Does SSRI use reduce COVID-19 mortality? A Study of EHR data in N3C Enclave using Targeted Learning for Causal Inference

Sonja S. Nakasian, MSc[1], Dan Housman[1], Yuval Koren, MSc[1], Daniel Poscover, MBA[1], Sergey Krikov, MS[4], Charles E. Barr, MD, MPH[1], Andy Wilson, PhD, MStat[2,4], Mark van der Laan, PhD[3], Ivana Malenica, PhD[5], on behalf of the N3C Consortium

Graticule, Inc., Newton, MA; [2]Parexel LLC, Salt Lake City, UT; [3]University of California, Berkeley; [4]University of Utah; [5]Harvard University

Background

- Published studies of selective serotonin reuptake inhibitors (SSRIs), specifically fluvoxamine, fluoxetine, and sertraline, showed mixed results for efficacy in COVID-19 despite plausible biological mechanisms[1]
- Double-blind, placebo-controlled trials (NCT 04727424; 04342663) and a prospective, open-label real-world evidence study indicated improvement in COVID-19 prognosis among fluvoxamine users; however, notable limitations were present such as the lack of clinically meaningful endpoints, small sample size, or failure to adequately address causal effects and confounding based on conditional exchangeability.[3]
- Targeted maximum likelihood estimation (TMLE), a family of advanced methods for causal inference, has been shown to handle complex observational real-world data very well compared to traditional parametric approaches [2]
- In an effort to better understand or clarify disparate findings from previous research, this study will utilize TMLE to compare risk of mortality in chronic vs. naive SSRI patients among a large COVID-19 population, leveraging EHR data in the National COVID Cohort Collaborative (N3C) Enclave (covid.cd2h.org/enclave)

Objectives

- To implement advanced causal machine learning methods via the Targeted Learning roadmap and evaluate the feasibility of using TMLE in the N3C Enclave
- To evaluate the potential protective effect of SSRI use in a large national cohort of COVID-19 patients by comparing risk of mortality in chronic SSRI (sertraline, fluvoxamine, fluoxetine) vs. naive patients, and determine whether there is an observable class effect among chronic SSRI users
- To develop insights and better understanding of disparate findings from previous research showing mixed efficacy of SSRI use in COVID-19 patients[1]

Study Design

- Retrospective, population-based case-control study using electronic health records data from the N3C database to identify a population of patients with COVID-19 based on the first record (index date) of a positive PCR or antigen test, or ICD-10 diagnosis codes for COVID-19
- Patients in the COVID-positive population were classified based on baseline use of SSRIs into two exposure groups: (1) chronic SSRI-users exposed at least 4 times to ≥1 SSRIs (fluvoxamine, fluoxetine, sertraline) within 1 year prior to the index date; (2) SSRI-naïve patients not exposed to any of the SSRIs at any time prior to the index date
- Baseline variables included age, gender, indications for SSRI use (on and off-label), comorbidities associated with adverse COVID-19 outcomes (e.g., cerebrovascular, cardiovascular, autoimmune complications), and baseline status proximal to index (Table 1)
- Patients in the two exposure groups were followed longitudinally after the index date of COVID-19 diagnosis. The primary outcome measure was a binary outcome of 180-day all-cause mortality

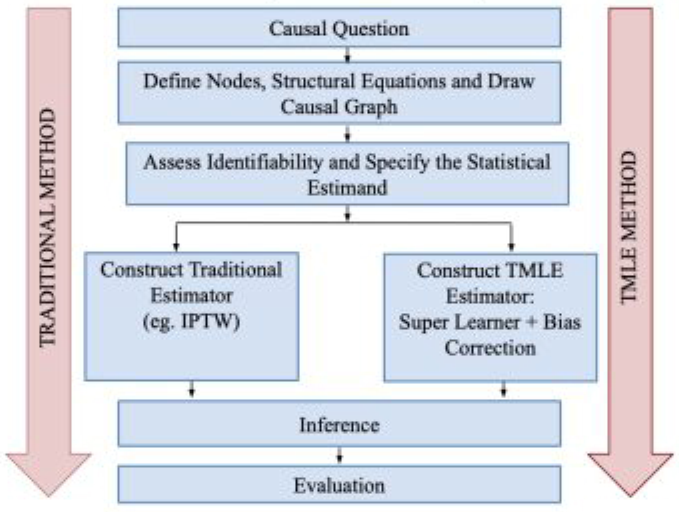

Figure 1. Statistical process overview

The image represents statistical process comparing the Traditional IPTW and TMLE method

Data Source

The N3C De-identified Dataset harmonizes EHR data from multiple data partners (77 sites) into the OMOP common data model, and is one of the largest and most secure resources available for COVID-19 research with about 6.3 million COVID-19 patients

