Drugs Effect on Covid-19

William Nguyen
University of Massachusetts Lowell
william_nguyen2@student.uml.edu

Tiffany Tran
University of Massachusetts Lowell
tiffany_tran2@student.uml.edu

***Abstract*— Covid-19 is one of the most significant pandemics in modern history, according to worldwide data. Even though we have developed several vaccines that work very well against Covid-19, finding out which medicines can help treat Covid-19 can be helpful for future research and development. In this paper, we discovered which medicines are effective to treat Covid-19 by using causal inference algorithms: Backdoor Adjustment and Propensity Score Matching (PSM) on observational data to calculate the Average Treatment Effect (ATE) and Cumulative Average Treatment Effect (CATE).**

Keywords— causal inference, Backdoor Adjustment, Propensity score Matching (PSM), Covid-19, Average Treatment Effect (ATE), Cumulative Average Treatment Effect (CATE)

I. Introduction

The COVID-19 pandemic boosted attempts to identify and develop effective drug treatments. In this research we applied causal inference methods to analyze the impact of medications on COVID-19 patient outcomes. Casual inference approaches take into consideration factors that could affect treatment assignment and results, they enable us to use observational data to simulate the conditions. Our analysis uses two causal inference techniques: propensity score matching and backdoor adjustment.

The backdoor adjustment approach to reduce influencing by conditioning variables that prevent backdoor paths between treatment and outcome. This method makes it possible to estimate Average Treatment Effects (ATE) for people. We implemented this method to calculate the causal effect of 16 different drugs on COVID-19 outcomes while controlling for patient characteristics and conditions.

Propensity score matching (PSM) is an alternate strategy that groups treat and untreated individuals with similar probabilities of receiving treatment. With this approach, we can estimate both Average Treatment Effects (ATE) and Conditional Average Treatment Effects (CATE) for particular group of people, and we can identify the causal effect of drugs on COVID-19 outcomes by finding matched pairs of patients, those who took treatment and those who did not.

II. Dataset

A. About the dataset

The dataset is a subset from National Covid Cohort Collaborative (N3C) with 10,000 instances. It contains patients' information during the COVID-19 pandemic, making it a great resource for researching the possible treatment effects of drugs. The dataset included:

1. **Patient demographics**: include age (represented as continuous variables), gender, race, and ethnicity (encoded), and geographic location (zip code). These demographic variables are crucial characteristics that could affect the process of treatment as well as the results of the illness.
2. **Medical conditions**: are represented in a "conditions" column containing lists of numerical condition code that are SNOMED ID. These codes include important details on patient characteristics that may influence treatment choices and results, such as heart diseases, fever, etc. Each patient will have a list of SNOMED ID in the condition column.
3. **Treatment variables**: consist of 16 kinds of different drugs (trazodone, amitriptyline, fluoxetine, citalopram, paroxetine, venlafaxine, vilazodone, vortioxetine, sertraline, bupropion, mirtazapine, desvenlafaxine, doxepin, duloxetine, escitalopram, nortriptyline)
4. **Outcome**: provides a binary value for patient survival or recovery (1) or non-survival or death (0), in addition to an additional variable that indicates the severity of COVID related deaths. The evaluation of treatment efficiency in improving patient survival is made possible by these outcome indicators.

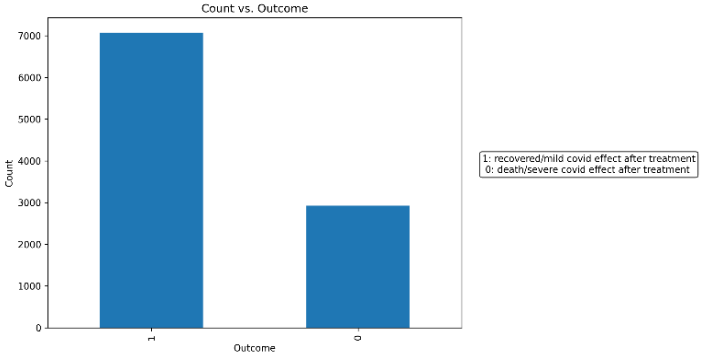

Fig. 1. The graph shows the number of people who recovered (outcome 1) and did not recover (outcome 0)

