Scholarship for Higher Education (SHE) Component under INSPIRE

Format for Project Completion Certificate

(To be filled by the Supervisor of Research Project)

1. IVR Number or Application Ref. No.: 202000009727
2. Name of INSPIRE Scholar: Aniket Tiwari
3. Name of College where Scholar is studying: Chennai Mathematical Institute
4. Name of University to which above college is affiliated: Chennai Mathematical Institute
5. Bank Account Number of Scholar: 39761421072
6. IFSC Code: SBIN0008314
7. Mobile No. and Email Address of Scholar: 8318138873, aniket9454578232@gmail.com
8. Title of Project: Propensity Score Analysis
9. Broad Subject Area of Research Project: Impact of D-Penicillamine in the treatment of Cirrhosis
10. Project Duration (in weeks): 7 weeks
    (Minimum 6 to 8 weeks)
11. Date of Start of Research Project: 04/12/23
12. Date of Completion of Research Project: 21/01/24
13. Supervisor Name: Dr. Prem Kumar
14. Designation of Supervisor: Assistant Professor
15. Affiliation of Supervisor
    (a) Department - Statistics
    (b) College - Gyanwati Mahavidyalaya, Pithanpura
    (c) University - Jiwaji University, Gwalior
    (d) Address - Syawali, Bhind, Madhya Pradesh
    (e) Mobile no. and Email Address - 8518252980, premnara024@gmail.com
16. Major Specialization area of Supervisor: Descriptive Statistics, Regression Modelling

(Seal & Signature of Supervisor)

Dr. Prem Narayan
Assistant Professor
Department Of Statistics
Gyanwati Mahavidyalaya, Pithanpura

Date: 10/09/24

Acknowledgement

First and foremost, I thank the Almighty for giving me the strength and perseverance to complete this project.

I am deeply grateful to my mentor, Dr. Prem Narayan, for his invaluable guidance and support throughout this project, "Propensity Score Analysis to Study the Impact of D-Penicillamine in the Treatment of Cirrhosis."

I also extend my heartfelt thanks to my parents for their unwavering support and encouragement, and to my friends for their constant encouragement and assistance.

Finally, I appreciate all the supporting personnel who provided the essential equipment and resources necessary for this project's success.

- Aniket Tiwari

Title : Propensity Score Analysis to study the impact of D-Penicillamine in the treatment of Cirrhosis

Abstract

In observational studies, evaluating treatment effects of a drug on a disease presents considerable challenges. The propensity score methodology offers a promising avenue to address such complexities by facilitating the comparison of treated and untreated subjects. This study focuses on investigating the impact of D-Penicillamine on Liver Cirrhosis utilizing propensity score matching techniques.

Through binary logistic regression, propensity scores were computed for 312 patients included in the study. These scores served as balancing scores, ensuring that the distribution of observed baseline covariates was similar between treated and untreated groups. Subsequently, matching was performed between the treatment and control cohorts employing the K-Nearest neighbors (KNN) algorithm.

By leveraging propensity score matching, this research endeavors to mitigate potential confounding biases inherent in observational studies, thereby providing a more rigorous assessment of the treatment effect. Finally, the average treatment effect (ATE) was calculated to quantify the impact of D-Penicillamine on the covariates associated with Liver Cirrhosis.

Keywords: Liver Cirrhosis, Observational Studies, D-Penicillamine, Propensity Score Matching, Average Treatment Effect

CONTENTS

INTRODUCTION

1.1 LIVER CIRRHOSIS

Liver cirrhosis is a significant global health concern, characterized by the replacement of healthy liver tissue with scar tissue (fibrosis), which impairs liver function. This chronic and progressive condition can result from various causes, including alcohol-related liver diseases, viral hepatitis, non-alcoholic fatty liver disease (NAFLD), metabolic syndrome, and autoimmune disorders. As cirrhosis advances, it severely hampers the liver's ability to perform its essential functions, leading to numerous complications.

Some of the serious complications associated with liver cirrhosis include increased pressure in the portal vein (portal hypertension), the buildup of fluid in the abdomen (ascites), gastrointestinal bleeding, confusion and cognitive issues (hepatic encephalopathy), and liver cancer. Common symptoms of cirrhosis include yellowing of the skin and eyes (jaundice), muscle loss, fatigue, loss of appetite, and general weakness. Cirrhosis can cause serious issues like increased pressure in the portal vein (portal hypertension), fluid buildup in the abdomen (ascites), gastrointestinal bleeding, brain dysfunction due to liver failure (hepatic encephalopathy), and liver cancer. Symptoms include yellowing of the skin (jaundice), muscle loss, fatigue, loss of appetite, and weakness. Prolonged liver damage leads to inflammation, causing liver cell death and the migration of inflammatory cells. While the liver tries to regenerate damaged cells, this process triggers scar tissue production by cells like fibroblasts, leading to fibrosis. As fibrosis progresses, it disrupts the liver's structure and function, causing portal hypertension and affecting blood flow within the liver. This increases the risk of liver cancer. Diagnosing cirrhosis involves clinical exams, laboratory tests, imaging studies, liver biopsy, and elastography. Early diagnosis and treatment can slow the disease's progression and reduce complications.

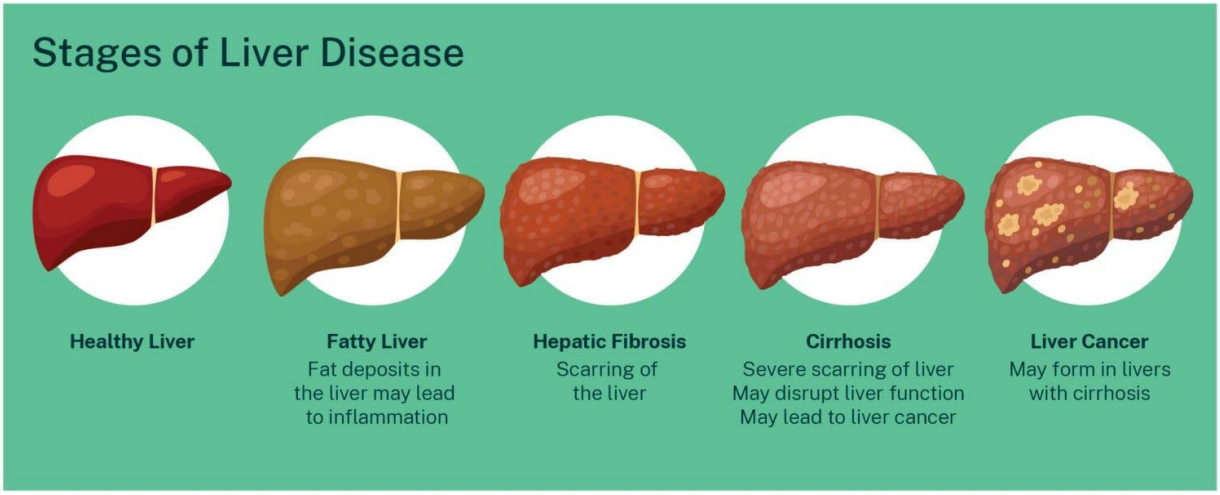

Figure: 1.1 Fig. shows different stages of condition of liver

1.2 OBSERVATIONAL STUDIES

According to Cochran (1965), an observational study is an empirical investigation whose objective is to elucidate causal relationships (i.e., cause and effect) when it is infeasible to use controlled experimentation and to assign participants at random to different procedures. At first, observational study concerns treatment effects. Second, observational studies can employ data from non experimental, non observational studies as long as the focus is on assessing treatment or the effects of receiving a particular service. By this definition, observational data refer to data that were generated by something other than a randomized experiment and typically include surveys, censuses, or administrative records.

