# ECN525_Proj1

Rhodes Kirkpatrick

4/28/2022

```
library(pacman)
p_load(haven, foreign, dplyr,tidyverse,tidyfast,modelsummary)
```

```
#Read in Data
OHP = read.dta('ohp.dta')
OHP[] = lapply(OHP,as.numeric)
```

#1: The variable “treatment” is having been chosen by the lottery within the experiment in the time span of conducting the experiment. The other variable in question “ohp_all_ever” is a dummy variable whether or not someone had ever been enrolled in Medicaid regardless of enrollment happened in the experiment. The reason why the former is the treatment (Di) variable is that we are specifically looking at the changes caused by having Medicaid. The Intent-to-Treat effect can only be systemically accurately measured with a direct counter-factual, which would be everyone not selected (the control group.) This is so because we need to know how individual behavior changed before and after coverage, however, with the same general environment to impede outside forces from confusing the comparison. Different administrative policies, pandemic post 2020, maybe one year there was extra fog around and it trapped more pm5 particulates in the air, ect.. these are all reason why we need “treatment” to be the treatment variable, in order to have a control group that can accurately reference the effects.

#2 and #3: In evaluating whether or not treatment was truly randomly assigned, we should look across characteristics that should be balanced when randomized across both control and treatment groups. Demographic data (race, age, gender, language) are the first and obvious place to look along with pre-treatment health measurements. This is important because if the treated start with higher blood pressure for example, the conclusion is likely to state that Medicaid was associated with higher ER visits when Medicaid might not have been the dominating force.

```
#2: Summary Control and Treatment Variables
control = subset(OHP, treatment ==1, select = c(age_inp, gender_inp, race_white_inp,race_black_inp, hispanic_inp,
itvw_english_inp,dia_dx_pre_lottery))

treat = subset(OHP, treatment ==2, select = c(age_inp, gender_inp, race_white_inp,race_black_inp, hispanic_inp,it
vw_english_inp,dia_dx_pre_lottery,ohp_all_ever_survey))

Csum = summary(control)
Csum
```

```
##     age_inp       gender_inp    race_white_inp race_black_inp   hispanic_inp
##  Min.   :19.00   Min.   :1.000   Min.   :1.00   Min.   :1.000   Min.   :1.000
##  1st Qu.:30.00   1st Qu.:1.000   1st Qu.:1.00   1st Qu.:1.000   1st Qu.:1.000
##  Median :41.00   Median :2.000   Median :2.00   Median :1.000   Median :1.000
##  Mean   :40.61   Mean   :1.569   Mean   :1.69   Mean   :1.107   Mean   :1.178
##  3rd Qu.:50.00   3rd Qu.:2.000   3rd Qu.:2.00   3rd Qu.:1.000   3rd Qu.:1.000
##  Max.   :68.00   Max.   :3.000   Max.   :2.00   Max.   :2.000   Max.   :2.000
##  NA's   :1                       NA's   :16     NA's   :16      NA's   :10
##  itvw_english_inp dia_dx_pre_lottery
##  Min.   :0.0000   Min.   :1.000
##  1st Qu.:1.0000   1st Qu.:1.000
##  Median :1.0000   Median :1.000
##  Mean   :0.8826   Mean   :1.072
##  3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.0000   Max.   :2.000
##
```

```
Tsum = summary(treat)
Tsum
```

```
##     age_inp        gender_inp    race_white_inp  race_black_inp  hispanic_inp
##  Min.   :20.00   Min.   :1.000   Min.   :1.000   Min.   :1.0    Min.   :1.000
##  1st Qu.:31.00   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.0    1st Qu.:1.000
##  Median :41.00   Median :2.000   Median :2.000   Median :1.0    Median :1.000
##  Mean   :40.99   Mean   :1.563   Mean   :1.687   Mean   :1.1    Mean   :1.182
##  3rd Qu.:51.00   3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:1.0    3rd Qu.:1.000
##  Max.   :71.00   Max.   :2.000   Max.   :2.000   Max.   :2.0    Max.   :2.000
##                                  NA's   :23      NA's   :23     NA's   :19
##  itvw_english_inp dia_dx_pre_lottery ohp_all_ever_survey
##  Min.   :0.00     Min.   :1.000      Min.   :1.000
##  1st Qu.:1.00     1st Qu.:1.000      1st Qu.:1.000
##  Median :1.00     Median :1.000      Median :1.000
##  Mean   :0.87     Mean   :1.071      Mean   :1.412
##  3rd Qu.:1.00     3rd Qu.:1.000      3rd Qu.:2.000
##  Max.   :1.00     Max.   :2.000      Max.   :2.000
##
```

```
#3: Regression of Control on Balance Variables
#Install and load sandwich and lmtest packages
p_load(sandwich, lmtest)

#Regression with homoskedasticity-only standard errors
mod1 = lm(treatment~age_inp+gender_inp+race_white_inp+race_black_inp+hispanic_inp+itvw_english_inp+dia_dx_pre_lot
tery, data = OHP)
summary(mod1)
```

```
##
## Call:
## lm(formula = treatment ~ age_inp + gender_inp + race_white_inp +
##     race_black_inp + hispanic_inp + itvw_english_inp + dia_dx_pre_lottery,
##     data = OHP)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5880 -0.5202  0.4486  0.4791  0.5331
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.6077960  0.0630438  25.503   <2e-16 ***
## age_inp             0.0007205  0.0003983   1.809   0.0705 .
## gender_inp         -0.0070847  0.0091473  -0.775   0.4386
## race_white_inp     -0.0065898  0.0145080  -0.454   0.6497
## race_black_inp     -0.0237923  0.0188073  -1.265   0.2059
## hispanic_inp       -0.0193502  0.0181336  -1.067   0.2860
## itvw_english_inp   -0.0397183  0.0198298  -2.003   0.0452 *
## dia_dx_pre_lottery -0.0084072  0.0179376  -0.469   0.6393
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4994 on 12168 degrees of freedom
##   (53 observations deleted due to missingness)
## Multiple R-squared:  0.0009586,  Adjusted R-squared:  0.0003839
## F-statistic: 1.668 on 7 and 12168 DF,  p-value: 0.1119
```

```
#Report heteroskedasticity robust standard errors
coeftest(mod1, vcov = vcovHC(mod1, type="HC1"))
```

```
##
## t test of coefficients:
##
##                       Estimate  Std. Error t value Pr(>|t|)
## (Intercept)         1.60779596  0.06283034 25.5895  < 2e-16 ***
## age_inp             0.00072053  0.00039831  1.8090  0.07048 .
## gender_inp         -0.00708472  0.00914595 -0.7746  0.43857
## race_white_inp     -0.00658978  0.01446123 -0.4557  0.64862
## race_black_inp     -0.02379226  0.01879981 -1.2656  0.20570
## hispanic_inp       -0.01935016  0.01807172 -1.0707  0.28431
## itvw_english_inp   -0.03971826  0.01974963 -2.0111  0.04434 *
## dia_dx_pre_lottery -0.00840724  0.01793015 -0.4689  0.63916
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#4: Yes, the results are completely consistent with randomized assignment. There is an extremely small mean difference across all demographic data points (race, age, language used, and gender). All of these variables had no statistical significance except the use of english which I'm sure had a small affect but as far as validity goes, the biases are impressively minimized. Additionally, the difference in mean of those with diabetes is about the same, therefore it is implied there exists approximately equalized risk factors across the treatment and control groups and the outcomes proceeding the experiment will not reflect any particular cluster of risk factors causing undue influence to our interpretation of the results.

