



HHS Public Access

Author manuscript

ICMHI 2021 (2021). Author manuscript; available in PMC 2023 November 10.

Published in final edited form as:

ICMHI 2021 (2021). 2021 May ; 2021: 296–303. doi:10.1145/3472813.3473206.

Causal AI with Real World Data: Do Statins Protect from Alzheimer’s Disease Onset?

Mattia Prosperi,

Department of Epidemiology, University of Florida

Marco Salemi,

Department of Pathology, University of Florida

Shantanu Ghosh,

Department of Epidemiology, University of Florida

Tianchen Lyu,

Department of Health Outcomes and Biomedical Informatics, University of Florida

Jiang Bian,

Department of Health Outcomes and Biomedical Informatics, University of Florida

Zhaoyi Chen,

Department of Health Outcomes and Biomedical Informatics, University of Florida

Jinying Zhao

Department of Epidemiology, University of Florida

Abstract

Causal artificial intelligence aims at developing bias-robust models that can be used to intervene on, rather than just be predictive, of risks or outcomes. However, learning interventional models from observational data, including electronic health records (EHR), is challenging due to inherent bias, e.g., protopathic, confounding, collider. When estimating the effects of treatment interventions, classical approaches like propensity score matching are often used, but they pose limitations with large feature sets, nonlinear/nonparallel treatment group assignments, and collider bias. In this work, we used data from a large EHR consortium –OneFlorida– and evaluated causal statistical/machine learning methods for determining the effect of statin treatment on the risk of Alzheimer’s disease, a debated clinical research question. We introduced a combination of directed acyclic graph (DAG) learning and comparison with expert’s design, with calculation of the generalized adjustment criterion (GAC), to find an optimal set of covariates for estimation of treatment effects –ameliorating collider bias. The DAG/CAC approach was assessed together with traditional propensity score matching, inverse probability weighting, virtual-twin/counterfactual random forests, and deep counterfactual networks. We showed large heterogeneity in effect estimates upon different model configurations. Our results did not exclude a protective effect of statins, where the DAG/GAC point estimate aligned with the maximum credibility estimate, although the 95% credibility interval included a null effect, warranting further studies and replication.

Keywords

Causal artificial intelligence; machine learning; biomedical informatics; electronic medical records; directed acyclic graph; Bayesian network; generalized adjustment criterion; treatment effect

1 INTRODUCTION

One of the desired capabilities of artificial intelligence (AI) in healthcare is to be ‘*interventional*’ in addition to predictive, i.e. being able to learn data-driven models that can evaluate different *what-if*s or actionable strategies –like treatment choices or behavioural changes [1], [2]. The main challenge with learning interventional models from observational data, including real-world data—those that are routinely collected outside of research settings such as electronic health records (EHR), is due to inherent bias, e.g., confounding, selection, indication bias. For instance, confounding happens when an exposure variable is spuriously thought to modify the odds of an outcome, due to correlation with an unmeasured factor (the confounder) that is the real cause of the outcome. Collider bias, instead, happens when an exposure and outcome are not directly related, but they independently cause a third variable, which in turn induces a mistaken causal effect of the exposure on the outcome [3]. In model learning, such biases can affect the feature selection process and lead to inclusion of non-causal variables or estimation of wrong causal effects [4]. As a result, a learnt model can have high accuracy in prediction but low performance in evaluating interventions, or calculating so-called counterfactuals [5]. One of the common biases in studies that attempt to investigate treatment effectiveness or drug repurposing from EHR data [6], is that the treatments were not randomized. The population that received treatment T for a disease A might differ substantially from the one that did not (e.g. the treated being at risk of another disease B , or the untreated having different healthcare access due to socioeconomic status), and any evidence of effectiveness on a subsequent health outcome might be affected by the population dissimilarities (e.g. the other disease B entails a care pathway with additional interventions that in turn reduce the risk of the target disease A , or the untreated population has poorer outcomes due to healthcare disparities). Thus, for a newly treated individual, who was not at risk for the other disease B or belongs to a health disparity group, the predicted health outcome for disease A would be favourable, whereas no treatment effect would be observed in reality.