Observational studies are ones where researchers observe the effect of a risk factor, diagnostic test, treatment or other intervention without trying to change who is or isn't exposed to it. Cohort studies and case control studies are two types of observational studies.

Cohort study: For research purposes, a cohort is any group of people who are linked in some way. For instance, a birth cohort includes all people born within a given time frame. Researchers compare what happens to members of the cohort that have been exposed to a particular variable to what happens to the other members who have not been exposed.

Case control study: Here researchers identify people with an existing health problem ("cases") and a similar group without the problem ("controls") and then compare them with respect to an exposure or exposures.

1.3 PROPENSITY SCORE

Drawing causal inferences in observational studies is challenging, and it is this task that has motivated statisticians and econometricians to explore new analytic methods. These methods aim to accomplish data balancing when treatment assignment is non-ignorable, to evaluate treatment effects using non-randomized or non-experimental approaches, and /or to reduce multidimensional covariates to a one-dimensional score called a propensity score. Rosenbaum and Rubin's (1983) defined a propensity score as the conditional probability of assignment to a particular treatment given a vector of observed covariates. When applied appropriately, these models can help solve the problem of selection bias and provide valid estimates of average treatment effects (ATEs). Conditioning methods commonly used to estimate the ATE include full matching, stratification, inverse probability of treatment weights (IPTW).

1.3.1 STRATIFICATION

Stratification divides individuals into many groups (or subclasses) on the basis of their propensity score values. It is similar to full matching but creates fewer groupings. The optimal number of strata depends on the sample size and the amount of overlap or common support between the treatment and control groups' propensity scores. However, five subclasses, purported to remove 90% of the bias due to measured confounders, have been used by the majority of propensity score studies based upon recommendations by Cochran (1968) and Rosenbaum and Rubin (1984).

1.3.2 MATCHING

The goal of matching is to obtain similar groups of treatment and control subjects by matching individual observations on their propensity scores. One of the most common matching methods used in propensity score analysis is 1:1 matching which forms pairs of treated and control subjects. Nearest neighbor (NN) or greedy matching selects a control unit for each treated unit based on the smallest distance from that treated unit in PS. The selection process can be done without replacement, i.e., subjects are not returned to the sample after being pair-matched, therefore many of the subjects in the dataset are discarded, reducing power and generalization.

1.3.3 INVERSE PROBABILITY OF TREATMENT WEIGHTS

In inverse probability of treatment weights(IPTW), individuals are weighted by the inverse probability of receiving the treatment that they actually received. Treated individuals receive an IPTW equal to $1/p_i$ and control individuals receive a weight equal to $1/(1-p_i)$. The weights are then used in a weighted least squares (WLS) regression model along with other predictor covariates. The IPTW method is inclusive of all subjects in a study, therefore no loss of sample occurs as in other conditioning methods, i.e., matching, stratification.

METHODS

2.1 DATA DESCRIPTION

Table 2.1 : Data description

ID	Unique identifier
N_Days	Number of days between registration and the earlier of death, transplantation, or study analysis time in July 1986.
Status	Status of the patient C (censored), CL (censored due to liver tx), or D (death).
Drug	Type of drug, D-penicillamine or placebo
Age	Age in [days].
Sex	M (male) or F (female).
Ascites	Presence of ascites is denoted by N (no) or Y (yes). Ascites is the abnormal build-up of fluid in the abdomen, often due to severe liver disease.
Hepatomegaly	Presence of hepatomegaly N (No) or Y (Yes). Hepatomegaly is enlargement of the liver, which indicates dysfunction in the liver.
Spiders	Presence of spiders N (No) or Y (Yes). A spider angioma or spider naevus, also nevus araneus, is a type of telangiectasis (swollen, spider-like blood vessels on the skin) found slightly beneath the skin's surface, often containing a central red spot and deep reddish extensions (see Blood color) which radiate outwards like a spider's web or a spider's legs.
Edema	Presence of Edema N (no edema and no diuretic therapy for edema), S (edema present without diuretics, or edema resolved by diuretics), or Y (edema despite diuretic therapy).
Bilirubin	serum bilirubin in [mg/dl]. Bilirubin is the main pigment in bile, is produced when red blood cells are broken down in the liver.
Cholesterol	serum cholesterol in [mg/dl]. An important function of the liver is to produce and clear cholesterol in the body but cholesterol is necessary

	for the creation of hormones, vitamin D, and enzymes needed for digestion.
Albumin	albumin in [gm/dl]. Albumin is a protein made in the liver. It is a non-specific indicator of the synthetic function of the liver with a long half-life (20 days).
Copper	urine copper in [ug/day]. Copper is an indispensable trace element which serves as a cofactor for enzymes involved in cellular energy metabolism, antioxidant defense, iron transport, and fibrogenesis.
Alk_Phos	alkaline phosphatase in [U/liter]. Alkaline phosphatase (ALP) is a non-specific liver enzyme mainly found in the bile ducts of the liver.
SGOT	SGOT in [U/ml]. An SGOT blood test measures levels of aspartate aminotransferase (AST) and helps determine liver function. Constant increases in SGOT/SGPT levels indicate chronic liver diseases.
Triglycerides	Triglycerides in [mg/dl]. High triglyceride levels can raise your risk for certain health conditions, including stroke, heart attack, and liver disease.
Platelets	platelets per cubic [ml/1000]. Low platelet count (thrombocytopenia) is the most common blood abnormality in chronic liver disease
Prothrombin	prothrombin time in seconds [s]. Prothrombin is a protein made by the liver to help blood clot. Prothrombin Time (PT) measures how long it takes for blood to clot.
Stage	histologic stage of disease (1, 2, 3, or 4)

D-Penicillamine:

D-penicillamine is used for patients with primary biliary cirrhosis due to its hepatic copper decreasing and immunomodulatory potentials. Penicillamine is a medication primarily used for the treatment for Wilson's disease which is a rare genetic disorder of copper metabolism. It's also used for people with kidney stones who have high urine cystine levels, rheumatoid arthritis and various heavy metal poisonings.

Adverse Effects: Common side effects include rash, loss of appetite, nausea, diarrhea, and low blood white blood cell levels, swollen and/or painful glands, joint pain. Other serious side effects include liver problems, obliterative bronchiolitis, and myasthenia gravis.

Component: Penicillamine is a trifunctional organic compound, consisting of a thiol, an amine, and a carboxylic acid. It is an amino acid structurally similar to cysteine, but with geminal dimethyl substituents α to the thiol.