Methods

- Directed Acyclic Graphs (DAGs) captured causal assumptions about 36 variables related to exposure and outcome, of which 28 were defined in N3C concept sets. Broader sets of confounders (including the complex associations between them) were considered according to the disjunctive cause criterion,[5] which simply requires potential confounders to be causes of the outcome, the exposure, or both
- The complete list of 28 candidate variables includes demographics, baseline health status, medication history, and intercurrent events
- Crude and conditional odds ratios (ORs) were calculated using traditional (multivariable logistic regression) methods
- To better address potential confounding, propensity scores were computed using ensemble super learners to predict treatment from potential confounders and used to inversely-weight observations (IPTW)
- The analysis was performed using open-source R software v 4.2. Each of the given SSRI treated patients were matched to an untreated subject using Nearest Neighbor propensity score matching (package 'MatchIt')
- Packages *'WeightIt'* and *'cobalt'* were used for propensity score weighting and package *'Survey'* was used for the estimation of treatment effect on the outcome
- The intervention effects (as average treatment effects (ATE)) were also estimated using Double-Robust TMLE, as this technique combats bias by employing data-adaptable algorithms to control for confounding factors
- The variables in the study have a complicated structure (they are non-normal, non-linear, and exhibit strong correlations), making it necessary to utilize ensemble super learners through the R TMLE package ('tmle').[2] This approach has proven to be more effective than traditional methods such as logistic regression.

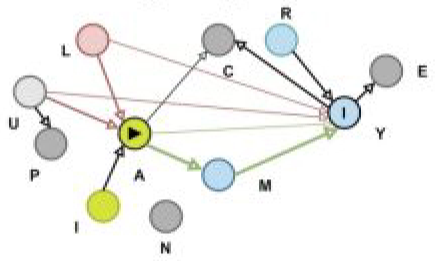

Figure 2. Directed Acyclic Graph to display causal relationships between variables[2]

Variable roles: A = exposure or treatment; Y = outcome; L= confounder, R= risk factor for Y, M= mediator; C = collider; E = effect of Y; I = instrument;u = unmeasured confounder, P = proxy of U: N = noise variable

The DAG provides a simplistic and transparent way to evaluate causal relationships between variables. In this study, a complex DAG was constructed to evaluate variables *L, R, E, M, C, U, P* that relate to propensity of fluvoxamine, fluoxetine or sertraline treatment (A) and all-cause mortality (Y) in patients diagnosed with COVID-19

Results and Discussions

- There were 34,473 patients in the chronic SSRI-exposed group, and 523,301 patients in the SSRI-naive group
- The unadjusted crude analysis (i.e., no control for confounding) yielded an OR of **2.40 (95% C.I. [2.28 - 2.53]**) for all-cause 180-day mortality comparing the chronic SSRI vs SSRI-naïve groups. This is not surprising given that psychiatric diagnoses are known to have higher risk of mortality. The mortality estimates in both adjusted models were much lower but still elevated: the IPTW OR was **1.35 (95% C.I. [1.22 - 1.51]**), and the TMLE OR was: **1.29, 95%CI [1.19, 1.39]**)
- The elevated 180-day mortality risk in adjusted analysis suggests there is no treatment benefit from chronic exposure to any of the three SSRIs during the baseline period of 1 year prior to COVID-19 diagnosis. This result is similar but complementary to findings from prior studies that found no treatment effects from individual SSRIs taken as acute treatment for mild or moderate outpatient COVID-19.[3,4] However, our population was also broader since some patients were hospitalized or in the ICU at the time of COVID-19 diagnosis
- Chronic SSRI users had greater comorbidities than the naive cohort: cerebrovascular disease (88.1%; 74.0%), anxiety disorders (72.5%; 18.2%), hypertension (52.3%; 34.1%), COPD (27.7%; 15.3%), depression (27.0%; 3.4%), respectively (see Table 1)
- Limitations of our study include misclassification due to missing data and unmeasured data (for example, vaccination status, SSRI dosage, unmeasured confounders).
- In summary, the elevated mortality rate is likely due to residual confounding due to unmeasured or unknown confounders. Future research could use TMLE with time-to-event outcomes, and longitudinal TMLE (lTMLE) to address intercurrent events and time-varying exposure. These methods were not supported by the current N3C Enclave

