Matt McCoy

EC 525

Prof: Eric Zou

4/29/21

**Empirical Project 1 Research Summary**

**Abstract**

In 2008, low-income and uninsured adults in Oregon were selected for the opportunity to apply for Medicaid through a lottery, and this lottery set the grounds for this randomized design. This lottery, along with some administrative data, allows us to evaluate the effects of Medicaid on health care usage like doctor visits and the health outcomes of low-income adults using a randomized control experiment. After treatment we found that the treatment group was about 25.4 % more likely to have insurance than the control group. We also found that the treatment group had significantly higher health care utilization, measured by doctor visits. We saw an increase of 156% for the number of annual doctor visits. They also self-reported improved mental and physical compared to the control group. Medicaid caused decreases in bp and cholesterol, and increased the diagnosis of depression and diabetes. We also found an increase in the number of prescriptions by about 50%.

The difference between the variables treatment and ohp_all_ever_survey is that treatment indicates whether the individual has been treated, which means they won the lottery and accepted the insurance, or if they are not treated. And ohp_all_ever_survey only tells us if the individual has ever been enrolled in medicaid during the period this study took place. It was possible for people in the OHP study to also qualify for medicaid through other means. Treatment is the treatment variable because they are using an intent to treat model to compare outcomes for everyone that got selected in the lottery with everyone who was on the list but was not selected. The variable ohp_all_ever_survey does not tell us if an individual won the lottery and was treated or not, so it does not make sense for it to be the treatment variable.

```
table1(~  edu_inp + age_inp + gender_inp + hispanic_inp + race_black_inp + race_white_inp | treatment, data=ohp)
```

```
## Warning in table1.formula(~edu_inp + age_inp + gender_inp + hispanic_inp + :
## Terms to the right of '|' in formula 'x' define table columns and are expected
## to be factors with meaningful labels.
```

| | 0<br>(N=5842) | 1<br>(N=6387) | Overall<br>(N=12229) |
|---|---|---|---|
| **Education: highest completed** | | | |
| Mean (SD) | 2.24 (0.904) | 2.26 (0.912) | 2.25 (0.908) |
| Median [Min, Max] | 2.00 [1.00, 4.00] | 2.00 [1.00, 4.00] | 2.00 [1.00, 4.00] |
| Missing | 3 (0.1%) | 8 (0.1%) | 11 (0.1%) |
| **Age - inperson survey** | | | |
| Mean (SD) | 40.6 (11.7) | 41.0 (11.7) | 40.8 (11.7) |
| Median [Min, Max] | 41.0 [19.0, 68.0] | 41.0 [20.0, 71.0] | 41.0 [19.0, 71.0] |
| Missing | 1 (0.0%) | 0 (0%) | 1 (0.0%) |
| **Gender - inperson survey** | | | |
| Mean (SD) | 0.569 (0.496) | 0.563 (0.496) | 0.566 (0.496) |
| Median [Min, Max] | 1.00 [0, 2.00] | 1.00 [0, 1.00] | 1.00 [0, 2.00] |
| **Hispanic/Latino** | | | |
| Mean (SD) | 0.178 (0.382) | 0.182 (0.386) | 0.180 (0.384) |
| Median [Min, Max] | 0 [0, 1.00] | 0 [0, 1.00] | 0 [0, 1.00] |
| Missing | 10 (0.2%) | 19 (0.3%) | 29 (0.2%) |
| **Race/Ethnicity is Black** | | | |
| Mean (SD) | 0.107 (0.309) | 0.0996 (0.300) | 0.103 (0.304) |
| Median [Min, Max] | 0 [0, 1.00] | 0 [0, 1.00] | 0 [0, 1.00] |
| Missing | 16 (0.3%) | 23 (0.4%) | 39 (0.3%) |
| **Race/Ethnicity is White** | | | |
| Mean (SD) | 0.690 (0.463) | 0.687 (0.464) | 0.688 (0.463) |
| Median [Min, Max] | 1.00 [0, 1.00] | 1.00 [0, 1.00] | 1.00 [0, 1.00] |
| Missing | 16 (0.3%) | 23 (0.4%) | 39 (0.3%) |

