

# Problem Set 4

Petra - pi2018 - Section 005

Due Dec 13, 2023

This homework must be turned in on Brightspace by Dec. 13 2023. It must be your own work, and your own work only – you must not copy anyone’s work, or allow anyone to copy yours. This extends to writing code. You may consult with others, but when you write up, you must do so alone.

Your homework submission must be written and submitted using Rmarkdown. No handwritten solutions will be accepted. **No zip files will be accepted. Make sure we can read each line of code in the pdf document.** You should submit the following:

1. A compiled PDF file named yourNetID\_solutions.pdf containing your solutions to the problems.
2. A .Rmd file containing the code and text used to produce your compiled pdf named your-NetID\_solutions.Rmd.

Note that math can be typeset in Rmarkdown in the same way as Latex. Please make sure your answers are clearly structured in the Rmarkdown file:

1. Label each question part
2. Do not include written answers as code comments.
3. The code used to obtain the answer for each question part should accompany the written answer. Comment your code!

## Problem 1 (100 points)

Despite the heated political and media rhetoric, there are a few causal estimates of the effect of expanded health insurance on healthcare outcomes. One landmark study, the Oregon Health Insurance Experiment, covered new ground by utilizing a randomized control trial implemented by the state of Oregon. To allocate a limited number of eligible coverage slots for the state's Medicaid expansion, about 30,000 low-income, uninsured adults (out of about 90,000 wait-list applicants) were randomly selected by lottery to be allowed to apply for Medicaid coverage. Researchers collected observable measure of health (blood pressure, cholesterol, blood sugar levels, and depression), as well as hospital visitations and healthcare expenses for 6,387 selected adults and 5,842 not selected adults.

For this problem, we will use the OHIE.dta file.

- treatment - selected in the lottery to sign up for Medicaid (instrument)
- ohp\_all\_ever\_admin - Ever enrolled in Medicaid after notification of lottery results (compliance)
- tab2bp\_hyper - Outcome: Binary indicator for elevated blood pressure (1 indicates a high blood pressure)
- tab2phqtot\_high - Outcome: Binary indicator for depression
- tab4\_catastrophic\_exp\_inp - Outcome: Indicator for catastrophic medical expenditure (1 if their total out-of-pocket medical expenses are larger than 30% of their household income)
- tab5\_needmet\_med\_inp - Outcome: Binary indicator of whether the participant feels that they received all needed medical care in past 12 months

```
# Load in the data
data <- haven::read_dta("OHIE.dta")

head(data)

## # A tibble: 6 x 59
##   weight_total_inp tab1_gender_inp tab2dia_dx_post_lott~1 tab2hbp_dx_post_lott~2
##           <dbl> <dbl+lbl>         <dbl+lbl>         <dbl+lbl>
## 1           1.15      1 [Female]           0 [No]           0 [No]
## 2           0.897      0 [Male]           0 [No]           1 [Yes]
## 3              0      NA              NA              NA
## 4              1      1 [Female]           0 [No]           0 [No]
## 5           1.21      0 [Male]           0 [No]           0 [No]
## 6              1      0 [Male]           0 [No]           0 [No]
## # i abbreviated names: 1: tab2dia_dx_post_lottery, 2: tab2hbp_dx_post_lottery
## # i 55 more variables: tab2chl_dx_post_lottery <dbl+lbl>,
## #   tab2dep_dx_post_lottery <dbl+lbl>, tab3_pcs8_score <dbl>,
## #   tab3_mcs8_score <dbl>, tab5_usual_clinic_inp <dbl+lbl>,
## #   tab5_needmet_med_inp <dbl+lbl>, tab5_chl_chk_inp <dbl+lbl>,
## #   tab5_pap_chk_inp <dbl+lbl>, tab5_fobt_chk_inp <dbl+lbl>,
## #   tab5_col_chk_inp <dbl+lbl>, tab5_psa_chk_inp <dbl+lbl>, ...
```

**Hint:** This was an experiment with imperfect compliance. Instead of creating a “participated” or “complied” variable, simply use “treatment” as the instrument and “ohp\_all\_ever\_admin” (enrollment in Medicaid) as the main independent variable of interest.

