HRP 203 Module 3 Project

Harry Koos - 06/07/2024

Link to GitHub Repository: https://github.com/harrykoos/HRP_203_Module_3_Project

Introduction

Heart disease is considered the leading cause of death in the United States and has been for many years. Heart attacks, also known as acute myocardial infarctions (AMIs), are one of the primary acute causes of death for people with heart disease. There is significant debate regarding how patients should be treated when they initially experience an AMI. Less intensive treatments, including drugs like Aspirin, Beta Blockers, and Statins are less costly, but may not have the same effectiveness as more intensive treatments like cardiac catheterization.

In order to determine the ideal course of treatment, it is critical to understand the marginal benefits and marginal costs of using cardiac catheterization as opposed to other forms of treatment. While many studies have aimed to assess the marginal benefits of this treatment, few have assessed the marginal costs.

In this study, we attempt to estimate the net marginal costs of receiving cardiac catheterization. The key challenge with this task is that the patients who get this more intensive treatment likely differ from those who do not in ways that may also affect the total cost of care. For example, AMI patients who are smokers may be more likely to receive cardiac catheterization, but also more likley to have longer hospitals stays regardless of whether they recieve this procedure.

To address this concern, we focus on a sample of AMI patients who have the same health plan and received care from the same health system. We also calculate propensity scores for cardiac catheterization based on observable patient characteristics and use these scores to identify for each treated patient the "nearest neighbor" untreated patient and compare their costs. While this method is not as robust as a randomized controlled trial, it can significantly reduce the bias coming from a raw comparison of treated versus untreated groups.

Methods

Data

The data for this analysis comes from an anonymous health system located very close to Stanford University in Stanford, California. We observe a random sample of 5,000 patients who all had the same private health plan and experienced a heart attack within the month of January 2024. For each patient we observe their age (ranging 18 - 65), gender (male / female), whether they were a smoker (yes / no), whether they received cardiac catheterization (yes / no), and the total cost of the care they received in the three months following the heart attack. We did not receive IRB approval to conduct this study, but decided to precede anyways.

Statistical Analysis

The goals of this analysis is to estimate a less-biased average marginal cost for treating AMI patients with cardiac catheterization. We begin by showing the differences in patient characteristics and costs between the patients who recevied this treatment and those who did not. We perform a t-test to compare the raw difference in mean costs between these populations.

Next, we fit the following logistic regression model on the full sample of patients:

$$Cardiac_i = \beta_1 Smoker_i + \beta_2 Female_i + \beta_3 Age_i + error_i$$

Cardiac is a binary indicator = 1 if patient (i) received cardiac catheterization and = 0 otherwise. Smoker and Female are binary indicators for whether patient (i) was a smoker or female, respectively. Finally, Age is the patient's age in years at the time of the AMI.

We use this model to estimate each patient's propensity score, or probability of receiving the treatment based on their age, gender, and smoking status. With these scores in hand, for each patient that received the treatment, we identify the untreated patient with the closest propensity score to theirs in absolute terms. We then evaluate the extent to which this matching process balanced the observable patient characteristics between treated and untreated groups. Finally, we plot the distribution of costs among the matched groups and perform another t-test to compare the difference in mean costs.