B. Data Preprocessing

In order to prepare the dataset for causal inference analysis, we converted the dataset into a suitable format. We preprocess data based on the following steps:

1. Encoding the condition: Since each patient have a list of SNOMED-ID in the condition columns, it could be thousands of different SNOMED-ID in the dataset and

doing one hot encoding to separate them may not be effective and will be computationally expensive later. To solve this issue, we used Word2Vec to convert each SNOMED-ID into a vector embedding of size 5. This gives us a list of vector embeddings for each patient. Then, we compute the pooled embedding that summarizes the conditions that a patient has by calculating the average of vector embeddings in the list of condition of that patient. This effectively gives us a single pooled embedding for each patient that represents all the medical conditions that they have.

2. We also removed 2 redundant columns named severity_covid_death and zip.
3. After obtaining the pooled embedding for each patient, since each pooled embedding vector has 5 values, we split them into 5 different columns named condition1, condition2, condition3, condition4, condition5 so that we can implement the algorithms for causal inference easier.
4. To further preprocess the dataset for Backdoor Adjustment, we remove the 5 condition columns. Also, we need to convert the continuous variable to categorical because Backdoor Adjustment does not work with continuous variable. For the 5 condition columns and age, we turn the upper half to 1 and lower half to 0. For the columns with more than 2 categories, we convert half of each category to 0 and the other half to 1.
5. To further preprocess the dataset for PSM, we do not do step 4 but create a copy of the dataset and do other preprocessing. We first numerically encode the categorical variables. Then, because our dataset has a mix of continuous and categorical features with different distribution and of value which is not good for training models like logistic regression, we decided to min-max scaling the variables (except for the 5 conditions columns since they are already in a small range of value)

III. Algorihtm Details

For this project, we will use 2 algorithms: Backdoor Adjustment and Propensity Score Matching

A. *Backdoor Adjustment*

Backdoor adjustment allows us to estimate the causal effect of a treatment by controlling the confounders. In observational data, the confounders like ages, gender, etc., can both influence the treatments that a patient would take and the outcome (whether the patient recovers or not). Controlling the confounders allows us to account for their effect on the treatments and outcomes and leads to more accurate results by removing the non-causal association between the treatments and the outcomes.

In order to compute ATE for a treatment using Backdoor Adjustment, we calculate the causal effect considering each confounder at a time.

$$\sum_i \frac{N(T,X=i)}{N(X=i)} * P(Y=1|T,X=i)$$

Fig. 2. The formula to calculate the causal effect of each confounder.

In the formula in Fig. 2. We iterate over each value (category) of the current confounder. N (T, X=i) is the number of patients who have taken treatment T and have the value i for this confounder. N(X=i) is the number of patients who have value i for the current confounder. P (Y = 1 | T, X= i) is probability that a patient survives given they take treatment T and have value i for this confounder. This essentially gives us the treatment effect of the treatment considering one confounder.

Since we have multiple confounders, we need to do the same thing for each of them and get multiple treatment effects. Then, we find the average of these treatments to get the average treatment effect (ATE) of a treatment.

Backdoor Adjustment is the most straightforward algorithm and easy to implement. However, this algorithm only works for categorical values. If the variables are continuous, then, we need to do use a discretization technique such as using logistic regression to turn the continuous values into categorical values.