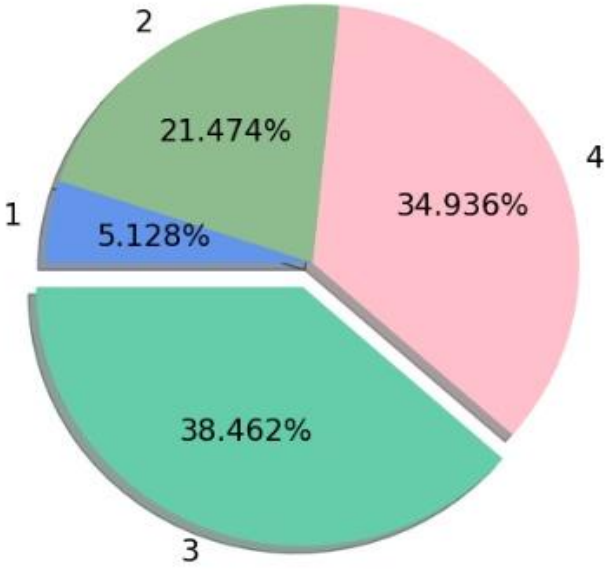

Figure 2.1: Stagewise proportion of cirrhosis

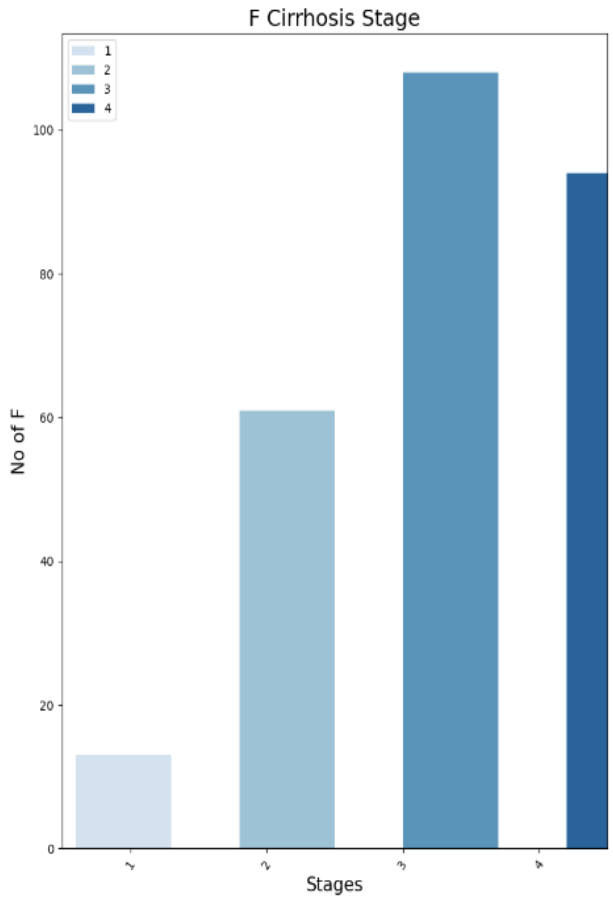

Figure 2.2 : Presence of cirrhosis stagewise in females

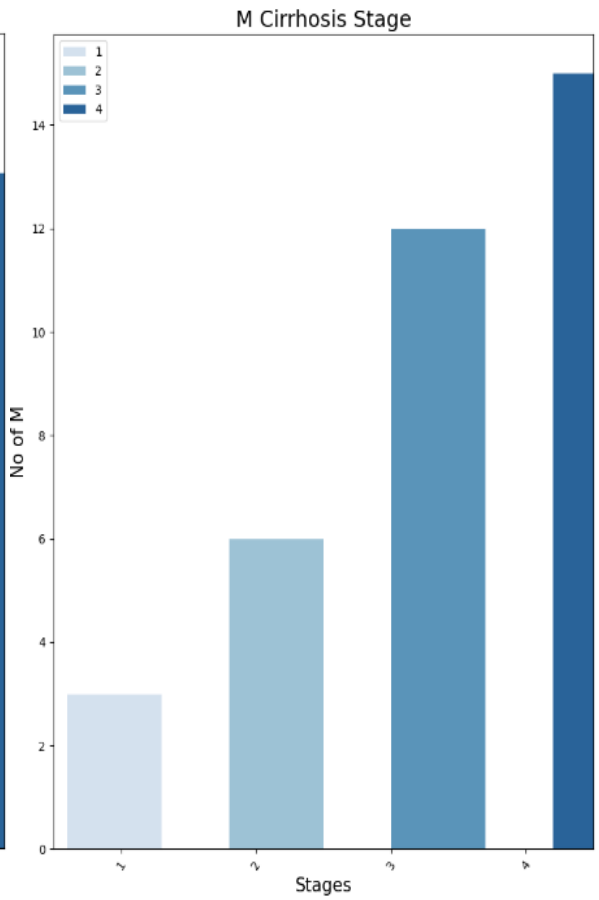

Figure 2.2 : Presence of cirrhosis stagewise in males

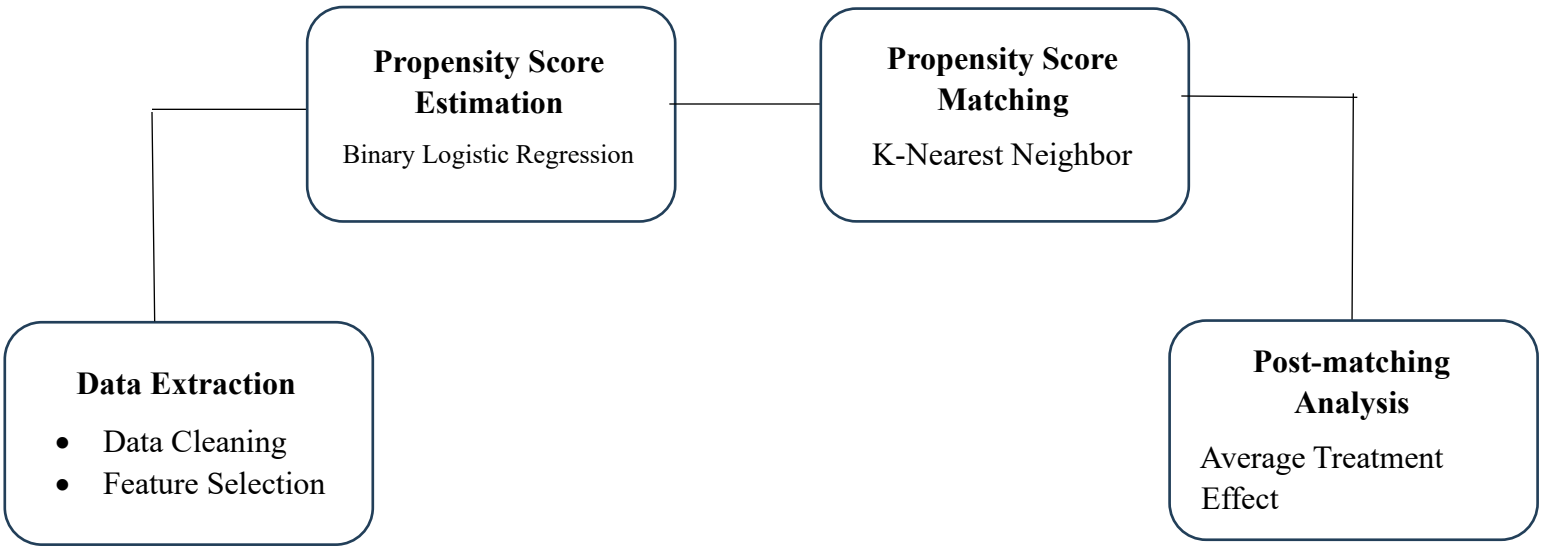

Figure 2.3 : Flow chart shows the general procedure for propensity score matching

2.2 DATA CLEANING

Data cleaning is an important early step in the data analytics process. This crucial exercise, which involves preparing and validating data, usually takes place before core analysis. Data cleaning is not just a case of removing erroneous data, although that's often part of it. The majority of work goes into dealing with missing data.