Traditional statistical methods for balancing treatment assignments include propensity score matching (PSM) and inverse probability weighting (IPW) [7], with versions that deal with large covariate sets [8], although in practice they might not be preferable to full covariate adjustment with/without regularization [9], [10]. In addition, PSM and IPW have been mostly used within linear regression settings, thus posing serious limitations to nonlinear/nonparallel treatment assignment. Finally, in the presence of colliders (i.e. variables that are caused by both the treatment exposure and the target outcome), PSM could increase rather than reduce the bias in estimating the treatment effect [11]. To overcome the linearly-bounded limitations of PSM, machine learning approaches have been proposed, including Bayesian additive regression trees [12], random forests [13] and deep learning methods

[14]. Nonetheless, these methods are not yet used widely in practice and do not explicitly address the problem of collider bias. An alternative approach is to assume *a priori* causal relationships among variables, draw a directed acyclic graph (DAG) as a causal diagram, and then apply procedures such as the front-door or back-door criterion [15] to identify an appropriate set of adjustment covariates. The problem with DAGs is that the *a priori* knowledge on causal relationships (including unmeasured confounders) might be poor, making the approach prone to errors due to model misspecification. Algorithms to discover DAGs causal structures from data are available [16], but they often cannot resolve causal arc directions and yield equivalence classes; yet, a complete generalized adjustment criterion (GAC) is available for graphs with partially directed arcs [17].

In this work, we combine DAGs –both pre-specified and mined from data– and GAC with the objective to identify bias-free covariate sets to help with estimating treatment effects (assuming no unmeasured confounders). As a significant use case, we investigate the effect of statins, a class of cholesterol-lowering drugs, in reducing the risk of developing Alzheimer’s disease. In 2019, an estimated 5.8 million Americans live with Alzheimer’s disease. By 2050, people living with Alzheimer’s disease in U.S. may grow to 13.8 million, fueled by the aging baby boomers [18]. Nevertheless, there are still no effective options for prevention and treatment of Alzheimer’s disease. Previous observational studies and meta-analyses often yielded a strong protective effect of statins [19], [20], [21], [22], but randomized clinical trials reported absence of evidence [23], [24]. Moreover, a systematic review showed that, in four out of ten epidemiologic studies, the apparent protective effect of statins on risk of Alzheimer’s disease could be, at least partially, explained by the confounding effect of cholesterol or hyperlipidemia [25].

Exploring whether statins truly have a protective effect over Alzheimer’s disease is significant and our capability to accurately estimate its effect in a real-world setting with minimized bias is critical. Our goal here is to compare the DAG/GAC approach with traditional PSM/IPW, regularized linear regression, and machine learning methods, and demonstrate how choices of methods and parameters can sensibly change causal effect estimates in the same dataset; yet, the DAG/GAC approach possibly provides more robust reduction in the bias and variance of effect estimation.

2 MATERIAL AND METHODS

2.1 Definition of treatment effects

Suppose that the effect of a treatment T on a health outcome Y is being investigated in a population sample of n individuals. Each subject i is defined by the tuple $\{X_i, T_i, Y_i\}$, where $X_i = \{x_{i1} \dots x_{im}\}$ is a vector of pre-treatment characteristics of the subject. Let Y_i^1 and Y_i^0 be the potential outcomes of subject i when administered treatment $T_i = 1$, or in absence of treatment $T_i = 0$, respectively. The individual treatment effect (ITE) $\tau(x)$ for individuals whose $X_i = x$ is defined as the average of the difference in potential outcomes under both treatments given x , i.e. $\tau(x) = E[Y^1 - Y^0 | X = x]$. This ITE formulation—counterfactual—is usually not directly calculable because individuals cannot be both treated and not treated at the same time. However, if the potential outcomes are independent of the

treatment conditional on background variables, i.e., $\{Y^1, Y^0\} \perp T \mid X$ an assumption of strongly ignorable treatment assignment (SITA), the ITE can then be calculated as $\tau(x) = E[Y^1 \mid T = 1, X = x] - E[Y^0 \mid T = 0, X = x] = E[Y \mid T = 1, X = x] - E[Y \mid T = 0, X = x]$. By averaging over the distribution of X , the average treatment effect (ATE) τ_{01} can be calculated as $\tau_{01} = E[\tau(X)] = E[Y \mid T = 1] - E[Y \mid T = 0]$. When the outcome is binary, e.g. diagnosis of an illness, the odds ratio is also often used in place of the ATE. Under SITA, ITE and ATE can be calculated only when the treatment groups have the same X , which becomes practically unfeasible as the dimension of X increases. PSM, through the conditional probability $\pi(x) = \Pr(T = 1 \mid X = x)$, tries to balance the probability of receiving T given $X = x$ by pairing individuals in treatment and control groups with similar propensity scores calculated on the basis of their baseline covariates. The pairing can be done in different ways, including nearest neighbor and Caliper matching, possibly leading to exclusion of unmatched samples, or all cases can be considered through IPW.