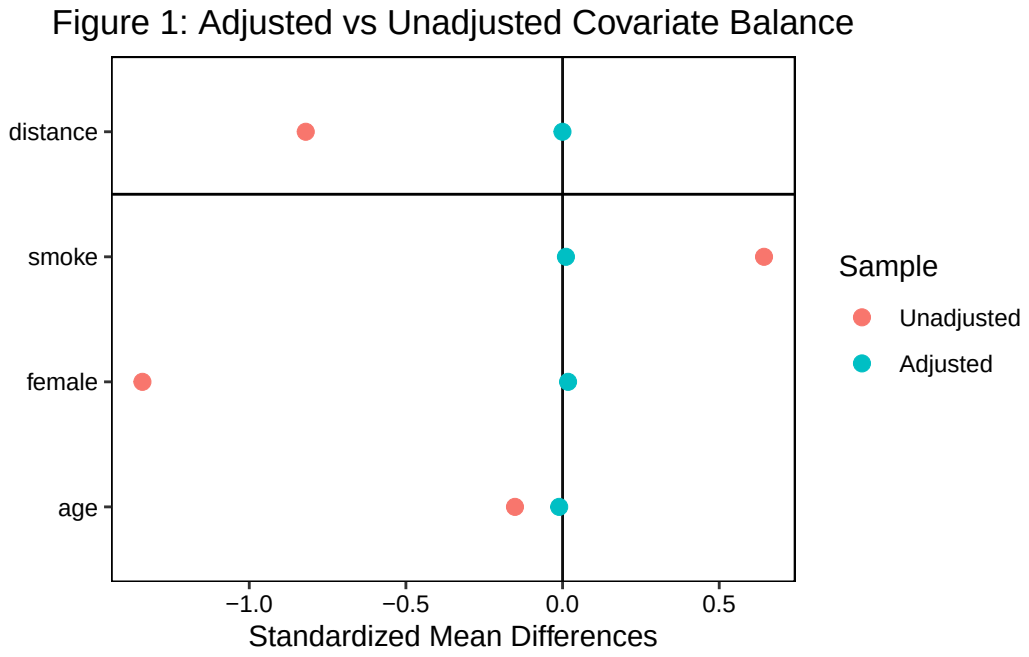
Results

A summary of the patient characteristics and costs for our full patient sample and the treated versus untreated subsamples is presented in Table 1. Only 190 patients received cardiac catheterization in our sample and these patients were disproportionately male and more likely to smoke. These patients were also slightly younger, but not by a large margin. In our t-test of the difference in mean costs, we find a statistically significant (P-value <0.001) difference of $556.67 (95% CI = [492.42, 620.92]) between the treated and untreated groups.

Table 1: Patient Characteristics

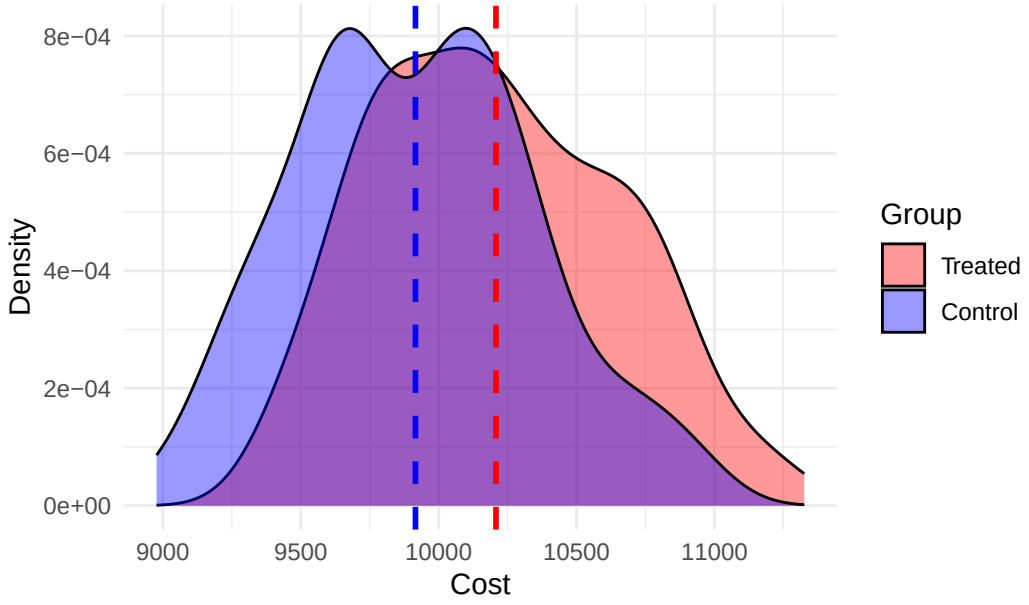
Data Subset	Full Sample	Cardiac Catheterization	No Catheterization
Patients	5000	190	4810
% Smokers	10	41	9
% Female	49	10	50
Mean Age (SD)	41.47 (13.54)	39.47 (13.66)	41.55 (13.53)
Mean Cost (SD)	9672.27 (402.63)	10207.79 (442.53)	9651.12 (386.06)

Figure 1 plots the standardized mean differences for the observable patient characteristics and propensity scores (distance) for the matched samples (adjusted) and the full samples (unadjusted). Consistent with the data presented in Table 1, the unadjusted differences are quite stark. However, the adjusted sample has negligible differences between the two groups, suggesting the matching process worked well.



Next, Figure 2 plots the distribution of treatments costs between propensity matched groups of patients who recevied cardiac catheterization (Treated) and those who did not (Control). The plot shows that there is significant overlap between the two groups, suggesting receiving the treatment does not necessarily imply the patient's costs will be hgiher. However, the treated distribution is notably shifted to the right of the control distribution. The dotted lines denote the mean costs of each distribution.

Figure 2: Density Plot of Cost by Treatment Group



Finally, we performed a t-test on the difference in mean costs between the matched treated and control groups. The result is a statistically significant (P-Value < 0.001) mean difference of \$288.17 with a 95% confidence interval of (199.33, 377.00). While cardiac catheterization is clearly associated with higher expenses than alternative forms of treatment, this estimate is considerably smaller than the difference identified prior to controlling for observeable patient characteristics via propensity score matching.