```
mod2 = lm(ohp_all_ever_survey~treatment, data = OHP)
summary(mod2)
```

```
##
## Call:
## lm(formula = ohp_all_ever_survey ~ treatment, data = OHP)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4119 -0.4119 -0.1583  0.5881  0.8417
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.904742   0.012650   71.52   <2e-16 ***
## treatment   0.253594   0.007896   32.12   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4361 on 12227 degrees of freedom
## Multiple R-squared:  0.0778, Adjusted R-squared:  0.07773
## F-statistic:  1032 on 1 and 12227 DF,  p-value: < 2.2e-16
```

#5: The population percent in the treatment who complied was 25%.

```
ITT = lm(treatment~hbp_dx_post_lottery+dep_dx_post_lottery+dia_dx_post_lottery+rx_num_mod_inp, data = OHP)
summary(ITT)
```

```
##
## Call:
## lm(formula = treatment ~ hbp_dx_post_lottery + dep_dx_post_lottery +
##     dia_dx_post_lottery + rx_num_mod_inp, data = OHP)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7042 -0.5140  0.4554  0.4834  0.4911
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          1.364612   0.046466  29.368  < 2e-16 ***
## hbp_dx_post_lottery -0.005088   0.020283  -0.251 0.801929
## dep_dx_post_lottery  0.018177   0.021120   0.861 0.389442
## dia_dx_post_lottery  0.136318   0.037383   3.647 0.000267 ***
## rx_num_mod_inp       0.002550   0.001668   1.529 0.126353
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4992 on 11508 degrees of freedom
##   (716 observations deleted due to missingness)
## Multiple R-squared:  0.00161,    Adjusted R-squared:  0.001263
## F-statistic: 4.638 on 4 and 11508 DF,  p-value: 0.0009674
```

