a. Comment on 2SLS bias in finite sample and its determinants. How does the first stage predictive strength of the instrument affect bias and consistency of the IV estimator?
b. What assumptions are required to estimate the LATE? Prove the LATE theorem under these maintained assumptions.
c. How does the monotonicity assumption help us meaningfully interpret the IV estimator? Discuss the benefit for both the first stage and reduced form estimates.
d. Show that for instruments that generate no always takers, LATE=ATT.
e. What are the benefits of estimating marginal treatment effects? What additional functional form restrictions do researchers *usually* make to estimate the MTE? How do they help? Discuss the steps to obtain the MTE.

This section asks you to replicate results from the Finkelstein et al. (*QJE*, 2012) paper using the public use data files from the Oregon study. These files are freely available through NBER at https://www.nber.org/oregon/4.data.html. You will need an NBER account (which you can easily create). The authors have done a great job of documenting their data and code. Replication code for their QJE paper is also included. In this exercise we will focus on health care utilization (Table 5) as the outcome and only replicate results on survey respondents since that data is public.

a. We have explicit randomization in this setting through the lottery. Present a balance table of individual characteristics between lottery winners and losers and comment on whether you think the randomization appears to be well done. Estimate the intent-to-treat effect of winning the experiment lottery on the likelihood of using different type of health care (extensive margin only). Ensure that you use sampling weights if there are any (follow the paper). Interpret the estimates. Are these estimates policy relevant?

b. Discuss potential identification concerns with using OLS to estimate the effects of Medicaid coverage on healthcare utilization. Present some evidence on the balance between those who received Medicaid (treated) and the remaining people (controls). Discuss your concerns building on this evidence.

c. Regress utilization outcomes from Table 5 on the indicators for Medicaid coverage and obtain OLS estimates. There are several different measures of coverage (see Table 3), use any two that you like. What treatment effect do these coefficients estimate in the context of the potential outcome framework?

d. What would be the arguments for monotonicity and exclusion in this setting? Comment on whether you find them reasonable.

e. Replicate the first stage (Table 3 Columns 5 & 6) for the two measures of Medicaid coverage you used in part (c) above.

f.   How are compliers defined in this setting? If we wanted to characterize the size and composition of the complier group, how would we do it? Use some of the background information available about the people (e.g., gender, age, whether they signed up for the lottery themselves) to characterize compliers. What is the size of the complier group?

g.   Follow the approach used by the authors to instrument for Medicaid coverage using the offer of Medicaid through the lottery. Generate 2SLS estimates for extensive margin health care utilization outcomes in Table 5. Interpret the 2SLS estimate in the context of heterogeneous treatment effects. Can we interpret the LATE as the ATT?