**Meeting attendees.**

Xia Jiang,  Garrett Barber

**Meeting time**

9:00 am – 10:07 am, Dec 20, 2022

**Meeting agenda (an addition meeting in response to an email question).**

1. Discussed some of the questions that Garrett had about the time aspect of before and after treatment, regarding some of the current papers about DID.
2. Explained Jiang’s idea of developing our package (iRCT) based on the original DID (matching estimators) in the research design section below.
3. Reviewed a bit on Garrett’s experiments using the existing package (DID.ipynb).
4. Work assignment.

**Research Design**

iRCT – an intelligent pseudo randomized controlled trial.

1. Implement the simple matching estimator method as described in Jiang’s slide (AboutDID.pptx).
2. Created a simple test dataset using the same example Jiang used in her slides.
3. Test 1) with the dataset created in 2).
4. Include a transform function in our iRCT (See the MBIL package) that can convert all the covariates into one variable (such as the X in the example).
5. Develop a function that convert multi-value variables into a binary variable and include it in the iRCT pacakge.
6. Apply iRCT to our LSM-15year.
7. Identify more interesting “treatment” variables such as Menopausal status in our LSM-15 year, use method developed in 5) to convert them into binary each respectively, if they are non-binary. Then apply iRCT each respectively.
8. Compare what you learned from using iRCT with what you can learn from our MBIL methods, and from the other causal learning methods that we have access to.
9. In terms of the completed causal network, such as the you (Garrett) learned using FCI with our LSM-15year, you can just retrieve the direct causes to the target variable (BCM) and compare with our MBIL and iRCT.

…..

**Thoughts about the transformation function.**

I am concerned that the current implementation (which I just emailed you about) might not work fine in our iRCT case due to a large number of variables we have, which may result in a very large number of possible combinations that can’t be handled by the computer.

Possible solutions:

1. Develop a second version of the transformation function (first fully understand the current java version that I emailed you). The overall idea of this second version is not to physically create all possible combinations, rather just compute the index of a record based on the value of the record.
2. Not to use transformation, rather, compute the distance between two data points of multiple variables. I am sure there are ways to do this out there.

**Progress made in the past week.**

**Issues/Questions and Comments**

Discuss LSM15-Year dataset, maybe in order to solve DID issue.

**Ongoing tasks that cover more than a week**

**Specific tasks for the coming week (the original task assignment for two weeks)**

1. Implement the “research design” section above.
2. Start putting together our iRCT methodology paper.

**Less urgent tasks**