There are three primary methods to handle missing values in a dataset

1. **Deletion**: This method involves removing any rows or columns that contain missing values.
2. **Imputation**: In this technique, missing values are replaced with substituted values, which can be derived from various statistical methods such as mean, median, mode.
3. **Prediction**: Here, missing values are predicted based on the existing data.

In our dataset, we have four covariates which consist of missing values, namely, 'Cholesterol', 'Copper', 'Triglycerides' & 'Platelets'. All these four covariates are continuous in nature. We opt for the second method as the missing values are meaningful in our analysis.

Table 2.2 : Missing values present in data

Feature	Missing Values
ID	0
N_Days	0
Status	0
Drug	0
Age	0
Sex	0
Ascites	0
Hepatomegaly	0
Spiders	0
Edema	0
Bilirubin	0
Cholesterol	28

Albumin	0
Copper	2
Alk_Phos	0
SGOT	0
Triglycerides	30
Platelets	4
Prothrombin	0
Stage	0

To handle missing values in our dataset, we used different imputation methods based on the extent and nature of the missing data for specific covariates. For 'Copper' (2 missing) and 'Platelets' (4 missing), we replaced missing values with the overall median.

For 'Cholesterol' (28 missing) and 'Triglycerides' (30 missing), we used stagewise medians, calculated separately for each stage of the response variable ('Stage' 1, 2, 3, 4). Missing values were replaced with the median corresponding to the same stage. This approach ensures transparent and stage-specific handling of missing data, potentially leading to more accurate and meaningful analyses.

2.3 FEATURE SELECTION

Feature selection is the process of identifying and selecting a subset of relevant features (variables, predictors) for use in model construction. The main objective of feature selection is to improve the model's performance by eliminating irrelevant or redundant data, which can lead to several benefits: Improved Model Performance, Reduced Overfitting, Faster Computation, Enhanced Interpretability, Reduced Storage and Maintenance Costs. We've employed two distinct methods for feature selection. Ordinal logistic regression was utilized to identify the continuous independent variables that influence the stage of cirrhosis. Additionally, Cohen's Kappa was applied to assess the impact of categorical independent variables.

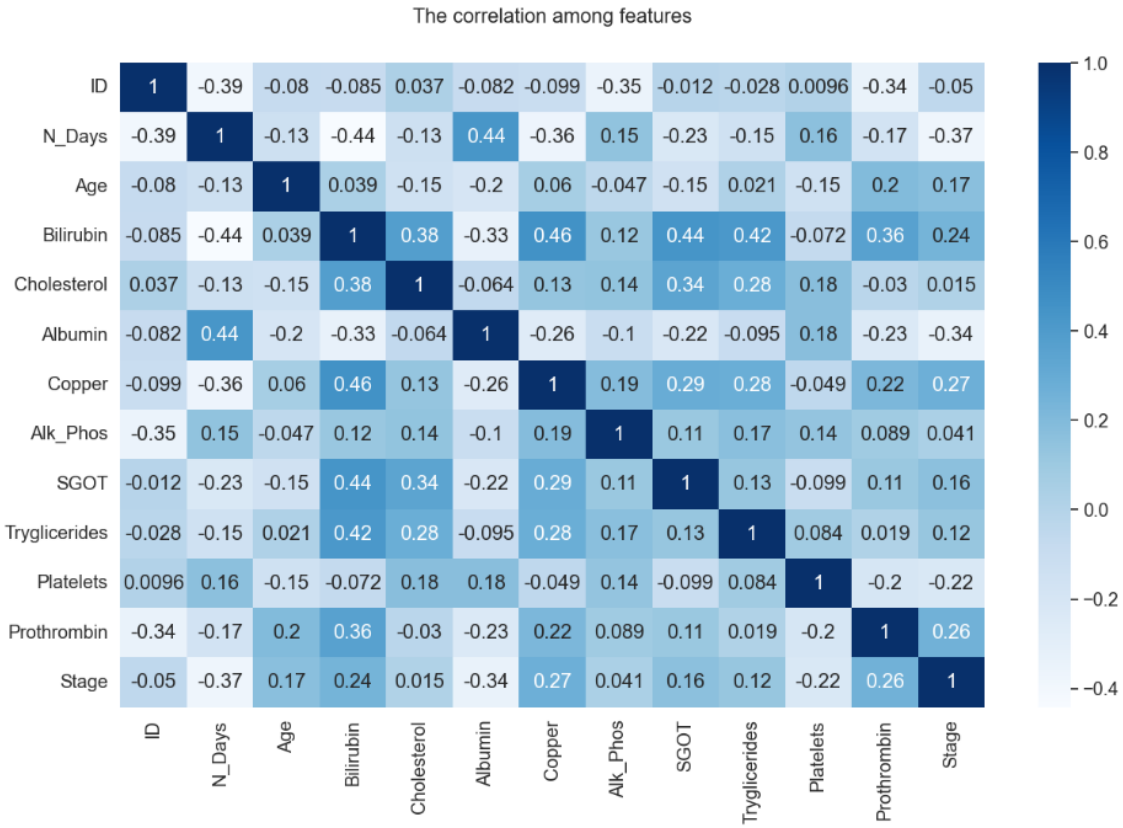

Figure 2.4 : Correlation Heatmap between continuous variables

The correlation coefficients are calculated for all the continuous variables to discard dependent covariates. However, none of the variables are highly correlated. Hence, we proceed with ordinal logistic regression to identify significant variables, i.e., those that impact the stage of cirrhosis.

2.3.1 ORDINAL LOGISTIC REGRESSION

Ordinal Logistic Regression is used when there are three or more categories with a natural ordering to the levels, but the ranking of the levels do not necessarily mean the intervals between them are equal. Ordinal logistic regression is a statistical method used when the dependent variable has ordered categories. It models the relationship between the ordinal dependent variable Y and one or more independent variables $X_1, X_2, \ldots, X_k$. The model assumes that the odds of moving to a higher category versus all lower categories are proportional across different thresholds of the dependent variable.

