# ECON 523: Program Evaluation for International Development
## In-Class Activity 1

This exercise makes use of the data set `E1-CohenEtAl-data.dta`, a subset of the data used in the paper "Price Subsidies, Diagnostic Tests, and Targeting of Malaria Treatment: Evidence from a Randomized Controlled Trial" by Jessica Cohen, Pascaline Dupas, and Simone Schaner, published in the American Economic Review in 2015. The authors examine behavioral responses to various discounts ("subsidies") for malaria treatment, called "artemisinin combination therapy" or "ACT."

Before you begin, save a new do file containing the Stata code below. Extend your do file as you answer the questions.

```
// LOAD DATA

clear all
set scheme s1mono
set more off
set seed 12345

** change working directory as appropriate to where you want to save
cd "C:\Users\pj\Dropbox\ECON-523\topics\1-selection\stata\results"

** load the data from the course website
webuse set https://pjakiela.github.io/ECON523/exercises
webuse E1-CohenEtAl-data.dta

** save the raw data so that you have a local copy
save E1-CohenEtAl-data-my-raw-data-copy, replace
```

1. How many variables are in the data set? (hint: use `describe`, `desc` for short)
2. How many observations are in the data set? (hint: use `describe` or `count`)
3. What does the variable `act_any` measure? (hint: use `describe` or `codebook`)
4. What is the mean of `act_any` to three decimal places? (hint: use `summarize`, `sum` for short)
5. How many people received subsidized malaria treatment? (hint: use `tabulate`, `tab` for short)
6. What does the variable `c_act` measure?
7. What is the standard deviation of the mean `c_act`?
8. What is the standard deviation of the variable `c_act`?
9. What is the standard error of the variable `c_act`? (hint: look at the help file for the `summarize` command)

10. What is the mean level of ACT use among those assigned to the treatment group? (hint: use an `if` statement)

11. Variables starting with `b_` are baseline characteristics (measured before the RCT). Use the summarize command to familiarize yourself with these variables. How many baseline variables are included in the data set? Which ones are missing data for some households in the sample? (hint: `sumb_*`)

12. We're going to look at selection bias by comparing the level of educational attainment among households that choose to use ACT treatment when they have malaria. Use the `cimeans` command to obtain the mean and standard error of `b_h_edu` when `c_act==1` and when `c_act==0`. Using these quantities, calculate the estimated difference in means and its standard error.

13. Now compare your results to what you obtain using the the `ttest` command.

14. Look at the help file for the `ttest` command. Can you figure out why the standard error you calculated does not match the results of the `ttest` command?

15. Confirm that you can replicate your results from Q12 using the `ttest` command.