**Assignment 1**

**Due: Tuesday, April 7, 5 PM (turn in via Canvas)**

**Overview**

In this assignment, you’ll be using actual survey data from the Oregon Health Insurance Experiment (OHIE). These data are used in Baicker et al. (2013). The assignment will walk you through getting set up with the data, comparing lottery winners and losers on observable characteristics, estimating causal effects, and drawing conclusions. If you need additional assistance, please refer to the Stata Boot Camp as well as Stata resource materials that are available on Canvas. And of course, you are encouraged to bring questions to office hours.

For this and other assignments, you are required to submit a “do file” or equivalent. This is a simple text file with a “.do” suffix, which contains a sequence of commands that can be executed in Stata to replicate your results. You are welcome to use any alternative statistical programming software, such as SAS, SPSS, or R. Please write up your results in a Word document or PDF, including your answers and any relevant output, and append your do file at the end of the document. For each answer, copy-paste relevant output directly into your write-up. See Canvas for a sample write-up.

**Part 1: Getting Acquainted with the Data**

From Baicker et al. (2013): “In 2008, Oregon initiated a limited expansion of its Medicaid program for low-income adults through a lottery drawing of approximately 30,000 names from a waiting list of almost 90,000 persons. Selected adults won the opportunity to apply for Medicaid and to enroll if they met eligibility requirements.” The 30,000 names randomly drawn from the waiting list are called “lottery winners” and those not drawn are called “lottery losers.” The “lottery winners” were invited to enroll in Medicaid if they met eligibility requirements. In this study, survey data were collected among a subset of lottery participants living in Portland, Oregon.

Note that in this study, the “treatment” is the opportunity to enroll in Medicaid. Not everyone who was “treated” ended up enrolling in Medicaid. Note that also that many of the calculations below should be similar but not necessarily identical to those reported in Baicker et al. (2013), since those tables make use of survey weights and additional control variables, which we omit here since they are not essential for pedagogical purposes.

Download the data set, which is called ohie.dta. You can open this directly in Stata. If you are using alternative statistical programming software, you can also download ohie.csv, which is a comma-delimited file with the same data.

1. How many observations are in the data set? What unit of analysis does each observation represent?
2. The variable “treatment” is a 0/1 dummy. It is equal to 1 for “lottery winners” and 0 for “lottery losers.” How many lottery winners are there? How many lottery losers are there?
3. The variable “ohp\_all\_ever\_survey” is a 0/1 dummy. Use the “describe” command to find the variable label for this variable and report it here. The variable is equal to 1 for those who enrolled in Medicaid and 0 for those who did not enroll in Medicaid. How many Medicaid enrollees are there? How many Medicaid non-enrollees are there?
4. What percent of “lottery winners” enrolled in Medicaid? What percent of “lottery losers” enrolled in Medicaid?
5. By how much (how many percentage points) did winning the lottery increase the probability of enrolling in Medicaid? Is this similar to what is reported in Baicker et al. (2013)?

**Part 2: Comparing Lottery Winners and Losers**

As mentioned above, Baicker et al. (2013) focuses on a sample of those living in Portland, Oregon, who responded to an in-person survey. The variable “sample\_inp\_resp” is a 0/1 dummy indicating those who responded to the in-person survey. For the remainder of the assignment, restrict your sample to the survey respondents studied in Baicker et al. (2013). In other words, keep observations only for those individuals who did respond to the in-person survey.

1. How many people are now in your sample? How many lottery winners are now in your sample? How many lottery losers are now in your sample? Is this similar to what is reported in Baicker et al. (2013)?
2. We can compare lottery winners and lottery losers on observable characteristics. Do you expect lottery winners and lottery losers to appear similar on observable characteristics? Why or why not?
3. The variable “gender\_inp” reports gender from the in-person survey. What percent of lottery winners are female? What percent of lottery losers are female?
4. The variable “age\_inp” reports age from the in-person survey, and the variables “age\_19\_34\_inp”, “age\_35\_49\_inp”, and “age\_50\_64\_inp” are 0/1 dummies for different age bins. What percent of lottery winners and lottery losers are in each of the following age bins: 19-34, 35-49, 50-64?
5. The variables “hispanic\_inp”, “race\_white\_inp”, “race\_black\_inp”, and “race\_nwother\_inp” report race/ethnicity indicators from the in-person survey. What percent of lottery winners and lottery losers are in each of the following categories: Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Other, and Hispanic?
6. Overall, would you conclude that lottery winners and lottery losers appear similar or dissimilar on the characteristics you examined? What do you conclude about the study’s internal validity?

**Part 3: Causal Estimates of Winning the Lottery and Enrolling in Medicaid**

1. The variable “bp\_hyper” is an outcome variable indicating those who had elevated blood pressure. What percent of lottery winners and lottery losers have elevated blood pressure?
2. The difference in means between lottery winners and lottery losers on the outcome variable “bp\_hyper” is an “intent-to-treat” (ITT) estimate of the causal effect of winning the lottery on the probability of having elevated blood pressure. Using your results from the previous question, what is your ITT estimate?

For a given outcome, the ITT estimate is the causal effect of winning the lottery. However, for this study, it is relatively straightforward to translate an ITT estimate into an estimate of the causal effect of enrolling in Medicaid. For example, if winning the lottery increased the chance of enrolling in Medicaid by 75%, and we assume that winning the lottery affected outcomes only by changing Medicaid enrollment, then the causal effect of being enrolled in Medicaid is simply (1/.75) as high as the causal effect of winning the lottery. In this example, we can multiply an ITT estimate by (1/.75) to obtain an estimate of the causal effect of enrolling in Medicaid.

1. Using your answer to question 5 and your ITT estimate from the previous question, what is your estimate of the causal effect of enrolling in Medicaid on the probability of having elevated blood pressure? Is this a large effect?
2. Using the same method, what is your estimate of the causal effect of enrolling in Medicaid on each of the following outcomes: total cholesterol is high (chl\_h), screened positive for depression (phqtot\_high), and health about the same or better since last year (health\_change\_noworse)?

**Part 4: Drawing Conclusions**

1. Summarize the results from your estimates of the causal effect of enrolling in Medicaid. On which outcomes do you find that Medicaid has no effect or a relatively small effect? On which outcomes do you find that Medicaid has a relatively large effect?
2. What do you conclude about the overall effectiveness of expanding Medicaid to low-income populations?