PUBLIC POLICY 596 - DIRECTED STUDY

The Effects of Oregon Health Plan on Income

Andrew Joonhyung Park

University of California, Los Angeles

Luskin School of Public Affairs

Department of Public Policy

June 6, 2020

Supervised by: Martin Gilens

Professor of Public Policy

Introduction

With a background in psychology, I have been interested in the collinear relation between mental health and social mobility, particularly for marginalized populations. In short, these interests have led me to pursue a master's degree in public policy at UCLA Luskin School of Public Affairs. Throughout my time in the program, my interests have remained consistent, but have broadened to include physical health and other factors affecting social mobility such as income inequality and education.

As part of an elective for my final quarter, I enrolled in a directed study where students submit a proposal of tasks to pursue during the quarter under the supervision of a faculty member of their choice. Although I have taken a few quantitative analytical courses at UCLA, I wanted to apply my knowledge into research. Under the supervision of Dr. Gilens, I decided to use the dataset (NBER, 2015) and methods of "The Oregon Health Insurance Experiment: Evidence From the First Year" (Finkelstein et al., 2012) to study the effects of the Oregon Health Plan (OHS) lottery on income after a year for individuals below the federal poverty level (FPL). This study was chosen for several reasons. It aligned with my research/policy interests, its data was publicly available, and it was a randomized control trial. As a beginner in quantitative analysis, a randomized control trial would theoretically allow me to circumvent more advanced statistical methods to overcome selection bias. In other words, if done correctly, random assignment eliminates selection bias enabling us to substitute the expected untreated potential outcome with the expected outcome of the control group and the expected treated potential outcome with the expected outcome of the treated group. This allows us to calculate the treatment effect either through a regression or a difference-in-differences approach.

The Oregon Health Insurance Experiment (Finkelstein et al., 2012)

Overview

In the first stages of Finkelstein et al.'s (2012) research, potentially eligible individuals were given the opportunity to sign up for a lottery that would allow them to apply for the Oregon Health Plan (OHP) Standard— a state-run insurance plan that offered extensive medical coverage. The Office for Oregon Health Policy and Research (2009) determined eligibility as

> Adults ages 19–64 not otherwise eligible for public insurance who are Oregon residents, are U.S. citizens or legal immigrants, have been without health insurance for six months, have income below the federal poverty level (FPL), and have assets below $2,000.

During the first step, eligibility screening was done using limited demographic data. Another screening took place during the application process, which used more extensive data such as credit reports etc. From the eligible applicants, individuals were randomly assigned to either the treatment or control group. If assigned to the treatment group, OHP would be retroactively applied.

Methods

To measure the outcomes of the lottery Finkelstein et al. (2012) estimated the intent-to-treat effect (ITT) and the local average treatment effect (LATE). However, the focus and scope of this report will only be regarding their ITT analysis. In their analysis, Finkelstein et al. (2012) estimated their outcomes (i.e. health) through an ordinary least square regression (OLS):

$$y_{ihj} = \beta_0 + \beta_1 \mathrm{LOTTERY}_h + X_{ih}\beta_2 + V_{ih}\beta_3 + \varepsilon_{ihj},$$

using *LOTTERY* as an instrument for their treatment, insurance. Subscript *j* in the outcome variable, *y*, represents an outcome within a larger set of related outcomes, *J*. For example, if *J* represented health, it would include outcomes such as the number of emergency visits or self-reported mental health. The subscript *i* represents each observation or individual, while *h* is their household identifier.

X_{ih} represents a set of variables that are correlated with treatment assignment and possibly the outcome variable. It was included as a control to alleviate potential bias for the estimate of β_1, the ITT estimate. Although the treatment group was randomly selected from a pool of eligible participants, many of the participants did not actually receive OHP for various reasons. 1)The first screening process used limited demographic data and created a handful of false positives. 2) Some participants may have failed to apply before the deadline. Lastly, V_{ih} represents a set of variables included to overcome possible selection bias attributable to chance. Excluding V_{ih} will not create a biased estimate of β_1 but including it will improve standard errors.

The Effect Oregon Health Plan on Income

Research Question: Did the opportunity to enroll in Oregon Health Plan (OHP) improve the income levels of individuals below the federal poverty level (FPL)?