## Question A (25 points)

Estimate the intent-to-treat effects of being selected to sign up for Medicaid on each of the four outcomes (elevated blood pressure, depression, catastrophic medical expenditure, and whether respondents had their

health care needs met). Provide 95% confidence intervals for each estimate and interpret your results. (Use `lm_robust`)

```
# Estimate the ITT on elevated blood pressure
itt_bp <- lm_robust(tab2bp_hyper ~ treatment, data = data)
summary(itt_bp)

##
## Call:
## lm_robust(formula = tab2bp_hyper ~ treatment, data = data)
##
## Standard error type: HC2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)   0.1591    0.00480  33.182 3.21e-231  0.1497  0.1685 12186
## treatment    -0.0016    0.00662  -0.242 8.09e-01 -0.0146  0.0114 12186
##
## Multiple R-squared:  4.8e-06 , Adjusted R-squared: -7.73e-05
## F-statistic: 0.0584 on 1 and 12186 DF, p-value: 0.809
```

We can see that our treatment coefficient is negative at -0.0016 with a standard error of 0.0066. This would indicate that being selected to sign up for Medicaid is negatively associated with likelihood of having elevated blood pressure. However, we can see that our p-value for our treatment is quite high (much higher than general threshold of 0.05) at 0.809. This indicates that being selected to sign up for Medicaid does not have a statistically significant effect on the likelihood of having elevated blood pressure. Further looking at our 95% confidence interval of [-0.0146, 0.0114] we can notice that it includes zero, reinforcing the lack of a significant effect that we mentioned.

```
# Estimate the ITT on depression
itt_depression <- lm_robust(tab2phqtot_high ~ treatment, data = data)
summary(itt_depression)

##
## Call:
## lm_robust(formula = tab2phqtot_high ~ treatment, data = data)
##
## Standard error type: HC2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)   0.3037    0.00603  50.33 0.00e+00  0.292  0.3155 12159
## treatment    -0.0349    0.00821  -4.26 2.09e-05 -0.051 -0.0188 12159
##
## Multiple R-squared:  0.00149 , Adjusted R-squared:  0.00141
## F-statistic: 18.1 on 1 and 12159 DF, p-value: 2.09e-05
```

We can see that our treatment coefficient is negative at -0.0349 with a standard error of 0.0082. We can also see that our p-value is sufficiently small at 2.09e-05 and that our 95% confidence interval [-0.051, -0.0188] does not contain 0 indicating a statistically significant association. This shows us that being selected to sign up for Medicaid is associated with a statistically significant decrease in the likelihood of depression. The fact that our coefficient is negative suggests a reduction in depression rates among those selected for Medicaid.

```
# Estimate the ITT on catastrophic expenditures
itt_catastrophic <- lm_robust(tab4_catastrophic_exp_inp ~ treatment, data = data)
summary(itt_catastrophic)
```

```
##
## Call:
## lm_robust(formula = tab4_catastrophic_exp_inp ~ treatment, data = data)
##
## Standard error type: HC2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)   0.0538    0.00300   17.92 6.77e-71  0.0479  0.05971 11793
## treatment    -0.0153    0.00388   -3.94 8.34e-05 -0.0229 -0.00766 11793
##
## Multiple R-squared:  0.00133 ,    Adjusted R-squared:  0.00124
## F-statistic: 15.5 on 1 and 11793 DF,  p-value: 8.34e-05
```

We can see that our treatment coefficient is negative at -0.0153 with a standard error of 0.00388. We can also see that our p-value is sufficiently small at 8.34e-05 and that our 95% confidence interval [-0.0229, -0.00766] does not contain 0 indicating a statistically significant association. This shows us that being selected to sign up for Medicaid is associated with a statistically significant decrease in the likelihood of experiencing catastrophic medical expenditures. This also means that Medicaid selection can help reduce the financial burden of medical expenses, as we can expect.

```
# Estimate the ITT on "needs met"
itt_needs_met <- lm_robust(tab5_needmet_med_inp ~ treatment, data = data)
summary(itt_needs_met)
```

```
##
## Call:
## lm_robust(formula = tab5_needmet_med_inp ~ treatment, data = data)
##
## Standard error type: HC2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
## (Intercept)   0.6124    0.00638   96.02 0.00e+00  0.5999  0.6249 12214
## treatment     0.0345    0.00875    3.94 8.19e-05  0.0173  0.0516 12214
##
## Multiple R-squared:  0.00127 ,    Adjusted R-squared:  0.00119
## F-statistic: 15.5 on 1 and 12214 DF,  p-value: 8.19e-05
```

We can see that our treatment coefficient is positive at 0.0345 with a standard error of 0.00875. We can also see that our p-value is sufficiently small at 8.19e-05 and that our 95% confidence interval [0.0173, 0.0516] does not contain 0 indicating a statistically significant association. This shows us that being selected to sign up for Medicaid is associated with a statistically significant increase in the likelihood of a person feeling that their health care needs were/are met. This means that there is a positive impact of Medicaid selection on whether participants believe they received all needed medical care.