B. *Propensity Score Matching (PSM) for ATE*

PSM is also a technique to estimate the causal effects of treatments in observational data by accounting for confounders. But unlike Backdoor Adjustment, PSM simulates randomized Controlled Trial (RCT) by matching each patient in the treatment group with another patient in the control group. Calculating. Calculating ATE using PSM consist of multiple steps:

1. Calculating the propensity score for the patients in the treatment group and control group. The propensity score is the probability that a patient takes the treatment given the covariate. Propensity scores are used to group or match patients with similar covariates, even if they didn't receive the same treatment. This can be computed by training a logistic regression model using the data of people who took the treatment (treatment group) and the control group and whether that patient took the treatment as labels.
2. Matching each patient who took the treatment with another patient who did not take the treatment but have a similar propensity score (closest to the propensity score of the patient who took the treatment). This is done using KNN to find 1 nearest neighbor. Matching allows us to get the counterfactual, which is the outcome of the patient who takes the treatment if we did not give them the treatment. The outcome of the matched patient in the control group will be used as the counterfactual.
3. Calculating ATE. The ATE can be calculated by finding the average of the outcome of people who took the treatment (the sum for the average is simply just adding all the 0 and 1 even though the outcome is a categorical feature) and subtract by the average of the outcome of people who did not take the treatment. The result of this subtraction is the ATE. The ATE will also be calculated

by using logistic regression. First, we train a logistic regression using the treatment and matched control group as training data with whether they recover (outcome 1) or not as labels. Then, we use the weight for the treatment variables in the logistic regression as ATE. If the ATE is positive, then the treatment is helpful but if the result is negative, then the treatment can be harmful.

C. *Propensity Score Matching (PSM) for CATE*

PSM is seem to be a more efficient technique than Backdoor adjustment. However, it has a major weakness when calculating the propensity scores using logistic regression. Due to the imbalance in the dataset, there are much more people who did not take the treatment compare to number of people who took the treatment. This imbalance make it really hard for the logistic model to learn to predict correctly for the minority class. To address this issue, we use boostrapping. We repeatedly choose a subset of datapoints from the control group randomly so that it would have the same number of datapoints as in the treatment group, then, we compute the propensity score, do matching and calculate ATE like before. Repeat this entire process 500 times, and we will obtain a list of ATE for the treatment. Next, we calculate the confident interval on 95%. With the mean as our best estimate for the CATE (cumulative average treatment effect). We will also obtain the lower bound and upper bound for the 95% Confident Interval. These information essentially says "we are 95% condident that the true treatment effect is within the lower bound and upper bound range with the best guess of what the treatment effect is the mean."

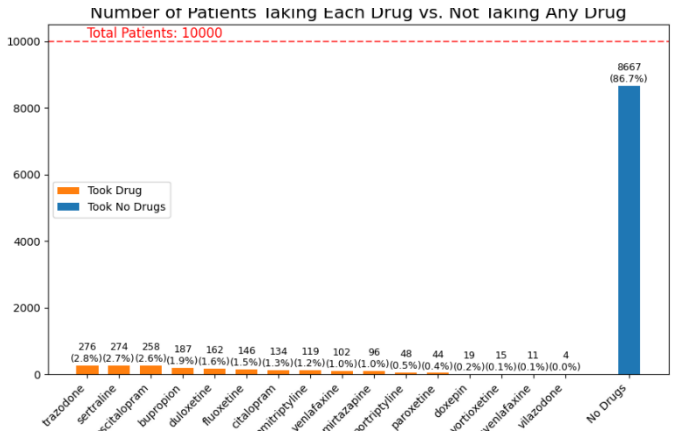

Fig. 3. Graphs that show the number of patients take each kind of drug and not taking any drugs

IV. Training/Implementing The Algorihtms

The algorithm described in section 3 is implemented using Python with the help of libraries:

1. Pandas: for manipulating the csv files
2. Numpy: for doing array operations and manipulations and working with the data in the dataframe from Pandas.
3. Scikit-learn: for training logistic regression models, doing KNN, and do minmax scaling.
4. Matplotlib: for creating graphs
5. Gensim: for implementing Word2Vec
6. Scipy: for calculating the confident interval

One major addition is that the algorithms details mentioned in section 2 is repeated (just use a for loop) for each of the 16 drugs (treatments).

