Empirical Project 1

Tillman Degens

Coauthored by Joey and Devin

4/29/2021

Contents

1 Abstract

It is extremely rare to be able to conduct a randomized experiment on the division of medical insurance and its impact on health care use and outcome. But, in March 2008, the state of Oregon did just that. In the Oregon Health Plan Experiment, 6387 individuals were selected to receive Medicaid in a unique lottery. Researchers were able to conduct interviews 25 months after the lottery occurred to determine its impact on lottery winner's health and usage of health care. This project aims to conduct an exploratory analysis on a potential selection bias for the treatment and control group, as well as provide a basic summary of the impact of the program on important health outcomes. The results indicate that for the most part participants were randomly assigned to the treatment and control group. They also indicate that the policy had a larger impact in increasing health care usage (such as doctors' visits, or prescriptions taken) compared to health outcomes. This study was significant due to the policy debate around public health care and its impact on the economy.

2 Questions to answer

2.1 Question 1

> Explain the difference between the variables treatment and ohp_all_ever_survey. Explain why treatment is the treatment variable (Di), rather than ohp_all_ever_survey.

The difference between the treatment and ohp_all_ever_survey is that treatment is individuals that were entered into medicaid specifically through the lottery, whereas ohp_all_ever_survey are those that have been enrolled in medicaid independent of the lottery.

```
#Load packages and libraries
library(pacman)
library(kableExtra)
```

```
## Warning: package 'kableExtra' was built under R version 4.0.5
```

```
p_load(readr,dplyr, tidyverse,
       ggplot2, skimr, haven, stargazer,
       tidymodels, skimr, janitor, magrittr,
       datasets, rpart.plot, baguette, glmnet,
       tune, haven, ranger, data.table, parallel,
       sandwich, modelsummary,  huxtable, hrbrthemes)
```

2.2 Question 2

> Provide evidence that the OHP lottery really did randomly assign individuals to treatment and control groups. Similar to Table 1 in Taubman et al. (2014), please create a nicely formatted table that reports means of 4 to 6 relevant characteristics for individuals in the control group. Note: Part of this question is to get you to think about which variables should be balanced in a randomized experiment. You need to read carefully through all the variables in the dataset (documentation attached at the end of this file) and decide which 4 to 6 you will summarize.

I selected Gender, Education, Race (non-white), Number of Doctors Visits, Medicine, and Cholesterol. I selected these variables because I believe that they all explain how an individual will interact with the health care system and their health outcome. Age was selected because it impacts the price of private health insurance as well as probability of an underlying condition. Gender was selected as it changes the distribution of health conditions (i.e. breast cancer). Race and education were chosen as societal factors and systematic discrimination have been proven to influence how individuals use health care.

2.3 Question 3

> For each of the variables you summarized above, calculate: (i) the difference between the mean in the treatment group and the mean in the control group. (ii) the standard error for the difference in means. Add these as columns two and three to the table you started in question 2.

```
setwd("C:/Users/tillm/OneDrive/Desktop/525")
#load data from downloads since Rconsole cannot download ".dta" files
ohp_data <- read_dta("ohp.dta")
```

```r
Gender = lm(gender_inp ~ treatment, data = ohp_data)
Age = lm(age_inp ~ treatment, data = ohp_data)
Not_White = lm(race_nwother_inp ~ treatment, data = ohp_data)
Education = lm(edu_inp ~ treatment, data = ohp_data)
Medicine= lm(rx_num_mod_inp ~ treatment, data = ohp_data)
Cholesterol = lm(chl_inp ~ treatment, data = ohp_data)

stargazer(Gender, Age, Not_White,Education, Cholesterol,
          title = "Overall Characterisitics for Explanatory Variables",
          font.size = "small",
          omit.stat = c("f", "ser"),
          column.sep.width = "6pt")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Apr 29, 2021 - 11:39:47 PM

Table 1: Overall Characterisitics for Explanatory Variables

	Dependent variable:				
	gender_inp	age_inp	race_nwother_inp	edu_inp	chl_inp
	(1)	(2)	(3)	(4)	(5)
treatment	−0.006	0.380*	0.003	0.022	−0.642
	(0.009)	(0.212)	(0.006)	(0.016)	(0.613)
Constant	0.569***	40.606***	0.142***	2.238***	205.769***
	(0.006)	(0.153)	(0.005)	(0.012)	(0.443)
Observations	12,229	12,228	12,190	12,218	12,174
R^2	0.00004	0.0003	0.00002	0.0001	0.0001
Adjusted R^2	−0.00004	0.0002	−0.0001	0.0001	0.00001
Note:					*p<0.1; **p<0.05; ***p<0.01

The Constant row in this table represents the mean of the control group. The treatment row represents the difference in means between the control group and the treatment group. The numbers in parenthesis below the treatment row represent the standard errors of the differences in means.The Difference in Group Means = Treatment Effect +Selection Bias. Here we are looking to see if the explanatory variables that would influence health outcomes are properly randomized. If that is the case we can assume that selection bias is nonexistent.

Table 2: Difference in Means of Characterisitics for Treatment and Control

Characteristic	Treatment	Control	difference	Standard_Error
Gender	0.5627055	0.5688121	-0.0061066	0.009
Age	40.9863786	40.6060606	0.3803180	0.212
Not_White	0.1458202	0.1424648	0.0033554	0.006
Education	2.2600721	2.2383970	0.0216751	0.016
Cholesterol	205.1270200	205.7691757	-0.6421557	0.613

2.4 Question 4

> Is the balance table consistent with individuals having been randomly assigned to treatment group and control groups? Why or why not?

The tables above show that the experiment is consistent with a random assignment for gender, race (not white), education, and cholesterol as the differences in means are not statistically significant. The difference in age is significant at the 10 percent level, which suggests that the age between the treatment and control groups may not have been properly randomized.

2.5 Question 5

> Estimate the compliance rate for the OHP experiment. That is, what is the effect of being assigned to the treatment group on the probability of being enrolled in Medicaid? Hint: For this question and question 7, you can use the same regression as in question 3, just changing the dependent variable.

