**Meeting attendees.**

Xia Jiang,  Garrett Barber

**Meeting time**

11:00 – 12:00 pm, Jan 9, 2023

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

1. Review the progress of the work assigned last week. Garrett did more tests by flipping the cause and effect using the COVID data. He made an assumption that the results should be same when switch the cause and effect, but based on the new tests, the results are different. This may help resolving the “causal direction” issue.
2. 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.

…..

**Progress made in the past week.**

**Issues/Questions and Comments**

Dr. Jiang’s comments in terms of the causal direction issue.

1. Consider set up a set of rules that we can use to differentiate the cause and effect. For example, based on the new test new results, when the assumption about cause and effect is correct, meaning, when it is consistent with the ground true, the score seems to be significantly higher than the other case. The rule could be as follows:
2. We will conduct tests for every pair of variables.
3. For each pairs, we will test in both directions. The higher score determines the causal direction.

Based on the test results, when the two variables are not correlated, the score is very low. The issue is to find a threshold for no correlation. This is a similar issue to the significance level in statistical testing, threshold of correlation test, or the hyperparameter tunning in machine learning. We may do more testing with more datasets to find such a threshold. In that case, we can establish a third rule as below:

3. If the score is below the threshold value of …, we can consider there are no causal relationship between the two variables.

2) 1) above is heuristic. I wonder whether there is any theoretical work out there to support it? Hope that we can do more literature searching regarding this question.

3) In terms of DID, I wonder whether any other researchers have already notice and discuss this “causal direction” issue.

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

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

1. See the assignment from last week, please complete all.
2. Consider compare with MBIL (Due in two weeks).

**Less urgent tasks**