Hypothesis 1: On average, individuals given the chance to apply for OHP will experience higher income levels after 12 months than the control group.

Hypothesis 2: The income levels of racial minorities will most likely benefit less than their counterparts.

Hypothesis 3: Individuals with serious chronic conditions are less likely to be better off due to insurance than individuals without any chronic conditions or people with less serious conditions.

Operational Definition of "Serious Chronic Conditions"

To test hypothesis 3, there needed to be a consistent criterion that defines "serious." The first thought that came to mind was ranking the conditions based on national mortality rate (Appendix 1). Although a high mortality rate would indicate a level of severity, it does not seem all that likely for the data to be able to capture any significant relationship between mortality and income within a year's time. Using the available variables from Finkelstein et al.'s (2012) study, I decided to rank the conditions based on mortality rate, number of emergency visits, and number of emergency visits that resulted in hospitalization. I ran three multivariate regressions with the pre-treatment medical conditions as the outcome and the ranking criteria as the independent variables.

Table 1

Regression Output

	Dependent variable:		
	Emergency Visits, Hospitalizations	Emergency Visits	Deaths
	(1)	(2)	(3)
Asthma	0.005	0.292^{***}	0.0001
	(0.010)	(0.047)	(0.0005)
Diabetes	0.116^{***}	0.175^{**}	0.0002
	(0.016)	(0.073)	(0.001)
Hypertension	0.046^{***}	0.208^{***}	0.0001
	(0.011)	(0.051)	(0.001)
High Cholesterol	0.010	-0.262^{***}	0.001
	(0.013)	(0.059)	(0.001)
Heart Attack	0.180^{***}	0.627^{***}	-0.002^{*}
	(0.029)	(0.138)	(0.001)
Congestive Heart Failure	0.189^{***}	-0.109	0.007^{***}

	(0.038)	(0.180)	(0.002)
Emphysema	0.115***	0.484***	-0.001
	(0.025)	(0.120)	(0.001)
Kidney Failure	0.058**	0.136	0.004***
	(0.028)	(0.134)	(0.001)
Cancer	-0.002	0.195**	-0.001
	(0.019)	(0.090)	(0.001)
Depression	0.069***	0.491***	-0.0004
	(0.008)	(0.039)	(0.0004)
Constant	0.031***	0.512***	0.0003
	(0.005)	(0.024)	(0.0002)
Observations	10,178	10,173	12,229
R^2	0.038	0.035	0.002
Adjusted R^2	0.038	0.034	0.002

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

I decided to define “serious chronic conditions” as conditions that were significantly associated with both emergency visits did and did not result in hospitalization: diabetes, hypertension, heart attack, emphysema and depression.

Balance Test

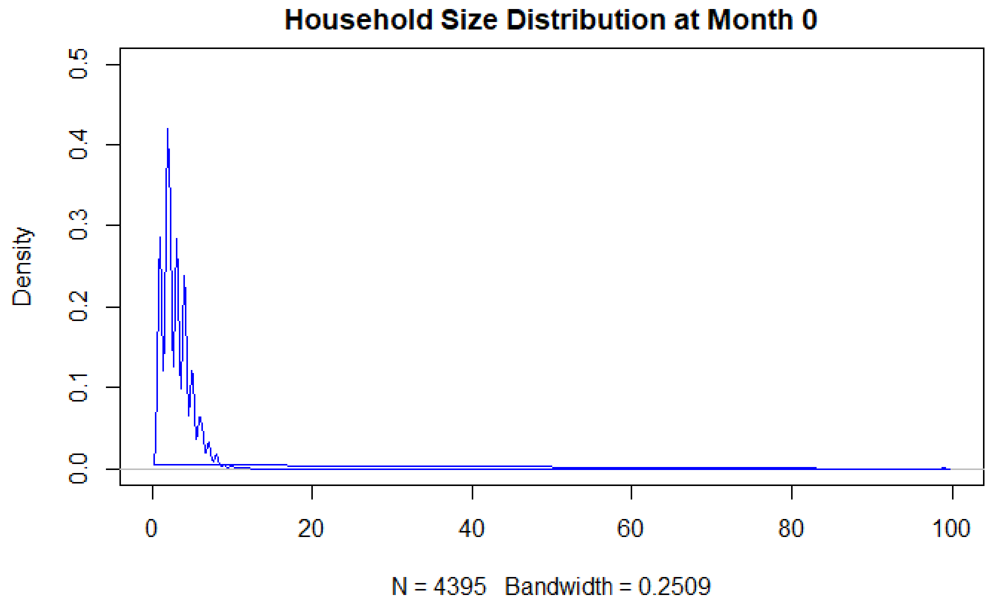
As aforementioned, because the participants were randomized into treatment groups, selection bias should have been mitigated. However, as an internal validity check, I ran a couple balance tests.

