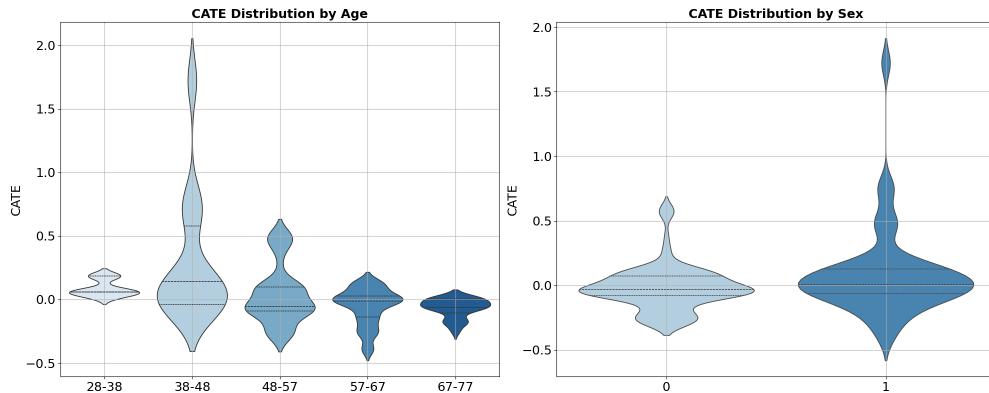


Figure 10: Violin plot of HTE by Heterogeneous Variables

- The analysis reveals significant variability in the treatment effect of trestbps on the target, indicating that the impact is not uniform across the population.
- The distribution of the heterogeneous treatment effects (HTE) suggests that certain subgroups may experience markedly different effects from the treatment, highlighting the importance of tailoring interventions.
- The violin plot illustrates that both age and sex are critical factors influencing the conditional average treatment effect (CATE). These factors reveal distinct patterns, with notable differences observed in the treatment effectiveness across the groups.
- Specifically, younger individuals tend to exhibit a more pronounced response to trestbps compared to older individuals, while the variation in effect based on sex indicates a potential interaction that warrants further investigation.
- Overall, these findings underscore the necessity for personalized approaches in treatment, as the effects of trestbps on the target are significantly influenced by individual characteristics.

6.3 Summary & Next Steps

6.3.1 Discussion

You did not conduct any discussion with causal copilot.

6.3.2 Next Steps Suggestions

To enhance the robustness of the findings, potential improvements to our analysis could involve the integration of larger and more diverse datasets, allowing for a more comprehensive assessment of the treatment effect across various demographic groups. **Incorporating additional covariates or exploring nonlinear relationships may also reveal further nuances in treatment effects, especially in subgroups with stark differences observed in age and sex.** Utilizing advanced causal inference techniques, like hierarchical modeling or machine learning methods, could improve the precision of the Average Treatment Effect (ATE) and the Average Treatment Effect on the Treated (ATT) estimates by accommodating complex interactions among variables.

- For future research directions, it would be valuable to validate these findings through experiments or longitudinal studies that track outcomes over time, which can help establish causal relationships more firmly.
- **Additionally, conducting subgroup analyses based on the identified heterogeneous treatment effects could facilitate a better understanding of individual variation, paving the way for personalized treatment strategies.**
- Exploring the mechanisms behind these treatment effects, particularly considering the influences of age and sex, would further enhance the applicability and effectiveness of the interventions.
- Ultimately, these next steps will not only refine our understanding of the current results but also contribute to the broader causal inference landscape by informing tailored approaches to interventions.