Yes, the data is consistent with individuals having been randomly assigned to treatment groups and control groups. We can see that the variable means are all very similar so we know the individual characteristics of the control and treatment groups are nearly identical. For example, we can see that the mean age in control is 40.6 and the mean age in treatment is 41. This is only a difference of .4 years. I believe you want characteristics like education, age, gender, and race to all look nearly identical between a treatment and control group. This will help let us know if the two groups are comparable on average.

```r
first_table <- left_join(table2, merge, by = "variable")

huxtable(first_table)
```

| treatment | variable | mean | difference | std.error |
|---|---|---|---|---|
| 0 | education | 2.24 | 0.0217 | 0.0164 |
| 0 | age | 40.6 | 0.38 | 0.212 |
| 0 | gender | 0.569 | -0.00611 | 0.00898 |
| 0 | hispanic | 0.178 | 0.00468 | 0.00696 |
| 0 | black | 0.107 | -0.00765 | 0.00552 |
| 0 | other | 0.142 | 0.00336 | 0.00637 |
| 0 | white | 0.69 | -0.00301 | 0.0084 |
| 1 | education | 2.26 | 0.0217 | 0.0164 |
| 1 | age | 41 | 0.38 | 0.212 |
| 1 | gender | 0.563 | -0.00611 | 0.00898 |
| 1 | hispanic | 0.182 | 0.00468 | 0.00696 |
| 1 | black | 0.0996 | -0.00765 | 0.00552 |
| 1 | other | 0.146 | 0.00336 | 0.00637 |
| 1 | white | 0.687 | -0.00301 | 0.0084 |

The balance table appears to show that individuals have indeed been randomly assigned to treatment and control groups. The difference between the means of most of the variable is smaller than two decimal places, so the differences are very small and the standard error for them is also less than two decimal spaces indicating high statistical accuracy of our estimates. This is untrue for age in particular and education to an extent. The difference in education is about .022 which is small and less that a one level increase in education. Both groups have very similar education on average but the standard error is slightly higher than most other variables. Age was the most inaccurate variable with a much higher standard error of .212. The difference in age between the two groups was the largest of any variable at .38. This is still a small difference of less than half a year. I believe the balance table is consistent with individuals having been randomly assigned to treatment and control because the descriptive variables have small differences with high accuracy. The treatment and control groups have nearly identical characteristics on average.

```
```{r compliance rate}
mod2 <- lm(ohp_all_ever_survey ~ treatment, data = ohp)
stargazer(mod2, type = 'text')
#compliance rate = .254 or 25.4%
```
```

```
length of NULL cannot be changedlength of NULL cannot
================================================
                        Dependent variable:
                    ----------------------------
                        ohp_all_ever_survey
------------------------------------------------
treatment                     0.254***
                              (0.008)

Constant                      0.158***
                              (0.006)

------------------------------------------------
Observations                   12,229
R2                             0.078
Adjusted R2                    0.078
Residual Std. Error      0.436 (df = 12227)
F Statistic         1,031.581*** (df = 1; 12227)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

The Compliance Rate is the fraction of treatment group units receiving treatment minus the fraction of control group units receiving treatment. In randomized experiments, compliance can be a problem as we cannot force people to comply with treatments. We can only offer them a treatment. We can see that the compliance rate in this study is 0.254 or 25.4%, so winning the ohp lottery increases the probability of being enrolled in medicaid by 25.4%.