Almost all the implementation for causal inference algorithm is coded from scratch except the Confident Interval is adapted from [1] in the references.

V. Results

A. *ATE from Backdoor Adjustment*

Table 1: Summary of ATE calculated using Backdoor Adjustment

Treatments	ATE
trazodone	0.579
amitriptyline	0.678
fluoxetine	0.737
citalopram	0.656
paroxetine	0.508
venlafaxine	0.659
vilazodone	0.180
vortioxetine	0.665
sertraline	0.645
bupropion	0.646
mirtazapine	0.359
desvenlafaxine	0.818
doxepin	0.679
duloxetine	0.552
escitalopram	0.678
nortriptyline	0.642

After calculating ATE using Backdoor Adjustment, desvenlafaxine is the one that has the highest ATE with a value of 0.818 and vilazodone is the one that has the lowest ATE with a value of 0.18. Most of the other treatments have ATE between 0.5 and 0.66. Looking at this result, we can say that desvenlafaxine is the most effective treatment for the patients and vilazodone is the least effective treatment for Covid-19.

B. *ATE from Propensity Score Matching (PSM)*

Table 2: Summary of ATE calculated using PSM

Treatments	ATE
trazodone	0.037
amitriptyline	0.019
fluoxetine	0.057
citalopram	0.014
paroxetine	0.003
venlafaxine	0.021

vilazodone	0.208
vortioxetine	-0.159
sertraline	-0.015
bupropion	0.013
mirtazapine	0.072
desvenlafaxine	-0.020
doxepin	0.023
duloxetine	0.084
escitalopram	0.057
nortriptyline	-0.001

After calculating ATE using PSM, duloxetine is the one that has the highest ATE with the value of 0.084 and nortriptyline is the one that has the lowest ATE with the value of -0.001. Nortriptyline, vortioxetine, sertraline, and desvenlafaxine all has negative ATE indicating a harmful effect on the patient. This is not aligned with the result from the Backdoor Adjustment which showed that desvenlafaxine is the most effective treatment.

Table 3: Summary of ATE calculated using PSM with logistic regression weight

Treatments	ATE
trazodone	-0.084
amitriptyline	0.016
fluoxetine	0.090
citalopram	-0.033
paroxetine	-0.188
venlafaxine	-0.022
vilazodone	-0.863
vortioxetine	-0.077
sertraline	-0.065
bupropion	-0.020
mirtazapine	-0.288
desvenlafaxine	0.148
doxepin	0.070
duloxetine	-0.056
escitalopram	0.037
nortriptyline	-0.009

After calculating ATE using PSM with logistic regression weight (using the weight of the treatment in the logistic regression model as the ATE), desvenlafaxine is the one what has the highest ATE with the value of 0.148. Nortriptyline, duloxetine, mirtazapine, bupropion, sertraline, vortioxetine, vilazodone, venlafaxine, paroxetine, citalopram, and trazodone have negative ATEs indicate that can has a harmful effect on the patient when using it to treatment COVID-19. This result aligns with the result from Backdoor Adjustment where desvenlafaxine is shown to be the most effective treatment and contradict with the other result of PSM with indicates that desvenlafaxine has a harmful effect.