Mathematically, the cumulative probability $P(Y \leq j)$ of the dependent variable being less than or equal to category j is expressed as:

$$log\frac{P(Y \leq j)}{P(Y > j)} = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \quad (2.1)$$

Where α_j represents the threshold (intercept) for category j, $\beta_1, \beta_2, \ldots, \beta_k$ are the coefficients for the independent variables, and $X_1, X_2, \ldots, X_k$ are the values of those independent variables. The model estimates the log-odds of being in or below a certain category. Parameters are typically estimated using maximum likelihood estimation. Ordinal logistic regression is useful in various fields such as psychology, social sciences, and epidemiology for analyzing ordered categorical outcomes.

Cumulative Probability: When response categories are ordered, the logits can utilize the ordering. You can modify the binary logistic regression model to incorporate the ordinal nature of a dependent variable by defining the probabilities differently. Instead of considering the probability of an individual event, you consider the probabilities of that event and all events that are ordered before it. A cumulative probability for Y is the probability that Y falls at or below a particular point. For outcome category j, the cumulative probability is:

$$P(Y \leq j) = \pi_1 + \pi_2 + \ldots + \pi_j \quad , \quad j = 1,2,\ldots,k \quad (2.2)$$

where π_j is the probability to chose category j.

Link Function: The link function is the function of the probabilities that results in a linear model in the parameters. There are some different link functions are available in the Ordinal Regression procedure: logit, log-log, probit.

Table 2.3 : Link functions

Function	Form	Typical Application
Logit	$\log(Y/1-Y)$	Evenly distributed categories
Log-Log	$\log(-\log(1-Y))$	Higher categories more probable

Probit	$\varphi^{-1}(Y)$	Analysis with explicit normally distributed latent variable.

The table below summarizes the results of the ordinal logistic regression performed in Python. To shortlist statistically significant variables, we examine the p-values corresponding to each covariate. A p-value less than 0.05 indicates statistical significance.

Albumin (p = 0.000), Copper (p = 0.021), Platelets (p = 0.008), and Prothrombin (p = 0.005) have p-values less than 0.05, indicating that these variables significantly affect the odds of progressing to a higher stage of cirrhosis. Other variables, such as Bilirubin, Cholesterol, SGOT, and Triglycerides, do not have statistically significant effects. Therefore, we have excluded these variables from further analysis.

Table 2.4 : Ordinal Model in Python

Dependent Variable	Stage
Model	OrderedModel
Method	Maximum Likelihood
Date	Mon, 15 Apr 2024
Time	05:11:03
No. Observations	312
Df Residuals	299
Df Model	10
Log-Likelihood	-337.96
AIC	701.9
BIC	750.6

Table 2.5 : Summary of Ordinal Logistic Regression in Python

	coef	std err	z	P > |z|	0.025	0.975
Bilirubin	0.0154	0.038	0.406	0.684	-0.059	0.090
Cholesterol	-0.0006	0.001	-1.098	0.272	-0.002	0.000
Albumin	-1.1486	0.307	-3.743	0.000	-1.750	-0.547
Copper	0.0038	0.002	2.313	0.021	0.001	0.007
SGOT	0.0022	0.002	0.959	0.338	-0.002	0.007
Triglycerides	0.0025	0.002	1.183	0.237	-0.002	0.007
Platelets	-0.0032	0.001	-2.638	0.008	-0.006	-0.001
Prothrombin	0.4105	0.146	2.803	0.005	0.123	0.698
½	-2.2998	2.135	-1.077	0.281	-6.484	1.885
2/3	0.7078	0.122	5.793	0.000	0.468	0.947
3/4	0.6911	0.082	8.471	0.000	0.531	0.851

2.3.2 COHEN'S KAPPA STATISTIC

Cohen's kappa (κ) is a statistic that measures inter-rater agreement for categorical items. It is particularly useful for determining the consistency between different raters or methods of classification when categorizing subjects into distinct groups. Here's a step-by-step explanation of how Cohen's kappa is calculated and interpreted.

Step-by-Step Calculation

1. Construct the Contingency Table:

Suppose there are n subjects and g distinct categories for both variables *X* and *Y*. Let f_{ii} represent the frequency of subjects classified in category *i* by *X* and in category *j* by *Y.*

The contingency table looks like this:

Table 2.6 : Contingency table for Cohen's Kappa Statistic

	$Y=1$	$Y=2$	...	$Y=g$
$X=1$	f_{11}	f_{12}	...	f_{1g}
$X=2$	f_{21}	f_{22}	...	f_{2g}
$:$	$:$	$:$	...	$:$
$X=g$	f_{g1}	f_{g2}	...	f_{gg}

2. Calculate the observed proportional agreement (P_0):

The observed agreement is the proportional of times the raters agree, given by the sum of the diagonal elements (where *X* and *Y* agree) divided by the total number of subjects *n*.

$$P_0 = \frac{1}{n}\sum_{i=1}^{g} f_{ii} \tag{2.3}$$

3. Calculate the expected agreement by chance (P_e):

The expected agreement is based on the assumption that the ratings are independent. It's calculated from the marginal totals (row and column sums).

$$P_e = \frac{1}{n^2}\sum_{i=1}^{g} f_{i+}\, f_{+i} \tag{2.4}$$

Here, $\boldsymbol{f_{i+}}$ is the total number of subjects in i^{th} row and $\boldsymbol{f_{+i}}$ is the total number of subjects in j^{th} column.

4. Compute Cohen's Kappa (κ):

Kappa is defined as the observed agreement corrected for chance agreement.

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \tag{2.5}$$

INTERPRETATION OF COHEN'S KAPPA

K=1, perfect agreement

K=0, agreement equivalent to chance

$0 \leq K \leq 1$, agreement less than chance (can occur in rare situations).

Table 2.7 : Kappa Scores of categorical variables

Feature 1	Feature 2	Kappa Statistic
Drug	Ascites	0.0
Drug	Hepatomegaly	0.0
Drug	Spiders	0.0
Drug	Edema	0.0
Ascites	Hepatomegaly	0.08397683
Ascites	Spiders	0.12135176
Ascites	Edema	0.39431818
Hepatomegaly	Spiders	0.26439024
Hepatomegaly	Edema	0.08418254
Spiders	Edema	0.17713547

2.4 PROPENSITY SCORE ANALYSIS

2.4.1 THE PROBLEM OF DIMENSIONALITY AND THE PROPERTIES OF PROPENSITY SCORES

With complete data, Rosenbaum and Rubin (1983) defined the propensity score for participant i (i = 1, . . . , N) as the conditional probability of assignment to a particular treatment (Wi = 1) versus nontreatment (Wi = 0) given a vector of observed covariates, xi :

$$e(x_i) = P(W_i = 1|X_{i=}x_i) \tag{2.6}$$

The advantage of the propensity score in matching, stratification, and weighting is its reduction of dimensions: The vector X may include many covariates, which represent many dimensions, and the propensity approach reduces all this dimensionality to a one-dimensional score.

1. Propensity scores balance observed differences between treated and control participants in the sample. Rosenbaum (2002b, p. 298) showed that a treated and control participant with the same value of the propensity score have the same distribution of the observed covariate X. This means

that in a stratum or matched set that is homogeneous on the propensity score, treated and control participants may have differing values for X, but the differences will be chance differences rather than systematic differences.