Table 1: Baseline Characteristics

Variables	COVID-19 patients	
	Naive Cohort	Chronic Cohort
	N=523,301	N=34,473
Age (SD)	42.4 (21.8)	47.7 (20.3)
Age (At the time of covid diagnosis) (%)		
<18 years	73,564 (14.05)	2,374 (6.88)
18–39 years	144,788 (27.66)	10,705 (31.05)
40–59 years	162,059 (30.96)	10,648 (30.88)
60–79 years	119,293 (22.79)	8,512 (24.69)
>79 years	23,597 (4.5)	2,234 (6.48)
Sex (%)		
Male	219,349 (41.91)	9,640 (28)
Female	303,903 (58.07)	24,830‡ (72)
UNK	49	<20‡
Comorbidities (Incidence prior to index event) (%)		
Diabetes, Type 1	6,254 (1.19)	991 (2.87)
Diabetes, Type 2	79,228 (15.14)	8,759 (25.4)
Hypertension	178,480 (34.1)	18,022 (52.27)
Chronic pulmonary disease (COPD)	80,198 (15.32)	9,536 (27.66)
Chronic Kidney disease	35,026 (6.69)	5,375 (15.59)
Asthma	61,192 (11.69)	7,005 (20.32)
Rheumatological disorder	7,844 (1.49)	866 (2.51)
Obesity diagnosis (may include index date)	151,517 (28.95)	16,101 (46.7)
BMI data available (may include index date)	452,599 (86.48)	26,222 (76.06)
Obesity derived from BMI (may include index date) *(BMI Value between 30 and 100)*	209,343 (40)	15,519 (45.01)
Autoimmune disease	32,544 (6.21)	3,500 (10.15)
Myocardial infarction	15,496 (2.96)	2,858 (8.29)
Cerebrovascular disease	387,250 (74)	30,357 (88.06)
Chronic liver disease	37,577 (7.18)	5,392 (15.64)
Baseline Status (on index date, +/- 1 day max) (%)		
Hospitalized	39,092 (7.47)	4,672 (13.55)
ICU	2,048 (0.39)	278 (0.8)
ER	89,808 (17.16)	7,565 (21.94)
Outpatient	372,147 (71.11)	24,574 (71.28)
Intubation/Ventilation	5,123 (0.97)	691 (2.0)

‡ To comply with N3C policy, counts below 20 are displayed as <20 and the number of females was skewed by up to 5 in order to render it impossible to back-calculate precise counts in the 'UNK' gender

Conclusion

- This large population-based observational study demonstrates the feasibility of implementing TMLE in a realistic setting within the N3C Enclave
- Furthermore, this combined approach provided advantages in sharing data, and accessing advanced analytic methods to address the complex causal relationships between variables motivated by the clinically-informed DAG (e.g., direct effect modification, indirect effect modification, effect modification by proxy and effect modification by a common cause)
- The advantages of TMLE including stable and reasonable finite-sample performance, removal of bias due to model misspecification, the use of machine learning for efficient modeling of high-dimensional and complex data sources such as the N3C Enclave, offer clear benefits in this setting

References

1. Hashimoto, Y., Suzuki, T., & Hashimoto, K. (2022). Mechanisms of action of fluvoxamine for COVID-19: a historical review. Molecular Psychiatry, 27(4), 1898-1907.
2. Mark van der Laan, Jeremy Coyle, Nima Hejazi, Ivana Malenica, Rachael Phillips, Alan Hubbard. Welcome | Targeted Learning in R. Tlverse.org, tlverse.org/tlverse-handbook/.[Cited 2022Sep15]
3. Memorandum explaining basis for declining request for emergency use authorization of fluvoxamine maleate [Internet]. [cited 2022Sep15]. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2020/EUA%20110%20Fluvoxamine%20Decisional%20Memo_Redacted.pdf
4. Rauchman, S. H., Mendelson, S. G., Rauchman, C., Kasselman, L. J., Pinkhasov, A., & Reiss, A. B. (2022). Ongoing use of SSRIs does not alter outcome in hospitalized COVID-19 patients: a retrospective analysis. Journal of Clinical Medicine, 11(1), 70
5. VanderWeele TJ, Shpitser I. A new criterion for confounder selection. Biometrics. 2011 Dec;67(4):1406-13. doi: 10.1111/j.1541-0420.2011.01619.x. Epub 2011 May 31. PMID: 21627630; PMCID: PMC3166439.

Acknowledgement Statement

"The analyses described in this poster were conducted with data or tools accessed through the NCATS N3C Data Enclave https://covid.cd2h.org and N3C Attribution & Publication Policy v 1.2-2020-08-25b supported by NCATS U24 TR002306. This research was possible because of the patients whose information is included within the data and the organizations (https://ncats.nih.gov/n3c/resources/data-contribution/data-transfer-agreement-signatories) and scientists who have contributed to the ongoing development of this community resource [https://doi.org/10.1093/jamia/ocaa196]."