Causal Learning with Large-Scale Propensity Scores to Predict Treatment Outcomes: A Study of Arrhythmia in Adolescents with Attention-deficit/hyperactivity disorder

Junhyuk Chang[1], Dong Yun Lee[2], Rae Woong Park[1,2]
[1]Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea
[2]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea

Background

When adolescents with attention-deficit/hyperactivity disorder (ADHD) are also diagnosed with depression, clinical guidelines recommend the concurrent use of ADHD agents (e.g., methylphenidate; MPH) and antidepressants, such as selective serotonin reuptake inhibitors (SSRIs). While MPH is effective for treating ADHD and SSRIs are effective for treating depression, their concurrent use should be prescribed with caution due to the potential increase in cardiovascular risk, including arrhythmia.[1,2] Although the general safety outcomes of the concurrent use of MPH and SSRIs are known, there is a lack of research analyzing the effects on individual patients.

The causal machine learning method is able to estimate treatment effects on individual patients by calculating average treatment effects and non-parametrically adjusting for potential confounding predictors inherent in observational data.[3,4] These techniques predict outcomes based on complex variable interactions and identify heterogeneity in treatment effects, enhancing the robustness and reliability of our findings. We aim to analyze the treatment effect of concomitant administrating SSRIs and MPH on arrhythmia occurrence in MPH patients with a causal forest model using common data model (CDM) and large-scale propensity score method which is CDM methodology.

Methods

In this study, we used the Health Insurance Review and Assessment Service - Attention Deficit/Hyperactivity Disorder (HIRA-ADHD) database, which was collected from January 1, 2016, to December 31, 2020, and contained ADHD patient data from nationwide claims data. The HIRA-ADHD database was converted to Observational Medical Outcomes Partnership-CDM. We defined MPH-used patients with an ADHD diagnosis based on the following criteria: 1) aged between 10 and 19; 2) patients with a depression record; 3) patients without other anti-ADHD agents and previous antidepressants.

The target outcome in this study is the occurrence of arrhythmia, we also extracted the SRI prescription record of the cohort population. For developing and validating the causal forest model, we divided the dataset into 70% for training and 30% for validation, ensuring the same outcome prevalence in both sets.

We extracted patient baseline covariates to employ a large-scale propensity score utilizing the FeatureExtraction, OHDSI open-source package. Initial screening was conducted to exclude rare covariates by 10-fold cross-validation in the training sample using a logistic regression model generated from the binomial outcome.

A large-scale propensity score approach was used to adjust for differences in baseline covariates between SSRI-treated and non-treated patients. Propensity scores (PS) were generated using random forests (RF) with screened covariates, and each patient was weighted using inverse-propensity weighting. With calculated PS and weights, the constructed causal forest model estimated the average treatment effect (ATE) via a doubly robust method. This method adjusts for baseline covariate differences by combining weights based on SSRI treatment probability and outcome regression using a random forest.[3,5]

We estimated treatment heterogeneity using the conditional ATE (CATE). Initially, CATEs were estimated in the 30% test sample by adopting a causal forest model trained from the training dataset. The validation dataset was divided into quintiles based on calculated CATEs, and ATEs were estimated within each quintile. This approach aimed to determine if the ATE was highest for those predicted to benefit most and lowest for those predicted to benefit least. To ensure robustness, a rank-average treatment effect (RATE) was used to estimate heterogeneity within quintile groups.

To identify characteristics of high and low CATE groups, we compared additional characteristics by analyzing the distribution of the top 5 variables based on variable importance from the causal forest model. Using the mean CATE of the validation dataset, we divided the groups into high and low CATE groups and analyzed whether there were statistically significant differences in the distribution of the top 5 variables between these groups.

Results

Among the total of 11,163 MPH-used patients, 7,873 patients were prescribed SSRIs and 58 patients had occurrences of arrhythmia. After dividing the outcome prevalence equally into training and validation datasets, the training dataset included a total of 7,813 patients, of whom 5,511 were treated with SSRIs, and the validation dataset included a total of 3,350 patients, of whom 2,362 were treated with SSRIs. From these patients, we initially extracted a total of 2,544 baseline covariates, which were then reduced to 1,766 covariates after 10-fold cross-validation.

The ATE for the validation dataset is 0.12. Based on the CATE calculated in the validation dataset, the cut-off values for the quantile groups Q1 to Q5 were as follows: -0.20, -0.06, -0.02, -0.01, 0.00, and 0.67. The validation dataset was divided equally into 5 quantile groups, which contained 670 patients respectively.

Figure 1 represents the ATE of the quantile groups. In order of increasing quantile, the calculated ATEs were -0.5, -0.1, 0.1, 0.1, and 0.4. Among ATE of quantile groups, the ATE of the Q5 group is statistically significant (95% CI: 0.1-0.8). The results in the Q5 group suggest that SSRIs may significantly affect