Looking at all 4 findings we can say that our analysis shows that being selected for Medicaid has no significant effect on elevated blood pressure but does have significant effect on reducing depression, reducing likelihood of catastrophic expenditures and increasing the likelihood of people feeling their healthcare needs are sufficiently met.

## Question B (25 points)

Suppose that researchers actually wanted to estimate the effect of Medicaid enrollment (ohp\_all\_ever\_admin) on each of the four outcomes. Suppose they first used a naive regression of each of the the outcomes on the indicator of Medicaid enrollment. Report a 95% confidence interval for each of your estimates and interpret your results. Why might these be biased estimates for the causal effect of Medicaid enrollment?

```
# Estimate the Naive OLS effect on elevated blood pressure
naive_bp <- lm_robust(tab2bp_hyper ~ ohp_all_ever_admin, data = data)
tidy(naive_bp)
```

```
##           term estimate std.error statistic p.value conf.low conf.high
## 1 (Intercept)   0.1634   0.00397    41.21 0.0000   0.1557   0.17122
## 2 ohp_all_ever_admin -0.0181   0.00716    -2.52 0.0117  -0.0321  -0.00401
##      df      outcome
## 1 12186 tab2bp_hyper
## 2 12186 tab2bp_hyper
```

From above output we can see that the estimate for the Medicaid enrollment is -0.0181 with a 95% confidence interval of [-0.0321, -0.00401]. We can also see that the p-value is sufficiently small and as the confidence interval does not contain 0 we can say that the effect is statistically significant. This means that the negative coefficient suggests that Medicaid enrollment is associated with a decrease in the likelihood of having elevated blood pressure. It is worth mentioning that while the effect is statistically significant it is very small and with a very narrow CI which means the association is modest at best.

```
# Estimate the Naive OLS effect on depression
naive_depression <- lm_robust(tab2phqtot_high ~ ohp_all_ever_admin, data = data)
tidy(naive_depression)
```

```
##           term estimate std.error statistic p.value conf.low conf.high
## 1 (Intercept)   0.2713   0.00477    56.83 0.00e+00   0.2619   0.2806
## 2 ohp_all_ever_admin 0.0493   0.00924     5.34 9.52e-08   0.0312   0.0674
##      df      outcome
## 1 12159 tab2phqtot_high
## 2 12159 tab2phqtot_high
```

From above output we can see that the estimate for the Medicaid enrollment is 0.0493 with a 95% confidence interval of [0.0312, 0.0674]. We can also see that the p-value is sufficiently small and as the confidence interval does not contain 0 we can say that the effect is statistically significant. This means that the positive coefficient suggests that Medicaid enrollment is associated with an increase in the likelihood of having depression. We can see that this result is unexpected and counterintuitive as we would expect that the person who got enrolled in Medicaid has more resources for treating depression. However, it can also be due to the fact that having Medicaid allows people to be diagnosed with depression which they could not be if they did not have that resource. As we cannot know which is the case we should further investigate this.

```
# Estimate the Naive OLS effect on catastrophic expenditures
naive_catastrophic <- lm_robust(tab4_catastrophic_exp_inp ~ ohp_all_ever_admin, data = data)
tidy(naive_catastrophic)
```

```
##           term estimate std.error statistic p.value conf.low conf.high
## 1 (Intercept)   0.0489   0.00235    20.81 1.66e-94   0.0443   0.05355
## 2 ohp_all_ever_admin -0.0107   0.00405    -2.65 8.13e-03  -0.0187  -0.00278
```

```
##           df           outcome
## 1 11793 tab4_catastrophic_exp_inp
## 2 11793 tab4_catastrophic_exp_inp
```