Balance table 1

The first balance test compared the coefficients of the pre-treatment variables in a series of bivariate regressions with lottery/treatment as the independent variable and controlling for household size at month 0 (Appendix 2.1). Household size was included to control for its inevitable effect on the outcome (income). The dataset provided by Finkelstein et al. (2012) measured income levels as categorical ranges of the federal poverty level, meaning income levels were adjusted according to an individual’s household size. For example, if two people made the same yearly income, the person with a larger household would be considered as having less income. Household size was also controlled for due to its effect on treatment assignment (lottery probability). All members of a household were eligible to apply for OHP if any member of the household was picked by the lottery. If household size were uncontrolled, the results would have suffered from omitted variable bias and selection bias.

A few of the household sizes were unusually large at month 0 (Graph 1). As Finkelstein et al. (2012) had no mention of what these values represented, household size was included as a fixed effect rather than a continuous variable and the extreme outliers were dropped (household size ≥ 99).

Graph 1



Balance Table (Fixed Effect: Household Size)

	Dependent variable:												
	Female	YOB	Education	Black	White	Hispanic	Non-White Other	Income 0mo	Diabetes	Hypertension	Heart Attack	Emphysema	Depression
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Treatment	-0.005	0.245	0.019	-0.002	-0.015	0.013	0.001	0.217**	-0.009	-0.002	-0.001	0.0001	-0.023*
	(0.011)	(0.259)	(0.025)	(0.008)	(0.012)	(0.010)	(0.010)	(0.107)	(0.007)	(0.011)	(0.004)	(0.004)	(0.013)
Observations	7,724	7,724	5,177	5,165	5,165	5,167	5,165	7,288	5,179	5,179	5,179	5,179	5,179
R^2	0.008	0.079	0.035	0.007	0.069	0.104	0.003	0.044	0.010	0.017	0.013	0.004	0.025
Adjusted R^2	0.006	0.076	0.032	0.003	0.066	0.101	-0.001	0.042	0.006	0.013	0.009	0.0001	0.022

Note: *p<0.1; **p<0.05; ***p<0.01

Unfortunately, as shown in the balance table above not all the pre-treatment variables are balanced— household income at month 0 and the number of participants with depression—indicating the need to control for these variables in the regression analysis.

Balance table 2

In the second balance table, household income and depression were also included as a fixed effect as a way of exact matching (Appendix 2.2).

Table 3

Balance Table (Fixed Effect: Household Size, Income, Depression)

	Dependent variable:												
	Female	YOB	Education	Black	White	Hispanic	Non-White Other	Income 0mo	Diabetes	Hypertension	Heart Attack	Emphysema	Depression
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Treatment	-0.008	-0.292	0.001	-0.001	-0.016	0.011	0.0001	0.000	-0.011	0.0001	-0.001	0.002	-0.000
	(0.014)	(0.324)	(0.026)	(0.008)	(0.012)	(0.010)	(0.010)	(0.000)	(0.008)	(0.011)	(0.004)	(0.004)	(0.000)
N	4,876	4,876	4,874	4,863	4,863	4,864	4,863	4,876	4,876	4,876	4,876	4,876	4,876
R^2	0.050	0.089	0.074	0.025	0.091	0.113	0.009	1.000	0.017	0.036	0.010	0.012	1.000
Adjusted R^2	0.042	0.081	0.067	0.018	0.083	0.106	0.001	1.000	0.009	0.029	0.002	0.004	1.000

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

One possible explanation for the difference between the control and treatment group, that I had originally thought to be true, was that there was an error in randomization. However, it was because I was using different pre-treatment variables from Finkelstein et al. (2012). I had used variables that were measured before the treatment was implemented, while they used more extensive, confidential dataset unavailable to the public. Nevertheless, balance was achieved through matching.