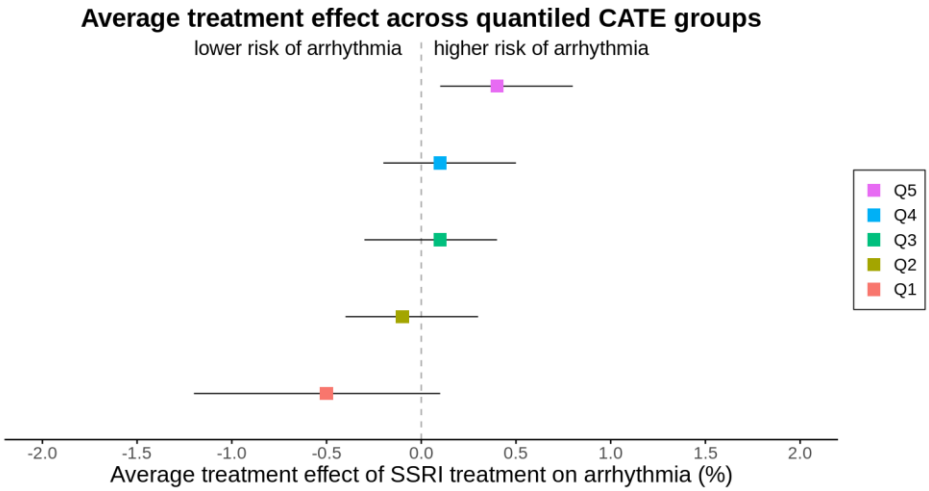

Figure 1. Average treatment effect of quantile groups

arrhythmia in this subgroup. The estimated RATE was 0.008 (95% CI: 0.002-0.015), which confirmed the heterogeneity between quantile groups.

Figure 2 represents the density of top 5 baseline covariates between high and low CATE groups. We extracted top 5 covariates from the variance importance of the constructed causal forest model: ' age group: 10-14', 'Individual psychotherapy',' index month: 9',' Mixed anxiety and depressive disorder',' age group: 15-19'. Differences in distribution for all covariates were statistically significant.

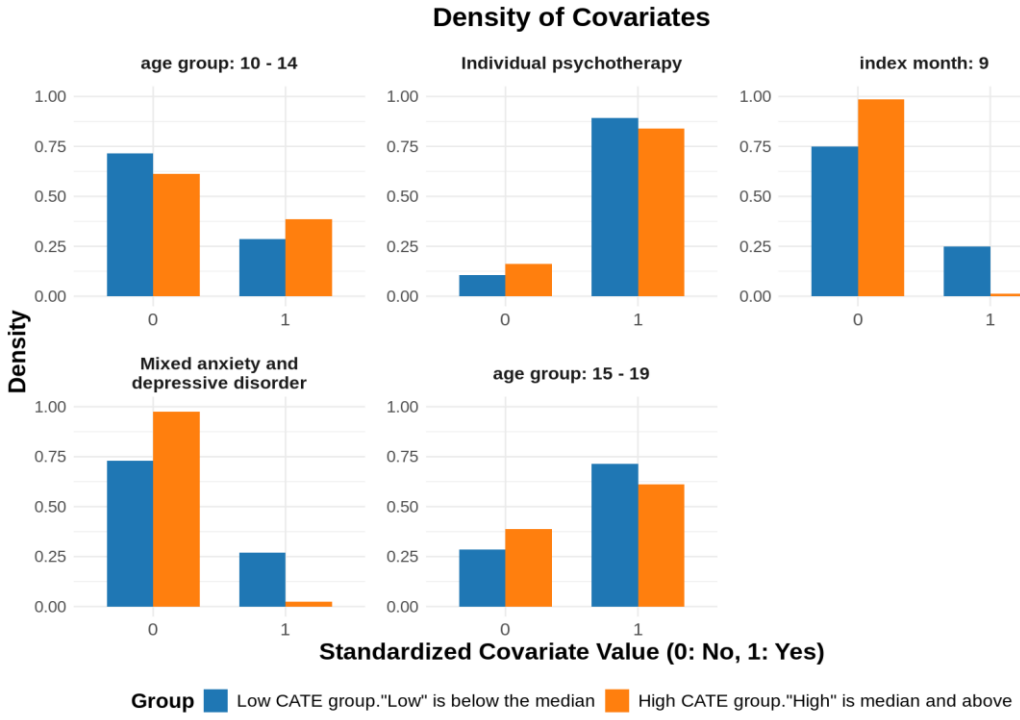

Figure 2. Density of top 5 covariates

Conclusion

The findings of this study suggest that while SSRI treatment did not significantly affect arrhythmia, certain populations can be affected higher occurrence of arrhythmia. Also, substantial heterogeneity exists across patients based on quantile CATE and covariates. An individualized treatment rule accounting for this heterogeneity could modify guidelines for concurrent use of MPH and SSRIs.

Acknowledgment

This research was funded a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR16C0001) and this research was supported by a Government-wide R&D Fund project for infectious disease research (GFID), Republic of Korea (grant number: HG22C0024, KH124685).

References

1. Kim C, Lee DY, Park J, et al. Safety outcomes of selective serotonin reuptake inhibitors in adolescent attention-deficit/hyperactivity disorder with comorbid depression: the ASSURE study. *Psychological Medicine*. 2023;53(10):4811-4819. doi:10.1017/S0033291723000120.
2. Zhao Y, Zhang Y, Yang L, Zhang K, Li S. Safety Profile of Selective Serotonin Reuptake Inhibitors in Real-

World Settings: A Pharmacovigilance Study Based on FDA Adverse Event Reporting System. Annals of Pharmacotherapy. 2024;0(0). doi:10.1177/10600280241231116
3. Ross EL, Bossarte RM, Dobscha SK, et al. Estimated Average Treatment Effect of Psychiatric Hospitalization in Patients With Suicidal Behaviors: A Precision Treatment Analysis. JAMA Psychiatry. 2024;81(2):135–143. doi:10.1001/jamapsychiatry.2023.3994
4. Feuerriegel, S., Frauen, D., Melnychuk, V. et al. Causal machine learning for predicting treatment outcomes. Nat Med 30, 958–968 (2024). https://doi.org/10.1038/s41591-024-02902-1
5. Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. Am J Epidemiol. 2011 Apr 1;173(7):761-7. doi: 10.1093/aje/kwq439. Epub 2011 Mar 8. PMID: 21385832; PMCID: PMC3070495.