From above output we can see that the estimate for the Medicaid enrollment is -0.0107 with a 95% confidence interval of [-0.0187, -0.00278]. We can also see that the p-value is sufficiently small and as the confidence interval does not contain 0 we can say that the effect is statistically significant. This means that the negative coefficient suggests that Medicaid enrollment is associated with a decrease in the likelihood of experiencing catastrophic medical expenditures. This is expected as Medicaid will indeed cover some of if not all medical expenses.

```
# Naive OLS estimate on needs met
naive_needs_met <- lm_robust(tab5_needmet_med_inp ~ ohp_all_ever_admin, data = data)
tidy(naive_needs_met)
```

```
##           term estimate std.error statistic  p.value conf.low conf.high
## 1 (Intercept)   0.6128   0.00522    117.40 0.00e+00   0.6026   0.6230
## 2 ohp_all_ever_admin 0.0613   0.00948     6.46 1.08e-10   0.0427   0.0799
##           df           outcome
## 1 12214 tab5_needmet_med_inp
## 2 12214 tab5_needmet_med_inp
```

From above output we can see that the estimate for the Medicaid enrollment is 0.0613 with a 95% confidence interval of [0.0427, 0.0799]. We can also see that the p-value is sufficiently small and as the confidence interval does not contain 0 we can say that the effect is statistically significant. This means that the positive coefficient suggests that Medicaid enrollment is associated with an increase in the likelihood of people feeling their health care needs were met. This is expected as Medicaid will can be used pay for their medical needs which they most likely could not afford prior.

Unlike in part A we can see that all the variables are statistically significant however there are some counter-intuitive points like the one with depression and Medicaid enrollment.

### Why might these be biased estimates for the causal effect of Medicaid enrollment?

In addition to general issues with measurements errors and unobserved confounders in our case we need to also consider the fact that Medicaid enrollment was a second step in the process - only those who have already won the lottery had the chance to enroll so it is not as representative of general population. While the lottery system being randomized helps with controlling for unobserved confounders when choosing who will get the ability to enroll into Medicaid, if we don't account for this and just do simple OLS regression we lose on the benefit of having a random lottery. Due to this, we can say that this randomized lottery is our instrumental variable. As mentioned, we are facing a selection bias in enrollment as only those chosen by lottery can enroll but we are also facing an issue with imperfect compliance - not all chosen to enroll will actually do it. This means that there might be (and will) be cases of people who were chosen to enroll but they have not done so for different personal reasons. Further, there might also be cases of people that got to enroll from other sources but the lottery. Due to this imperfect compliance and selection bias, the simple OLS regression might not be good enough for making valuable predictions. As seen in class, using the instrumental variable like two-stage least squares or using `iv_robust` to calculate the effect might be a better approach as it will hopefully isolate the effect of our Medicaid enrollment from other confounders we might be facing.

### Question C (25 points)

Suppose we were to use assignment to treatment as an instrument for actually receiving Medicaid coverage.

Consider that not everyone who was selected to apply for Medicaid actually ended up applying and receiving coverage. Likewise, some applicants who were not selected to receive the treatment nevertheless were eventually covered. What were the compliance rates (the level of Medicaid enrollment) for subjects who were selected and subjects who were not selected? Use a “first stage” regression to estimate the effect of being selected on Medicaid enrollment to estimate the compliance rates. Is the instrument of assignment-to-treatment a strong instrument for actual Medicaid enrollment?

```
# First Stage OLS
```

```
first_stage=lm_robust(ohp_all_ever_admin ~ treatment, data=data, se_type = "stata")
tidy(first_stage)
```

```
##           term estimate std.error statistic p.value  conf.low  conf.high    df
## 1 (Intercept)   0.145   0.00347     42.0      0      0.139    0.152 20743
## 2  treatment    0.236   0.00589     40.1      0      0.225    0.248 20743
##           outcome
## 1 ohp_all_ever_admin
## 2 ohp_all_ever_admin
```

```
# null model (compliance given an intercept only model)
```

```
null_mod<-lm_robust(ohp_all_ever_admin ~ 1, data=data)
tidy(null_mod)
```

```
##           term estimate std.error statistic p.value  conf.low  conf.high    df
## 1 (Intercept)   0.264   0.00306     86.3      0      0.258    0.27 20744
##           outcome
## 1 ohp_all_ever_admin
```

```
# F - Stat for Instrument Strength (use waldtest)
```

```
first_stage$fstatistic
```

```
## value numdf dendif
## 1610      1 20743
```

```
waldtest(first_stage, null_mod)
```

```
## Wald test
##
## Model 1: ohp_all_ever_admin ~ treatment
## Model 2: ohp_all_ever_admin ~ 1
##   Res.Df Df Chisq Pr(>Chisq)
## 1  20743
## 2  20744 -1  1610    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Looking at the First Stage OLS above we can see that the intercept estimate is 0.145. This means when not receiving treatment (winning the lottery) 14.5% of people will still get to enroll in Medicaid. We can also see that the treatment estimate is 0.236 which indicates that out of those that won the lottery 23.6% will also enroll in Medicaid. As this is quite a large percentage compared to our intercept, this indicates that effectiveness of the lottery selection in strongly influencing Medicaid enrollment. Further, as we want to show total compliance rates, we can say that out of those not winning the lottery 14.5% still enrolled in

Medicaid, while out of those who did win the lottery 38.1% (14.5% + 23.6%) enrolled in Medicaid. As we can see that out p-value for both of this estimates is 0, 0 is not in either of their CI and they both have low standard errors we can say that this effect is statistically significant.

It is also worth mentioning that when looking at our null model - not accounting for treatment variable (lottery results) - we can see that our intercept estimate is 0.264 which indicates 26.4% of all population (in our data) enrolled into Medicaid regardless of the lottery results. As this estimate also has 0 p-value and no 0 in the confidence interval, and low standard error we can say that this result is also statistically significant.

Above we can see that our F-statistic is quite high with value of 1610 and that our Wald test comparing the First Stage OLS model against the Null Model yields the same chi-squared statistic of 1610 with very low p-value result. This results indicate that our instrument (assignment to treatment) is very significantly associated with the Medicaid enrollment. This means that the instrument can be used in our IV analysis to be able to estimate the causal effect of Medicaid enrollment on our 4 different outcomes.

## Question D (25 points)

Now estimate the effect of Medicaid enrollment on each of the four outcomes using an instrumental variables strategy. Report a 95% confidence interval for your estimates and interpret your results. Compare the estimates to those you obtained in Question B.