Statistical Analysis

Multivariate Regression

To test my hypotheses, I ran a multivariate regression (Appendix 3.1), which included the pre-treatment variables as controls, the interactions terms (testing hypothesis two and three) and, income at month 12 as the outcome.

Table 4

Multivariate Regression Output (Interactions: Race and Health Conditions)

	Dependent variable:
	Income 12mo
Treatment	0.340
	(0.540)
Female	0.125
	(0.142)
YOB	0.006
	(0.006)
Education	0.515^{***}
	(0.077)
Black	-0.464
	(0.485)
White	0.317
	(0.385)
Hispanic	0.521
	(0.403)
Non-White Other	-0.497
	(0.347)
Diabetes	-0.300
	(0.364)
Hypertension	-0.287
	(0.263)
Heart Attack	-0.719
	(0.624)
Emphysema	0.515
	(0.684)
Treatment*Black	-0.046
	(0.660)
Treatment*White	-0.409
	(0.518)
Treatment*Hispanic	-0.794

	(0.539)
Treatment*Non-White Other	0.571
	(0.476)
Treatment*Diabetes	0.351
	(0.519)
Treatment*Hypertension	0.095
	(0.356)
Treatment*Heart Attack	0.242
	(0.894)
Treatment*Emphysema	-1.478
	(0.908)
Treatment*Depression	-0.073
	(0.284)
Observations	3,185
R^2	0.444
Adjusted R^2	0.434

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Surprisingly, none of coefficients besides education had a significant relationship with household income at month 12. It seemed particularly odd that race had no effect on income level at 12 months, so instead of including serious chronic conditions as interactions, they were included as control variables (Appendix 3.2). The results were consistent with the first regression's results.

Table 5

Multivariate Regression Output (Interaction: Race)

	Dependent variable:
	Income 12mo
Treatment	0.335
	(0.532)
Female	0.121
	(0.142)
YOB	0.006
	(0.006)
Education	0.516^{***}
	(0.077)
Black	-0.488
	(0.484)
White	0.330
	(0.384)

Hispanic	0.517
	(0.403)
Non-White Other	-0.485
	(0.346)
Diabetes	-0.134
	(0.259)
Hypertension	-0.236
	(0.185)
Heart Attack	-0.627
	(0.448)
Emphysema	-0.314
	(0.451)
Treatment*Black	-0.0005
	(0.658)
Treatment*White	-0.419
	(0.516)
Treatment*Hispanic	-0.774
	(0.539)
Treatment*Non-White Other	0.559
	(0.476)
Observations	3,185
R^2	0.443
Adjusted R^2	0.434

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Conclusion: At p-value < 0.1, I failed to reject all three null hypotheses. The lottery had no significant effect on income levels after 12 months of treatment, nor did it have any differing effect on the income levels of different race/ethnicity and individuals with serious chronic health conditions.

Discussion

Difference-in-Differences & Matching

To verify my results, I have also analyzed the results using difference-in-differences and matching, which all results in the same outcome (Appendix 4.1 & 4.2). Although, it was disappointing to not be able to reject any of the null hypotheses, I found slight solace in that the results were at least consistent.

Instrumental Variable

With the guidance of Dr. Gilens, I have decided not to implement an instrument variable in my ITT equation. Although the use of instruments in an ITT is quite common and often very useful, there are a few drawbacks. On one hand, instruments allow us to effectively measure the effects of a policy or program where there are many non-compliers. However, for the measure to be valid, the instrument variable needs to validate several assumptions one of which is untestable –exclusion restriction. The assumption is that the instrumental variable (i.e. *LOTTERY)* only affects the outcome through the treatment variable (i.e. insurance).