C. CATE from Propensity Score Matching (PSM)

Table 4: Summary of ATE calculated using PSM

Treatments	Mean	Lower bound	Upper bound
trazodone	-0.057	-0.060	-0.054
amitriptyline	-0.022	-0.026	-0.018
fluoxetine	-0.023	-0.027	-0.020
citalopram	-0.015	-0.020	-0.011
paroxetine	0.024	0.017	0.031
venlafaxine	-0.077	-0.081	-0.072
vilazodone	-0.292	-0.292	-0.292
vortioxetine	-0.030	-0.043	-0.018
sertraline	-0.055	-0.058	-0.053
bupropion	-0.029	-0.032	-0.053
mirtazapine	-0.064	-0.070	-0.060
desvenlafaxine	-0.105	-0.120	-0.090
doxepin	0.032	0.019	0.047
duloxetine	-0.028	-0.032	-0.024
escitalopram	-0.052	-0.054	-0.049
nortriptyline	-0.003	-0.009	0.004

After calculating CATE using PSM, we can see that doxepin have the highest CATE with the value of 0.032, and vilazodone has the lowest CATE with a value of -0.292. Also, we can see that only paroxetine and doxepin has a positive effect while all other has a negative effect. This is very different from the other results.

Table 5: Summary of ATE calculated using PSM with logistic regression weight

Treatments	Mean	Lower bound	Upper bound
trazodone	-0.056	-0.058	-0.053
amitriptyline	0.054	0.050	0.058
fluoxetine	0.039	0.035	0.042
citalopram	0.026	0.022	0.030
paroxetine	0.007	-0.0002	0.014
venlafaxine	0.004	-0.001	0.008
vilazodone	-0.918	-0.942	-0.892
vortioxetine	0.122	0.095	0.148
sertraline	-0.027	-0.030	-0.024
bupropion	-0.021	-0.025	-0.018
mirtazapine	-0.251	-0.258	-0.245
desvenlafaxine	-0.744	-0.852	-0.637
doxepin	0.062	0.046	0.078
duloxetine	-0.036	-0.040	-0.031
escitalopram	-0.033	-0.036	-0.030
nortriptyline	0.102	0.094	0.108

After calculating CATE using PSM, we can see that vortioxetine has the highest CATE with the value of 0.12, and vilazodone has the lowest CATE with a value of -0.918. Both

methods show that vilazodone has the most harmful effect on the patient when using it to treat COVID-19.

All the data above show quite inconsistent results, and it is hard to say which treatment is truly effective when looking at them. However, we can somewhat say that desvenlafaxine, duloxetine, doxepin, and vortioxetine has the potential of being effective when using to treat COVID-19.

VI. Limitations and Future Works

The small amount of data may have affected the accuracy of the results. Having more data would be about to give a higher accuracy. Also, Backdoor Adjustment was used on a binary encoded dataset. This can also affect the results since it simplifies many things in distribution and variety. Additionally, using Word2Vec to create pooled embeddings to summarize the patient condition may have affected the results.

In the future, we would like to try a more advanced machine learning method for this task like using neural networks as they are the most chosen model for many tasks nowadays. We would also like to expand the dataset to have more instances and variables to increase the accuracy of the results.

VII. Conclusion

In this paper, we have implemented the causal inference algorithms: Backdoor Adjustment and Propensity Score Matching (PSM) to calculate the ATE and CATE of the 16 different treatments (drugs) on treating COVID-19. Even though the results showed are inconsistent most likely due to the dataset size and data preprocessing steps, we can at least say that desvenlafaxine, duloxetine, doxepin, and vortioxetine has the potential to be effective when treating COVID-19.

References

[1] Rahman1, Md Mahmudur, et al. “Effects of Antidepressants on COVID-19 Outcomes: Retrospective Study on Large-Scale Electronic Health Record Data.” *Interactive Journal of Medical Research*, JMIR Publications Inc., Toronto, Canada, www.i-jmr.org/2023/1/e39455/. Accessed 8 May 2025.

[2] “Word2Vec Model.” *Gensim*, 10 Aug. 2024, radimrehurek.com/gensim/auto_examples/tutorials/run_word2vec.html.

[3] Szubelak, Lukasz. “Causal Inference with Python: A Guide to Propensity Score Matching.” *Towards Data Science*, 5 Mar. 2025, towardsdatascience.com/causal-inference-with-python-a-guide-to-propensity-score-matching-b3470080c84f/.