```
# Estimate the IV effect on elevated blood pressure (use iv_robust())
iv_bp <- iv_robust(tab2bp_hyper ~ ohp_all_ever_admin | treatment, data = data,
                  se_type = "stata", diagnostics = TRUE)
tidy(iv_bp)
```

	term	estimate	std.error	statistic	p.value	conf.low	conf.high
## 1	(Intercept)	0.1601	0.00818	19.567	5.71e-84	0.1440	0.1761
## 2	ohp_all_ever_admin	-0.0063	0.02606	-0.242	8.09e-01	-0.0574	0.0448
##	df	outcome					
## 1	12186	tab2bp_hyper					
## 2	12186	tab2bp_hyper					

From the above output we can see that the IV estimate for Medicaid enrollment on elevated blood pressure of -0.0063 is not statistically significant as the p-value is 0.8 (larger than the general threshold of 0.05) and the 95% confidence interval of [-0.0574, 0.0448] includes zero. This indicates that Medicaid enrollment does not have a clear (statistically significant) impact on elevated blood pressure.

These results are different than those we got from Naive OLS as in that case estimate was also small but significant. In this analysis, IV estimate is not statistically significant which indicates that the Naive OLS estimate may have been biased (as it did not take into account treatment (lottery result) variable) and the true effect of Medicaid enrollment on blood pressure is unclear.

```
# Estimate the IV effect on depression
iv_depression <- iv_robust(tab2phqtot_high ~ ohp_all_ever_admin | treatment,
                          data = data, se_type = "stata", diagnostics = TRUE)
tidy(iv_depression)
```

	term	estimate	std.error	statistic	p.value	conf.low	conf.high
## 1	(Intercept)	0.325	0.0104	31.26	2.38e-206	0.304	0.3452
## 2	ohp_all_ever_admin	-0.138	0.0329	-4.19	2.84e-05	-0.202	-0.0732
##	df	outcome					
## 1	12159	tab2phqtot_high					
## 2	12159	tab2phqtot_high					



From the above output we can see that the IV estimate for Medicaid enrollment on depression is negative, with an estimated effect of -0.138 and a 95% confidence interval of [-0.202, -0.0732] which does not contain 0 and thus we can say that the estimate is statistically significant. This indicates that Medicaid enrollment is associated with a decrease in the likelihood of depression as we would expect it to be.

This result is quite different than the Naive OLS result in Part B as in that analysis we found that Medicaid enrollment was associated with an increased likelihood of having depression which was quite counter intuitive. We also found that this relationship was statistically significant. In IV analysis we found that IV estimate was also statistically significant but in opposite direction - it indicates that Medicaid enrollment is associated with a *decrease* in the likelihood of depression, which is more intuitive as people with Medicaid have more resources to treat the depression or any other medical issues. This inaccuracy and bias in Naive OLS result may be due to the fact that it fails to account for the causal effect that is accounted for in IV analysis by taking into consideration treatment variable (lottery results).