Limitations

Although Finkelstein et al. (2012) randomized the lottery participants into treatment or control groups, individuals had to first sign up to be included in the lottery. While the researchers did attempt to remove as many barriers to signing up as possible, sampling bias may have affected the result's external validity. Another source of limitation is the fact that the individuals in the treatment group had to reverify their eligibility every 6 months. Once the participants in the treatment group are deemed ineligible, they may stop responding to the researcher's surveys. Since one of the ways an individual can become ineligible is by exceeding the federal poverty level, there could have been an underestimation of the treatment group's income level.

Bibliography

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., & Oregon Health Study Group (2012). THE OREGON HEALTH INSURANCE EXPERIMENT: EVIDENCE FROM THE FIRST YEAR. *The quarterly journal of economics*, *127*(3), 1057–1106. https://doi.org/10.1093/qje/qjs020

NBER. (2015, August 13). The Oregon Health Insurance Experiment. Retrieved June 10, 2020, from https://www.nber.org/oregon/1.home.html

Office for Oregon Health Policy Research. Trends in Oregon's HealthCare Market and the Oregon Health Plan: A Report to the 75th Legislative Assembly. 2009 Feb

Appendix 1: Defining “serious chronic conditions”

```
# Dependent Variables
hosp_visits_df = ed_vars %>%
  dplyr::select (person_id,
                num_hosp_pre_cens_ed,
                num_visit_pre_cens_ed)

deaths_df = descriptive_vars %>%
  dplyr::select (person_id,
                postn_death)

# Independent Variable
conditions_df= inperson_vars %>%
  dplyr::select (person_id ,
                ends_with("pre_lottery_inp"))

# Combining hospital visits and death
df_a = merge.data.frame(hosp_visits_df,
                       deaths_df,
                       by="person_id",
                       all.y = TRUE)

# Combining previous with conditions
df_b= merge.data.frame(df_a,
                      conditions_df,
                      by="person_id",
                      all.y = TRUE)
```

```r
# Regression to find more serious conditions
cond_reg = map(c("num_hosp_pre_cens_ed",
                 "num_visit_pre_cens_ed",
                 "postn_death"), function(x)
  str_c(x, "~ast_dx_pre_lottery_inp+
        dia_dx_pre_lottery_inp+
        hbp_dx_pre_lottery_inp+
        chl_dx_pre_lottery_inp+
        ami_dx_pre_lottery_inp+
        chf_dx_pre_lottery_inp+
        emp_dx_pre_lottery_inp+
        kid_dx_pre_lottery_inp+
        cancer_dx_pre_lottery_inp+
        dep_dx_pre_lottery_inp") %>%
    as.formula() %>%
    lm(data = df_b)
)
stargazer(cond_reg, type="text", out="C://Users//qkra0//testing.htm")

# Combining the conditions with regression data frame
df_reg = merge.data.frame(df_reg,
                          df_b,
                          by="person_id")

# Checking to see if participants no longer alive are still in data set
sum(df_reg$postn_death=="Dead")

#Ordering columns
df_reg=df_reg %>%
  dplyr::select(person_id,
```

```
household_id,
treatment,
hhsize_0m,
hhsize_12m,
everything())
```

Appendix 2: Balance tests

2.1 Household size fixed effect

```
balance_1 = map(c("female",
                  "yob",
                  "educ",
                  "race_black",
                  "race_white",
                  "hispanic_inp",
                  "race_nonwhite_other",
                  "hhinc_cat_0m",
                  "dia_dx_pre_lottery_inp",
                  "hbp_dx_pre_lottery_inp",
                  "ami_dx_pre_lottery_inp",
                  "emp_dx_pre_lottery_inp",
                  "dep_dx_pre_lottery_inp"), function(x)
  str_c(x, "~treatment | hhsize_0m") %>%
    as.formula() %>%
    felm(data = df_reg)
)
stargazer(balance_1, type="text", flip=TRUE, out="C://Users//qkra0//testing.htm")
```

2.2 Household size, income, and depression fixed effects

```
balance_2 = map(c("female",
                  "yob",
                  "educ",
                  "race_black",
                  "race_white",
                  "hispanic_inp",
                  "race_nonwhite_other",
                  "hhinc_cat_0m",
```

```
                        "dia_dx_pre_lottery_inp",
                        "hbp_dx_pre_lottery_inp",
                        "ami_dx_pre_lottery_inp",
                        "emp_dx_pre_lottery_inp",
                        "dep_dx_pre_lottery_inp"), function(x)
  str_c(x, "~treatment | hhsize_0m + hhinc_cat_0m + dep_dx_pre_lottery_inp") %>%
    as.formula() %>%
    felm(data = df_reg)
)
stargazer(balance_2, type="text", flip=TRUE, out="INPUT FILE PATH")
```