2. Treatment assignment and the observed covariates are conditionally independent given the propensity score; that is,

$$x_i \perp w_i | e(x_i) \tag{2.7}$$

In other words, conditional on the propensity score, the covariates may be considered independent of assignment to treatment. Therefore, for observations with the same propensity score, the distribution of covariates should be the same across the treated and control groups. Furthermore, this property means that, conditional on the propensity score, each participant has the same probability of assignment to treatment, as in a randomized experiment.

3. If the strongly ignorable treatment assignment assumption holds and e(xi) is a balancing score, then the expected difference in observed responses to the two treatment conditions at e(xi) is equal to the ATE at e(xi). This property links the propensity score model to the counterfactual framework and shows how the problem of not observing outcomes for the treated participants under the control condition can be resolved. It follows that the mean difference of the outcome variable between treated and control participants for all units with the same value of the propensity score is an unbiased estimate of the ATE at that propensity score. That is,

$$E[E(Y_1|e(X_i), W_i = 1) - E(E(Y_0|e(X_i), W_i = 0) = E[Y_1 - Y_0|e(X_i)] \tag{2.8}$$

4. Rosenbaum and Rubin (1983, p. 46) derived corollaries to justify key approaches using the propensity scores.

Pair matching: The expected difference in responses of treatment and control units in a matched pair with same value of propensity score e(x) equals the ATE at e(x), and the mean of matched pair differences obtained by this two-step sampling process is unbiased for the ATE

$$\tau = E(Y_1|W_i = 1) - E(Y_0|W_i = 0) = E[Y_1 - Y_0|e(X_i)] \tag{2.9}$$

2.4.2 ESTIMATION OF PROPENSITY SCORE

Several methods for estimating the conditional probability of receiving treatment using a vector of observed covariates are available. These methods include logistic regression, the probit model, and discriminant analysis. Of these methods, logistic regression is the most prevailing approach.

Binary Logistic Regression

The conditional probability of receiving treatment when there are two treatment conditions (i.e., treatment vs. control) is estimated using binary logistic regression. Denoting the binary treatment condition as Wi (Wi = 1, if a study case is in the treatment condition, and Wi = 0, if the case is in the control condition) for the ith case (i = 1, . . . , N), the vector of conditioning variables as Xi , and the vector of regression parameters as , a binary logistic regression depicts the conditional probability of receiving treatment as follows:

$$P(W_i|X_I = x_i) = E(W_i) = \frac{e^{x_i\beta_i}}{1+e^{x_i\beta_i}} = \frac{1}{1+e^{-x_i\beta_i}} \quad (2.10)$$

This is a nonlinear model, meaning that the dependent variable Wi is not a linear function of the vector of conditioning variables x_i . However, by using an appropriate link function such as a logit function, we can express the model as a generalized linear model (McCullagh & Nelder, 1989). Although W_i is not a linear function of x_i , its transformed variable through the logit function (i.e., the natural logarithm of odds or $\log_e\{P(W_i)/[1 - P(W_i)]\}$) becomes a linear function of x_i :

$$\log_e\left(\frac{P}{1-P}\right) = X_i\beta_i \quad (2.11)$$

where *P* denotes $P(W_i)$.

2.4.3 MATCHING

After propensity scores are estimated, the next step of analysis often entails matching treated to control participants based on the estimated propensity scores. The core idea of matching, after obtaining estimated propensity scores, is to create a new sample of cases that share approximately similar likelihoods of being assigned to the treatment condition. Perhaps the most common matching algorithm is the so-called greedy matching. It includes Mahalanobis metric matching,

Mahalanobis metric matching with propensity scores, nearest neighbor matching, caliper matching, nearest neighbor matching within a caliper, and nearest available Mahalanobis metric matching within a caliper defined by the propensity score. All methods are called greedy matching.

Propensity matching has multiple strengths. First, because propensity matching requires only the single propensity score to select a comparison group, the sample size limitations associated with covariate matching are avoided. Second, the independence of the propensity score from the outcome variable of interest makes it a good proxy for random assignment. Third, research shows that estimated propensity scores are better than the true propensity score at removing bias.

Nearest Neighbor Matching

P_i and P_j are the propensity scores for treated and control participants, respectively; I_1 is the set of treated participants; and I_0 is the set of control participants. A neighborhood $C(P_i)$ contains a control participant j (i.e., $j \in I_0$) as a match for treated participant i (i.e., $i \in I_1$), if the absolute difference of propensity scores is the smallest among all possible pairs of propensity scores between i and j, as

$$C(P_i) = min\left\|P_i - P_j\right\| \ , \quad j \boldsymbol{\epsilon} \boldsymbol{I_0} \tag{2.12}$$

Once a j is found to match to i, j is removed from I_0 without replacement. If for each i there is only a single j found to fall into $C(P_i)$, then the matching is nearest neighbor pair matching or 1-to-1 matching. If for each i the analyst defines n participants who fall into $C(P_i)$ as matches, then the matching is 1-to-n matching.

2.4.4 POSTMATCHING ANALYSIS

In the context of Propensity Score Matching (PSM), effect size refers to the estimated impact of the treatment on the outcome variable. This effect size is calculated by comparing the outcomes of the treated and control groups after adjusting for differences in covariates that predict treatment assignment. The primary goal is to isolate the effect of the treatment itself, thereby reducing the bias that may arise from confounding variables. By accounting for these covariate differences, PSM aims to mimic the conditions of a randomized controlled trial, providing a more accurate

estimation of the treatment's true effect. The importance of effect size in Propensity Score Matching (PSM) lies in its ability to reduce bias, enhance causal inference.

By matching treated and control subjects with similar propensity scores, reducing bias from confounding variables. This process enables a clearer understanding of the causal impact of the treatment, as the effect size represents the difference in outcomes between comparable individuals who differ only in their treatment status.

There are three types of effect sizes:

1. Average Treatment Effect (ATE): It is the average effect of the treatment across the entire population. It shows how much, on average, the outcome would differ if everyone in the population were treated versus if no one were treated.

2. Average Treatment effect on the Treated (ATT): It is the average effect of the treatment on those who actually received the treatment. It shows, how much, on average, the outcome differs for the treated group compared to what their outcome would have been without the treatment.

3. Average Treatment effect on the Controls (ATC): It is the average effect of the treatment on those who did not receive the treatment. It shows how much, on average, the outcome would differ for the control group if they had received the treatment compared to their actual outcome without the treatment.

RESULTS

We performed binary logistic regression using the variables ID, N_Days, Age, Albumin, Copper, Platelets, Prothrombin, Sex, Ascites, Hepatomegaly, Spiders, and Edema as predictors, and Drug as the outcome variable, to compute the propensity score. The propensity score represents the estimated probability that a patient will be prescribed D-Penicillamine, based on the observed covariates.

To account for the categorical nature of Sex, Ascites, Hepatomegaly, Spiders, and Edema, we introduced dummy variables. This transformation allowed us to include these categorical predictors in the logistic regression model appropriately.