```
#regress ohp all ever survey on treatment
lm(data= ohp_data, ohp_all_ever_survey ~ treatment)
```

```
##
## Call:
## lm(formula = ohp_all_ever_survey ~ treatment, data = ohp_data)
##
## Coefficients:
## (Intercept)    treatment
##      0.1583       0.2536
```

Individuals who were assigned to the treatment group were 25.36% more likely to be enrolled in Medicaid.

2.6 Question 6

> What is the intent-to-treat (ITT) effect of the OHP experiment on health outcomes? Please create a nicely formatted table that reports ITT estimates on 4 to 6 relevant health outcomes. Again, part of this question is to get you to think about which 4 to 6 variables could be used as health outcome variables.

I used diabetes diagnosis post treatment, medicine taken, hypertension post treatment, medical visits, and cholesterol. I tried to select characteristics that were important for medical use (i.e. doctors visits or prescriptions) as well as actual outcomes such as diabetes. This was done to properly evaluate the multiple impacts of the study.

```
dia <- lm(data=ohp_data, dia_dx_post_lottery~treatment)


doc <- lm(data = ohp_data, doc_num_mod_inp~treatment)


hdp <- lm(data=ohp_data, hbp_dx_post_lottery~treatment)
```

```r
meds <- lm(data=ohp_data,rx_num_mod_inp~treatment)


chl <- lm(data=ohp_data,chl_inp~treatment)



stargazer(dia, doc, hdp,meds, chl,
          title = "Overall Regression Results for Outcome Variables",
          font.size = "small",
          omit.stat = c("f", "ser"),
          column.sep.width = "1pt")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Apr 29, 2021 - 11:39:47 PM

Table 3: Overall Regression Results for Outcome Variables

	Dependent variable:				
	dia_dx_post_lottery	doc_num_mod_inp	hbp_dx_post_lottery	rx_num_mod_inp	chl_inp
	(1)	(2)	(3)	(4)	(5)
treatment	0.009^{***}	0.396^{*}	0.002	0.128^{**}	-0.642
	(0.002)	(0.216)	(0.004)	(0.053)	(0.613)
Constant	0.012^{***}	5.746^{***}	0.057^{***}	1.838^{***}	205.769^{***}
	(0.002)	(0.156)	(0.003)	(0.038)	(0.443)
Observations	12,186	12,158	11,945	11,912	12,174
R^2	0.001	0.0003	0.00003	0.0005	0.0001
Adjusted R^2	0.001	0.0002	-0.0001	0.0004	0.00001
Note:					*p<0.1; **p<0.05; ***p<0.01

2.7 Question 7

What is the "treatment on the treated" effect (ATET) of the OHP experiment, i.e. the effect among those who applied for Medicaid? Estimate it for every health outcome you chose in question 6 and provide some intuition for the calculation of this estimate.

```r
#List ITT coefficients
ATET = list(c(0.009, 0.396, 0.002, 0.128, -0.642))

#List outcome
Outcome = list(c("Diabetes Post", "Medical Visits", "Hypertension Post", "Medication","Cholesterol" ))

ATET = as.data.frame(ATET)

#Create data frame
Outcome = as.data.frame(Outcome)
Outcome = Outcome %>% mutate(ATET = ATET/.2536)%>% mutate(Outcome = Outcome)
Outcome = Outcome %>% select(ATET, Outcome)
```

Table 4: ATET

ATET	Outcome
0.035488959	Diabetes Post
1.561514196	Medical Visits
0.007886435	Hypertension Post
0.504731861	Medication
-2.531545741	Cholesterol

```
#Output table
kable(Outcome, format = "latex", booktabs = TRUE,
        caption = "ATET",
        format.args = list(big.mark= ","))
```

To calculate the ATET the coefficients in Table 4 (the intent to treat effect) were divided by the compliance rate. This was to determine the impact on treated individuals that actually participated in in the experiment.

2.8 Question 8

Do you have to worry about attrition bias in analyzing this data? Explain why or why not.

We should always be worried about attrition bias, as participants could theoretically move out of Oregon or stop participating. However, because the study was conducted over the relatively short period of the 10th of march 2008 to September, 30th 2009, there is not enough time for participants to develop survey fatiugue. For this reason, we do not have to be that worried about attrition bias.