```
ITTmodels = list(
  "Hypertension" = lm(hbp_dx_post_lottery~treatment + age_inp + race_white_inp +race_black_inp + age_inp, data =
OHP),
  "Depression" = lm(dep_dx_post_lottery~treatment + age_inp + race_white_inp +race_black_inp + age_inp, data = OH
P),
  "Diabetes" = lm(dia_dx_post_lottery~treatment + age_inp + race_white_inp +race_black_inp + age_inp, data = OHP)
,
  "RX Prescriptions" = lm(rx_num_mod_inp~treatment + age_inp + race_white_inp +race_black_inp + age_inp, data = O
HP)
)
msummary(ITTmodels,,stars=TRUE,title = "ITT Models")
```

ITT Models

| | Hypertension | Depression | Diabetes | RX Prescriptions |
|---|---|---|---|---|
| (Intercept) | 0.968*** | 1.052*** | 0.993*** | -3.356*** |
| | (0.017) | (0.016) | (0.009) | (0.202) |
| treatment | 0.002 | 0.005 | 0.008*** | 0.106* |

| | | | | |
|---|---|---|---|---|
| | (0.004) | (0.004) | (0.002) | (0.050) |
| age_inp | 0.002*** | 0.000** | 0.001*** | 0.078*** |
| | (0.000) | (0.000) | (0.000) | (0.002) |
| race_white_inp | -0.003 | 0.005 | -0.010*** | 0.711*** |
| | (0.005) | (0.005) | (0.003) | (0.060) |
| race_black_inp | 0.010 | 0.003 | -0.004 | 0.650*** |
| | (0.008) | (0.007) | (0.004) | (0.091) |
| Num.Obs. | 11909 | 12058 | 12148 | 11878 |
| R2 | 0.010 | 0.001 | 0.007 | 0.116 |
| R2 Adj. | 0.010 | 0.000 | 0.007 | 0.116 |
| AIC | -882.0 | -2162.8 | -15688.5 | 57483.2 |
| BIC | -837.7 | -2118.4 | -15644.0 | 57527.5 |
| Log.Lik. | 447.018 | 1087.393 | 7850.226 | -28735.600 |
| F | 30.900 | 2.206 | 21.646 | 390.625 |
| RMSE | 0.23 | 0.22 | 0.13 | 2.72 |

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

#7: The “treatment on the treated” (ATET) is equal to the Intent to Treat with full compliance. Since the compliance rate is less than 100%, we must divide tbe ITT by the compliance rate to get ATET. If treatment is causing a change in outcomes, it is those who comply who are actually making a difference as they are distinguished from the others in their actual participation. Therefore, we need to generalize the effect of this population across the entire treatment group to get the approximate effect over the group if compliance did equal 1.

```
H = ITTmodels[["Hypertension"]][["coefficients"]][["treatment"]]/mod2[["coefficients"]][["treatment"]]
Dep = ITTmodels[["Depression"]][["coefficients"]][["treatment"]]/mod2[["coefficients"]][["treatment"]]
Dia = ITTmodels[["Diabetes"]][["coefficients"]][["treatment"]]/mod2[["coefficients"]][["treatment"]]
Rx = ITTmodels[["RX Prescriptions"]][["coefficients"]][["treatment"]]/mod2[["coefficients"]][["treatment"]]


tibble(H,Dep,Dia,Rx,title = "ATET Effects")
```

```
## # A tibble: 1 x 5
##         H    Dep    Dia    Rx title
##     <dbl>  <dbl>  <dbl> <dbl> <chr>
## 1 0.00747 0.0185 0.0321 0.417 ATET Effects
```

#8: I am concerned that attrition bias happened because the compliance rate was extremely low. We should look to see if those who complied in the treatment group were more likely to fit some demographic category than not. I will use the treatment data subset and regress medicaid enrollment on demographic characteristics.

```
Mod4 = lm(ohp_all_ever_survey~age_inp+race_white_inp+race_black_inp+gender_inp+
            itvw_english_inp+dia_dx_pre_lottery,data = treat)
msummary(Mod4,stars=TRUE,title = "Attrition Bias?")
```

Attrition Bias?

| | **Model 1** |
|---|---|
| (Intercept) | 1.092*** |
| | (0.056) |
| age_inp | 0.000 |
| | (0.001) |
| race_white_inp | 0.036* |
| | (0.018) |
| race_black_inp | 0.086*** |

| | |
|---|---|
| | (0.024) |
| gender_inp | 0.076*** |
| | (0.012) |
| itvw_english_inp | 0.063** |
| | (0.022) |
| dia_dx_pre_lottery | -0.026 |
| | (0.024) |
| Num.Obs. | 6364 |
| R2 | 0.012 |
| R2 Adj. | 0.011 |
| AIC | 8977.9 |
| BIC | 9032.0 |
| Log.Lik. | -4480.957 |
| F | 12.806 |
| RMSE | 0.49 |

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

#There seems to be some systematic attrition across demographic characteristics. Those who are Black and pre-diagnosed with diabetes seem to go through with enrollment more ofter as well as those who did not interview in English.

#In 2008, Oregon found funds to expand Medicaid to some ten thousand participants and the other tens of thousands without the expansion made for a social scientists dream of perfect conditions to conduct a randomized controlled experiment. Holding a lottery, participants were randomly selected and according to our research, no selection bias was detected. We found that treatment status led to a significant increase in access to medication. Additionally, I found the “Treatment on the Treated” effect of increased depression diagnosis of 1.8% and diabetes diagnosis of 3%, which is a socially significant increase. This increase in diagnosis is actually positive because these patients are actually seen and treated. Overall, this experiment falls just short of platinum but stands strongly at a gold standard, the only downfall being a small compliance rate. Very little attrition bias was found however, and the main findings of this study should not be discarded.