**DATA SCIENCE 6500**

**Case Study 4:**

**Comparison Targeted maximum likelihood, G-computation and Inverse Probability Weighting for Examining correlates of cannabis users’ engagement with a digital intervention for substance use disorder: An observational study of clients in UK services delivering Breaking Free**

**Professor:**

**Dr. Alysha Cooper**

**Student:**

**Mohsen Selseleh**

**Student no:**

**1207477**

**Summer 2022**

**Background**

Cannabis is one of the most popular drugs, and research suggests that cannabis use disorder (CUD) may be more common than previously believed. To satisfy treatment needs, research should examine novel behavioral support strategies, such as computer-assisted therapies like Breaking Free Online (BFO). Elison-Davies et al (2021) sought to understand the relationships between baseline sociodemographic and clinical characteristics of participants and their engagement with BFO, as well as the potential relationships between baseline characteristics and engagement and cannabis use and biopsychosocial functioning at follow-up(Elison-Davies et al ,2021).Elison-Davies et al (2021) have Examined correlates of cannabis users’ engagement with a digital intervention for substance use disorder: An observational study of clients in UK services delivering Breaking Free Online.They used 1830 people in an observational study. Data analysis was done using an ANOVA to assess gender differences in BFO involvement, cannabis usage severity, and quality of life. Regression models are used to examine the proportion of behavioral change strategies that are successfully implemented given the confounding variables. After accounting for confounding, they have conducted further regression models to examine the impact of BFO on outcomes (quality of life, cannabis usage, etc.) (Elison-Davies et al ,2021).

A brief overview of the observational study: They characterized the treatment as how frequently the 12 behavior modification strategies gauge participants' level of program participation (continuous treatment). Confounding factors in their study included sociodemographic information (such as age and gender), clinical traits, and the frequency of other medication use. They used ANOVA to examine how each confounding variable affected the outcomes differently. They identified potential unmeasured confounding factors such as education level, family dynamics, income, participation in other programs, and social networks (who they are hanging out with). Because the causal inference may account for various confounding factors, it may aid in obtaining a more accurate response to the study question. (Elison-Davies et al ,2021). They define the potential outcomes as:Cannabis use (1), cannabis use(2),…,cannabis(12)

Mental health (1), mental health (2),…, mental health(12).

Their study did not use causal inference methods and therefore could only make claims about the associations between the variables.The emphasis on explicitly defining the causal question of interest and estimating well-defined statistical parameters motivated by the scientific question is the foundation of the causal inference framework's strength. The causal inference paradigm also emphasizes careful confounding correction, maybe by using adaptable algorithms that make few functional form assumptions (Schuler&Roes,2017). Certain presumptions call for the use of causal inference: Although several behavioral change approaches may be used inside the application (i.e., different versions of therapy), it may be assumed that the outcome will be the same regardless of which behavioral change technique is used, so the consistency is not fully met in this case. Although there is almost no unmeasured confounding, some factors that may affect both BFO involvement and cannabis usage or mental health were not taken into consideration. Because each person theoretically might have taken part in any level of treatment, the research is positive.

Therefore, I will assume that necessary causal assumptions are met. I employ the causal inference approach to explore the causal relationship between a digital intervention for substance use disorder and cannabis users' participation in the UK to overcome the shortcomings of prior studies.

**A statement of the specific project objective**

I will try to estimate causal effects using observational data with TMLE, G-computation and inverse probability weighting with using a super learner (ensemble algorithm).

**Analysis Plan (Overview of proposed methods for statistical analysis)**

In 12 behavioral change strategies, the statistical analysis will try to determine the causal impact of a digital intervention for substance use disorder. The results of the mediation analysis will help us determine whether the 12 behavioral change strategies have an impact on a person's social network and whether social networks after treatment have an impact on drug use disorders.

I will use the Targeted Maximum Likelihood Estimation method (TMLE) to measure the causal effect of interest. I will examine the effect of optimal participation in 12 behavioral change techniques at 1-and 3-month follow-up times on substance use disorder at 12-month follow-up.

The TMLE, G-computation, and Inverse Probability Weighting were estimated using machine learning algorithms (super learner). The confounders included in the TMLE estimation were any measured variable that was not balanced between those who participated optimally in 12 behavioral change techniques and those who did not. I have calculated ATE, average treatment effect; CI, confidence interval, mean bias for TMLE, G-computation, and Inverse Probability Weighting. The results of all three methods show a positive effect of 12 behavioral change techniques on substance use disorder at 12 months, with those who participated optimally being more likely to abstain from substances at follow-up. Fundamentally, machine learning methods can be used to create TMLE and the other estimators presented, which may be useful for complex observational data. As shown by our simulation findings, the super learning algorithm (ensemble algorithm) outperformed parametric regression for all 3 estimators. In comparison to main-term parametric regression, machine learning algorithms, in particular ensemble methods like super learning, can empirically identify interaction, nonlinear, and higher-order relationships among variables; as a result, the corresponding ATE estimate is less likely to be biased due to an incorrect functional form. Additionally, even though parametric regression and TMLE with super learning performed similarly in our simulation study, TMLE with super learning may surpass G-computation and Inverse Probability Weighting with super learning (Schuler & Rose,2017).

**Ethical Considerations**

Even though it is unlikely that the suggested analysis would have any unfavorable effects, it is crucial to think about the research's ethical consequences. Most significantly, patient privacy will be protected by anonymizing the data needed for studies. All the participants whose data will be used in the suggested analysis gave their consent to have their information used for the study. Participants shouldn't be in danger from the suggested analysis because the data has already been gathered. Furthermore, considering that most patients were white, male, and heterosexual, the data might have been drawn from an unrepresentative sample. As a result, it is possible that the results of the suggested analysis would not apply to minority communities, and more research will be required (Cooper,2022).

**Significance of Work and Conclusion**

Past research has demonstrated positive associations between users’ engagement with a digital intervention with substance use disorder (Schuler&Rose,2017). While TMLE has not been used as frequently in epidemiologic research, we strive to convey it in an understandable manner. I showed that TMLE outperforms G-computation, and Inverse Probability Weighting with super learning(ensemble algorithm) to demonstrate a causal effect of 12 behavioral change techniques on substance disorder. Meanwhile, my goal was comparison these methods. TMLE is another cutting-edge tool that may be added to the statistical toolbox of the applied practitioner because it is a flexible estimate approach that can easily combine nonparametric machine learning techniques.

**Reference**

Cooper, Alysha (2022). Natural (In)direct effects of 12-steps treatment on abstinence

from substance use: The role of social networks, DATA\*6600.University of

Guelph.

Elison-Davies, S., Wardell, J. D., Quilty, L. C., Ward, J., & Davies, G. (2021).

Examining correlates of cannabis users' engagement with a digital

intervention for substance use disorder: An observational study of clients in

UK services delivering Breaking Free Online. *Journal of Substance Abuse*

*Treatment, 123*, 108261.

Schuler, M. S., & Rose, S. (2017). Targeted maximum likelihood estimation for

causal inference in observational studies. *American journal of epidemiology,*

*185(1)*, 65-73.