Table 3.1 : Propensity Score and Propensity Logit

ID	N_ Days	Age	Albumin	Copper	Platelets	…	Drug	Sex	Propensity Score	Propensity Logit
43	4556	17850	3.64	36.0	203.0	…	D-Penicillamine	F	0.425154	-0.30165
52	2386	18460	3.70	158.0	363.0	…	D-Penicillamine	M	0.491873	-0.03250
53	1000	24621	3.10	94.0	214.0	…	D-Penicillamine	F	0.673086	0.72217
81	2540	23107	3.65	34.0	385.0	…	D-Penicillamine	F	0.579954	0.32258
94	750	19693	3.11	178.0	188.0	…	D-Penicillamine	F	0.458014	-0.16834
:	:	:	:	:	:	…	:	:	:	:
312	788	12109	3.79	186.0	200.0	…	Placebo	F	0.464156	-0.14362
304	1230	12979	3.93	22.0	246.0	…	D-Penicillamine	F	0.274639	-0.97121
:	:	:	:	:	:	…	:	:	:	:
311	839	13879	3.16	69.0	335.0	…	D-Penicillamine	F	0.521098	0.08444

The resulting propensity scores reflect the likelihood of each patient receiving the drug D-Penicillamine, given their specific combination of covariates.

We performed 1-Nearest Neighbor matching using propensity logit values for both the treatment and control groups to conduct propensity score matching as Rosenbaum and Rubin (1985) suggested using the logit of the predicted probability as a propensity score i.e. equation

$$\boldsymbol{q\hat{}(x) = \log[(1 - e\hat{}(x))/e\hat{}(x)]} \qquad (3.1)$$

because the distribution of $q\hat{}(x)$ approximates to normal. Note that in the literature, the quantity $q\hat{}(x)$ is also called an estimated propensity score, although $q\hat{}(x)$ differs from $e\hat{}(x)$ as given by the previous equation. This results in the following matched pairs.

Table 3.2 : Matched ID's of patients

ID	MATCHED ID
16	247
21	50
26	213
29	186
30	123
:	:
306	207
307	147
312	116

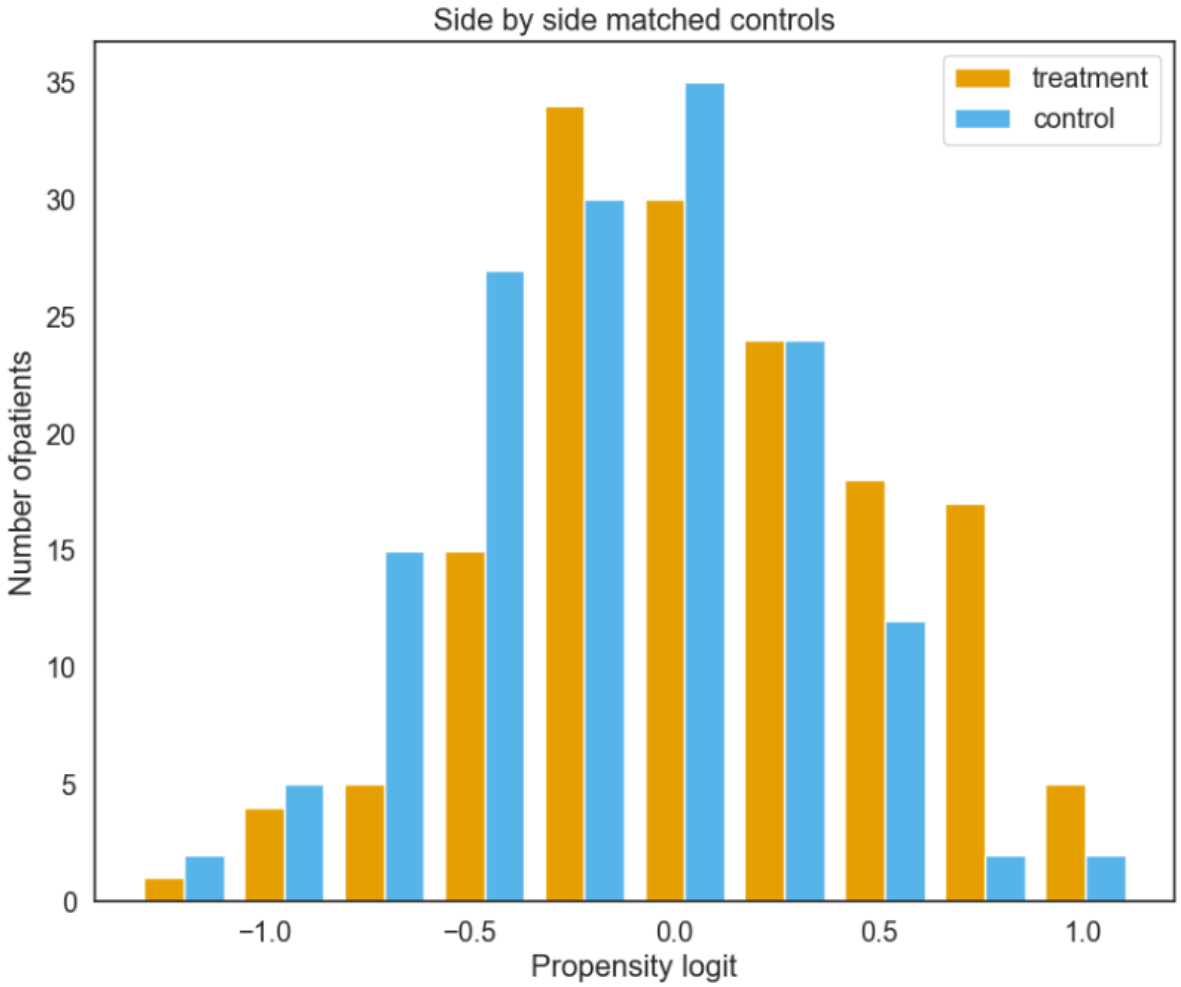

Figure 3.1 Number of patients in each group based on Propensity Logit

The graph above displays the number of patients in Treatment(D-Penicillamine) and Control(Placebo) group based on the values of Propensity Logit.

During the matching process, not all treated units were able to find a corresponding control unit with a similar propensity score, and vice versa. Consequently, in our obtained output, 4 units out of did not have any matches and will therefore be excluded from the analysis.

DISCUSSION

The impetus for developing propensity scores and matching techniques arises from the inherent imbalance often present in observational data, which prevents us from assuming that treatment assignment is ignorable. By matching on the estimated propensity logit, we achieve a balanced sample in terms of observed covariates between treated and control participants. This balance allows us to estimate the treatment effect using the matched sample.

Table 4.1 : Before and After effect size

Variable	Before	After
N_Days	0.016672	0.018487
Age	0.270088	0.235156
Albumin	0.017989	0.011533
Copper	0.000076	0.001976
Platelets	0.065092	0.052940
Prothrombin	0.146361	0.138579
Sex	0.110958	0.104630

Variable	Before	After
Ascites	0.088640	0.075900
Hepatomegaly	0.206322	0.202226
Spiders	0.016277	0.010163
Edema(N)	0.041668	0.056042
Edema(S)	0.578707	0.068744
Edema(Y)	0.006690	0.001714

The effect size table provides the standardized mean differences (SMDs) for various covariates before and after Propensity Score Matching (PSM).