Appendix 3: Multivariate regression

3.1 Regression with both race and health condition interactions

```
inter_reg = felm(hhinc_cat_12m~ treatment+
           female+
           yob+
           educ+
           race_black+
           race_white+
           hispanic_inp+
           race_nonwhite_other+
           treatment*race_black+
           treatment*race_white+
           treatment*hispanic_inp+
           treatment*race_nonwhite_other+
           dia_dx_pre_lottery_inp+
           hbp_dx_pre_lottery_inp+
           ami_dx_pre_lottery_inp+
           emp_dx_pre_lottery_inp+
           treatment*dia_dx_pre_lottery_inp+
           treatment*hbp_dx_pre_lottery_inp+
           treatment*ami_dx_pre_lottery_inp+
           treatment*emp_dx_pre_lottery_inp+
           treatment*dep_dx_pre_lottery_inp
          |hhsize_0m + hhinc_cat_0m + dep_dx_pre_lottery_inp|0|0,  # Added depression
          na.action = na.omit,
          data=df_reg)
stargazer(inter_reg, type="text", na.rm=TRUE, out="INPUT FILE PATH")
```

3.2 Regression with race interactions

```
inter_reg = felm(hhinc_cat_12m~ treatment+
```

```
          female+
          yob+
          educ+
          race_black+
          race_white+
          hispanic_inp+
          race_nonwhite_other+
          treatment*race_black+
          treatment*race_white+
          treatment*hispanic_inp+
          treatment*race_nonwhite_other+
          dia_dx_pre_lottery_inp+
          hbp_dx_pre_lottery_inp+
          ami_dx_pre_lottery_inp+
          emp_dx_pre_lottery_inp+
          dep_dx_pre_lottery_inp
         |hhsize_0m + hhinc_cat_0m + dep_dx_pre_lottery_inp|0|0,  # Added depression
         na.action = na.omit,
         data=df_reg)
stargazer(inter_reg, type="text", na.rm=TRUE, out="INPUT FILE PATH")
```

Appendix 4: Verifying analysis (code and output)

4.1 Difference-in-differences

```
# Creating empty list
m=list()

# Control differences
m[[1]]=lm(hhinc_cat_0m~treatment|hhsize_0m + hhinc_cat_0m + dep_dx_pre_lottery_inp|0|0,data=df_did)
m[[2]]=lm(hhinc_cat_12m~treatment|hhsize_0m + hhinc_cat_0m + dep_dx_pre_lottery_inp|0|0,data=df_did)

# Treatment differences
m[[3]]=lm(hhinc_cat_0m~control|hhsize_0m + hhinc_cat_0m + dep_dx_pre_lottery_inp|0|0,data=df_did)
m[[4]]=lm(hhinc_cat_12m~control|hhsize_0m + hhinc_cat_0m + dep_dx_pre_lottery_inp|0|0,data=df_did)

# Difference-in-differences
DiD = (m[[4]]$coefficients["(Intercept)"] - m[[3]]$coefficients["(Intercept)"]) - (m[[2]]$coefficients["(Intercept)"] -
m[[1]]$coefficients["(Intercept)"])

summary(DiD)

## Output:

##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -0.08515 -0.08515 -0.08515 -0.08515 -0.08515 -0.08515

# Verfying mean estimates
stargazer (m, type="text", out="C://Users//qkra0//matching.htm")

## Output:
```

	Dependent variable:			
	Income 0mo, Control	Income 12mo, Treatment	Income 0mo, Treatment	Income 12mo, Treatment
	(1)	(2)	(3)	(4)
Constant	6.246***	6.823***	6.212***	6.704***
	(0.057)	(0.070)	(0.056)	(0.067)
Observations	6,447	5,219	6,747	5,485
R^2	0.000	0.000	0.000	0.000
Adjusted R^2	0.000	0.000	0.000	0.000
Note:				*p<0.1; **p<0.05; ***p<0.01