```
stargazer(bp_mod, visits_mod, rx_mod, dep_mod, chol_mod, dia_mod, hbp_mod, type = 'text')
```

```
the condition has length > 1 and only the first element will be usedthe condition has length > 1 and only the first element will be usedlength of NULL cannot be changedlength of NU
changedlength of NULL cannot be changedlength of NULL cannot be changedlength of NULL cannot be changed
```

| | Dependent variable: | | | | | | |
|---|---|---|---|---|---|---|---|
| | bp_sar_inp | doc_num_mod_inp | rx_num_mod_inp | dep_difference | chl_inp | dia_difference | hbp_difference |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| treatment | -0.058 | 0.396* | 0.128** | 0.023** | -0.642 | 0.009* | 0.004 |
| | (0.300) | (0.216) | (0.053) | (0.010) | (0.613) | (0.005) | (0.009) |
| Constant | 119.130*** | 5.746*** | 1.838*** | -0.305*** | 205.769*** | -0.060*** | -0.130*** |
| | (0.217) | (0.156) | (0.038) | (0.007) | (0.443) | (0.004) | (0.006) |
| Observations | 12,188 | 12,158 | 11,912 | 12,095 | 12,174 | 12,186 | 11,945 |
| R2 | 0.00000 | 0.0003 | 0.0005 | 0.0004 | 0.0001 | 0.0002 | 0.00001 |
| Adjusted R2 | -0.0001 | 0.0002 | 0.0004 | 0.0003 | 0.00001 | 0.0002 | -0.0001 |
| Residual Std. Error | 16.550 (df = 12186) | 11.895 (df = 12156) | 2.891 (df = 11910) | 0.557 (df = 12093) | 33.792 (df = 12172) | 0.292 (df = 12184) | 0.478 (df = 11943) |
| F Statistic | 0.038 (df = 1; 12186) | 3.357* (df = 1; 12156) | 5.855** (df = 1; 11910) | 4.939** (df = 1; 12093) | 1.097 (df = 1; 12172) | 3.017* (df = 1; 12184) | 0.170 (df = 1; 11943) |
| Note: | | | | | | | *p<0.1; **p<0.05; ***p<0.01 |

ITT is the effect of being selected for treatment. The ITT regressions are in the table and to see the effect of OHP on people's actual health we wanted to see how being in the OHP experiment affected the health of the individuals. The effect of the OHP experiment on the number of doctor visits, prescriptions, depression, and diabetes were all significant.

```{r, eval = FALSE}
atet <- bind_rows(bp_mod, visits_mod, rx_mod, dep_mod, chol_mod, dia_mod, hbp_mod)

merge2 <- atet %>%
  filter(term == "treatment") %>%
  select(variable, estimate, std.error)

atet_final <- merge2 %>% mutate(ATET = estimate/.254)
huxtable(atet_final)
```

| variable | estimate | std.error | ATET |
|---|---|---|---|
| bp | -0.0583 | 0.3 | -0.229 |
| visits | 0.396 | 0.216 | 1.56 |
| prescriptions | 0.128 | 0.053 | 0.505 |
| depression | 0.0225 | 0.0101 | 0.0887 |
| cholesterol | -0.642 | 0.613 | -2.53 |
| diabetes | 0.00918 | 0.00529 | 0.0362 |
| hypertension | 0.00361 | 0.00875 | 0.0142 |

Column names: variable, estimate, std.error, ATET

ATET is the effect of having Medicaid. The causal effect of Medicaid on health is displayed for each variable in the table, and we can see there were varying effects across variables. For example, being on Medicaid causes a given individual to increase annual doctor visits by 156%. ATET is the true effect of being treated when there is non-compliance, which occurs when an individual does not take up the treatment even if they are offered one or when they take up the treatment after not being offered one. When this occurs ATET is equal to the ITT divided by the compliance rate (ATET = ITT/Compliance Rate). If there is no non-compliance issue, then ITT=ATET. For this study, we must adjust for non-compliance.

Attrition within the study population may contaminate the random design, if people drop out of the treatment and control groups at different rates. If respondents in the control and treatment groups differ, then the comparison is irrelevant. This is a particular concern for those surveyed by mail, whose response rates are lowest. I do not believe that this was an issue in analyzing this data, but it is hard to tell as we do not know how many people dropped out of the experiment in each group. We should certainly worry about the possibility of attrition bias in a study like this. If there was attrition in this study, I would expect it to primarily be those who did not win the lottery as they will not receive health insurance and therefore have less to gain from completing the experiment.

*Abstract: Placed at the top of the paper*