As displayed in Table 4.1, the SMDs for the covariates Age, Albumin, Platelets, Prothrombin, Sex, Ascites, Hepatomegaly, Spiders, and Edema(Y) have decreased post-matching. However, the reductions are relatively marginal, indicating that PSM has not had a profound effect on balancing these covariates.

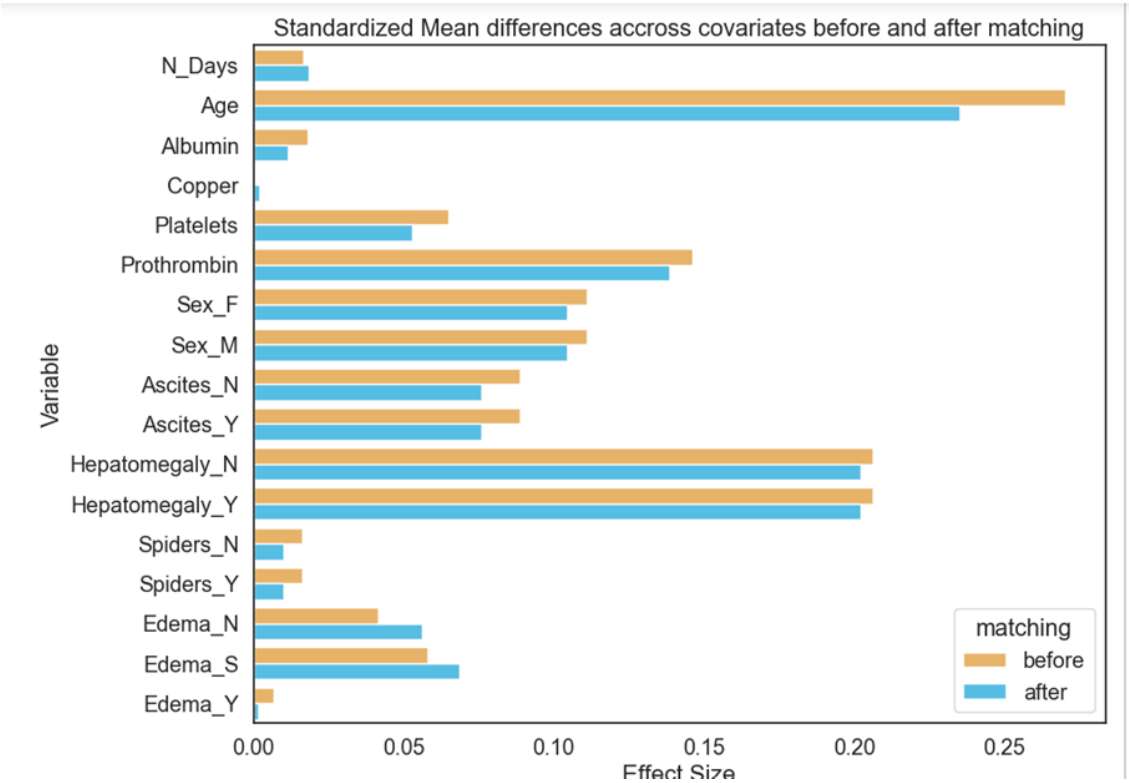

Figure 4.1 : Average treatment effect before and after matching

We aimed to study the impact of D-Penicillamine on Cirrhosis. Ideally, for the drug to demonstrate a meaningful impact, the Average Treatment Effect (ATE) should have been significantly high.

However, the results suggest that while there is a trend towards a reduction in cirrhosis severity due to the treatment, the evidence is not statistically significant at the conventional 0.05 significance level. Specifically, the ATE estimate of -0.171, with a p-value of 0.188 and a 95% confidence interval ranging from -0.426 to 0.084, indicates that the observed reduction is not statistically robust. Similarly, the effects observed in the control group (ATC) and the treated group (ATT) also show non-significant reductions.

Table 4.2 : Outputs of average treatment effect in python

	EST.	S.E.	Z	P>|Z|	[95% CONF. INT]
ATE	-0.171	0.130	-1.316	0.188	(-0.426,0.084)
ATC	-0.238	0.142	-1.677	0.094	(-0.517,0.040)
ATT	-0.106	0.152	-0.698	0.485	(-0.404,0.192)

LIMITATIONS

Propensity score matching (PSM) is a widely used statistical technique for estimating the effect of a treatment or intervention in observational studies. Despite its popularity, it has several limitations that researchers must be aware of. According to Rubin (1997), the primary limitations of PSM include that it works better in larger samples. In smaller samples, it may be challenging to find appropriate matches for each treated unit, leading to imbalances between the treatment and control groups. Larger samples provide a greater pool of potential matches, increasing the likelihood of finding good matches and improving the quality of the causal inference.

Matching at times may lead to significant data loss, especially if matches are not found for many treated or control units, reducing the sample size and statistical power.

Additionally, the high dimensionality of medical data, including various biochemical markers and clinical features, at times complicates the matching process making it difficult to find exact matches.

PSM results are specific to the matched sample and may not generalize to the broader population, limiting the applicability of findings. These limitations necessitate careful consideration and possibly the use of additional methods, such as sensitivity analysis, to ensure robust and reliable conclusions.

CONCLUSION

The findings of this study indicate that D-Penicillamine exhibits minimal to no efficacy in the treatment of cirrhosis. Nonetheless, this investigation highlights the utility of propensity score analysis in evaluating the effects of a drug within a clinical trial context. When appropriately implemented, this method facilitates a more efficient and robust examination of treatment effects, provided that rigorous steps and methodologies are carefully followed.

REFERENCES

1. Propensity Score Analysis, Statistical Methods and Applications by Shenyang Guo, Mark W. Fraser

2. Assessing liver function and interpreting liver blood tests
https://www.sps.nhs.uk/articles/assessing-liver-function-and-interpreting-liver-blood-tests

3. Penicillamine
https://en.wikipedia.org/wiki/Penicillamine

4. D-penicillamine for primary biliary cirrhosis
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8846335

5. Penicillamine Side Effects
https://www.drugs.com/sfx/penicillamine-side-effects.html

6. A Complete Guide to Causal Inference in Python
https://analyticsindiamag.com/a-complete-guide-to-causal-inference-in-python/

7. Detecting Individuals at High Risk of Liver Cirrhosis with Artificial Intelligence Deep Learning Algorithms
https://www.clinicsearchonline.org/article/detecting-individuals-at-high-risk-of-liver-cirrhosis-with-artificial-intelligence-deep-learning-algorithms

8. Cohen's Kappa Statistic for Measuring Agreement
https://online.stat.psu.edu/stat509/lesson/18/18.7

9. Interrater reliability: the kappa statistic
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052

10. Research and scholarly methods: Propensity scores
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9980423

11. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3144483

12. psmpy 0.3.13 documentation
https://pypi.org/project/psmpy

13. Ordinal Logistic Regression
https://bookdown.org/chua/ber642_advanced_regression/ordinal-logistic-regression.html

Declaration

Declaration by the Scholar:

I Aniket Tiwari hereby declare that the details/facts mentioned above are true to the best of my knowledge and I solely be held responsible in case of any discrepancies found in the details mentioned above.

Date: 10/09/2024 (Signature of Scholar)

Aniket

Place: Bhind