4.2 Matching

```
form=as.formula("hhinc_cat_12m~ treatment+
        female+
        yob+
        educ+
        race_black+
        race_white+
        hispanic_inp+
        race_nonwhite_other+
        treatment*race_black+
        treatment*race_white+
        treatment*hispanic_inp+
        treatment*race_nonwhite_other+
        dia_dx_pre_lottery_inp+
        hbp_dx_pre_lottery_inp+
        ami_dx_pre_lottery_in+
        emp_dx_pre_lottery_inp+
        dep_dx_pre_lottery_inp+
        treatment*dia_dx_pre_lottery_inp+
        treatment*ami_dx_pre_lottery_inp+
        treatment*chf_dx_pre_lottery_inp+
        treatment*kid_dx_pre_lottery_inp+
        treatment*dep_dx_pre_lottery_inp")

pre_treat_raw="female+yob+educ+race_black+race_white+hispanic_inp+dia_dx_pre_lottery_inp+ami_dx_pre_lottery_inp+chf_dx_pre_lottery_inp+kid_dx_pre_lottery_inp+dep_dx_pre_lottery_inp+dia_dx_pre_lottery_inp+hbp_dx_pre_lottery_inp+ami_dx_pre_lottery_inp+emp_dx_pre_lottery_inp+dep_dx_pre_lottery_inp"

# Creating a binary treat variable just in case you need it
df_match = df_reg %>%
  mutate(treat = ifelse(treatment==0, "FALSE", "TRUE"))


# Ordering the data frame
df_match = df_match %>%
  dplyr::select (c(person_id,
        household_id,
        treatment,
        treat,
        everything()))

# Splitting the string by the + sign
pre_treat=strsplit(pre_treat_raw, "[+]")
# Removing the levels in the list
pre_treat=unlist(pre_treat)
# Removing starting and ending spaces
pre_treat=trimws(pre_treat)

# Ordering columns
df_match = df_match %>%
```

```
  dplyr::select (c(person_id,
                household_id,
                treatment,
                treat,
                everything()))


# Getting rid of unnecessary variables
df_match_dropped = df_match %>% dplyr::select (-c(approved_app, postn_death.x, num_hosp_pre_cens_ed,
num_visit_pre_cens_ed, postn_death.y), -c(starts_with ("ins_")))

# Dropping all NAs in df_match_dropped to use the matching function
df_match_dropped=drop_na (df_match_dropped)


# Exact Matching
match_exact <- Matching::Match(Y = df_match_dropped$hhinc_cat_12m,
                              # Outcome Variable
                              Tr = df_match_dropped$treatment,
                              # Treatment Variable
                              X = dplyr::select(df_match_dropped, pre_treat),
                              # Xs to Match On
                              M = 1,
                              # Number of Matches
                              exact = TRUE,
                              estimand = "ATT")

## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(pre_treat)` instead of `pre_treat` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.

summary (match_exact)

##
## Estimate...  -0.030976
## AI SE......  0.15156
## T-stat.....  -0.20438
## p.val......  0.83805
##
## Original number of observations.............  3053
## Original number of treated obs...............  1615
## Matched number of observations...............  809
## Matched number of observations  (unweighted).  1685
##
## Number of obs dropped by 'exact' or 'caliper'  806

# Non-Exact Matching
match_exact <- Matching::Match(Y = df_match_dropped$hhinc_cat_12m,
                              # Outcome Variable
                              Tr = df_match_dropped$treatment,
                              # Treatment Variable
```

```
                          X = dplyr::select(df_match_dropped, pre_treat),
                          # Xs to Match On
                          M = 1,
                          # Number of Matches
                          exact = FALSE,
                          # Standard deviation tolerance
                          distance.tolerance = 0.25,
                          estimand = "ATT")

summary (match_exact)

##
## Estimate...  0.14336
## AI SE......  0.19917
## T-stat.....  0.71978
## p.val......  0.47166
##
## Original number of observations..............  3053
## Original number of treated obs...............  1615
## Matched number of observations...............  1615
## Matched number of observations  (unweighted).  17968
##
## Number of obs dropped by 'exact' or 'caliper'  0
```