```
# Estimate the IV effect on catastrophic expenditures
iv_catastrophic <- iv_robust(tab4_catastrophic_exp_inp ~ ohp_all_ever_admin | treatment,
                             data = data, se_type = "stata", diagnostics = TRUE)
tidy(iv_catastrophic)
```

```
##               term estimate std.error statistic  p.value conf.low conf.high
## 1      (Intercept)   0.0631   0.00508     12.43 3.03e-35   0.0532   0.0731
## 2 ohp_all_ever_admin -0.0604   0.01543     -3.91 9.16e-05  -0.0906  -0.0301
##      df      outcome
## 1 11793 tab4_catastrophic_exp_inp
## 2 11793 tab4_catastrophic_exp_inp
```

From the above output we can see that the IV estimate for Medicaid enrollment on catastrophic expenditures is negative, with an estimated effect of -0.0604 and a 95% confidence interval of [-0.0906, -0.0301] which does not contain 0 and thus we can say that the estimate is statistically significant. This indicates that Medicaid enrollment is associated with a decrease in the likelihood of experiencing catastrophic medical expenditures as we would expect since these expenditures are at least partially covered by Medicaid when enrolled.

These results from the IV analysis are similar to those in part B from the Naive OLS. Just like in the Naive OLS, the association is negative and significant which indicates that Medicaid enrollment may protect against catastrophic expenditures. In the Naive OLS, the effect was much less negative (6 times less) which indicates that in the IV results having enrolled in Medicaid decreases the chance of experiencing catastrophic medical expenditures much more than we calculate in Naive OLS.

```
# IV estimate on needs met
iv_needs_met <- iv_robust(tab5_needmet_med_inp ~ ohp_all_ever_admin | treatment,
                          data = data, se_type = "stata", diagnostics = TRUE)
tidy(iv_needs_met)
```

```
##               term estimate std.error statistic  p.value conf.low conf.high
## 1      (Intercept)   0.592   0.0109     54.45 0.00e+00   0.570   0.613
## 2 ohp_all_ever_admin   0.135   0.0344      3.94 8.35e-05   0.068   0.203
##      df      outcome
## 1 12214 tab5_needmet_med_inp
## 2 12214 tab5_needmet_med_inp
```

From the above output we can see that the IV estimate for Medicaid enrollment on the likelihood of respondents feeling their healthcare needs were met is positive, with an estimated effect of 0.135 and a 95% confidence interval of [0.068, 0.203] which does not contain 0 and thus we can say that the estimate is statistically significant. This indicates that Medicaid enrollment is associated with an increased likelihood of

individuals feeling that their medical needs are being met which again is expected as they will have more opportunity to get medical attention when enrolled into Medicaid which will subsidise the cost of their necessary treatment.

These results from the IV analysis are similar to those in part B from the Naive OLS. Just like in the Naive OLS, the association is positive and significant which indicates that Medicaid enrollment may improve individuals' perceptions of their healthcare needs being fulfilled. In the Naive OLS, the effect was much smaller at 0.06 while in IV analysis it was much higher at 0.14 which indicates that people having improved perception of healthcare is much likelier than we have expected when looking only at Naive OLS.

Looking at all of these results and comparing IV estimates to those we got from the Naive OLS we can see that while most relationships were in the same direction (except for the depression variable) there are some significant differences. The original Naive OLS OLS estimates could have been biased due to issues we mentioned before such as unobserved confounders and the fact that Medicaid enrollment is dependent on winning the lottery first which introduces a selection bias, as the sample we are working with may not be representative of the general population. Due to this, in the IV analysis approach we use lottery as an instrument so we can address this selection bias by isolating the effect of Medicaid enrollment from other confounders. The significant differences between the Naive OLS and IV estimates for some variables of interest, like depression, suggest that the Naive OLS may have been biased. As we learned in class the IV estimates are generally more reliable for causal inference if we know our instrument is good. On that note, large F-statistic for the first-stage regression and significant Wald test p-values that we calculated in part C show us that our instrument is indeed strong. From all of this we can say that Naive OLS was a good first approach / attempt at getting some insight but in general IV estimates give us more accurate representation of the causal effects of Medicaid enrollment on the outcomes we are interested on. The IV approach showed us that that Medicaid enrollment may have positive effects on reducing depression and catastrophic expenses and improving the perception that healthcare needs are met, with no clear impact on blood pressure.