2.2 Data source, study design, ethical statement

We extracted data from the OneFlorida Clinical Research Consortium, a statewide clinical research network and database (<https://www.ctsi.ufl.edu/ctsa-consortium-projects/oneflorida/>) contributing to the national Patient-Centered Clinical Research Network (PCORnet). OneFlorida's partners –hospitals, practice/clinic settings and physicians– provide healthcare to more than 15 million Floridians (60%+ of the state population). The OneFlorida data contain robust longitudinal and linked patient-level RWD, including data from claims, cancer registry, vital statistics, and EHRs from its clinical partners. The OneFlorida data is a HIPAA limited data set and follows the PCORnet Common Data Model (CDM) that contains detailed patient and clinical variables, including demographics, encounters, diagnoses, procedures, vitals, medications, and labs. Based on the PCORnet CDM, we used the International Classification of Diseases, 9th/10th revision, Clinical Modification (ICD-9/10-CM) to encode clinical conditions, RxNorm and National Drug Code (NDC) for medications, and Logical Observation Identifiers Names and Codes (LOINC) for laboratory tests. From the OneFlorida database, we included people who lived in Florida, had at least three prior years of medical records (with one or more visit per year), and who were not yet diagnosed with Alzheimer's disease (ICD-9: 331.0; ICD-10: G30.0, G30.1, G30.8, G30.9) as of their 65th birthday –index date. We extracted the most recent information associated to the index date for each subject, including demographics (gender, race/ethnicity, zip-code of residence), insurance type (private, public, or uninsured), smoking status (current, former, never smoker), body-mass index (BMI), Charlson's comorbidity index (CCI)[36], vitals and laboratory findings –high-density lipoprotein (HDL, LOINC 2085–9) and low-density lipoprotein (LDL, 2089–1), total cholesterol (2093–3), triglycerides (2571–8), and Glycated hemoglobin (HbA1c, 4845–4). Other covariates included end-stage renal disease (ICD-9 585.6; ICD-10 N18.6) and Lou Gehrig's disease (335.20; G12.21), both making eligibility for Medicare prior to age 65, and relevant comorbidities associated with Alzheimer's onset –sleep disorders (327; G47) anxiety (300; F41), depression (311; F32, F33), hypertension (401; I10), diabetes (250, E10, E11), heart disease (402, 416, 429; I11, I27; I51), alcohol use disorders (291, 303; F10). We recorded usage (number of prescriptions and time frame) of statins,

nonsteroidal anti-inflammatory drugs (NSAID) –aspirin, ibuprofen, naproxen, paracetamol– and anti-hypertension drugs (49 active ingredients including bumetanide, chlorthalidone, chlorothiazide, ethacrynate, furosemide, hydrochlorothiazide, indapamide, methyclothiazide, metolazone, and torsemide). The statins examined were atorvastatin, simvastatin, and sitagliptin. Statins, NSAID and anti-hypertension drugs needed to be administered at least for 6 weeks before the index date, prescribed at least twice in the patient’s EHR. The outcome was set as the diagnosis of Alzheimer’s after the index date, with associated time, and latest follow up time for those who were not diagnosed. Since the sample size of those who did not have a diagnosis after index date was largely surpassing the number of Alzheimer’s cases, we matched subjects by zip-code of residence, limiting the control-to-case ratio to a max of 20-to-1. Missing values in the dataset were not imputed, but coded into an additional unknown category; laboratory variables with missing data were discretized using multi-valued clinical thresholds for normal, borderline, or abnormal levels.

The authors abide to the Declaration of Helsinki; the study was approved under University of Florida’s institutional review board (IRB), protocol no. IRB201900182.

2.3 Analysis

We estimated the crude (unadjusted) ATE of statin with respect to Alzheimer’s disease onset, regardless of time to diagnosis, and reported the odds ratio. Then we estimated adjusted odds ratios using: (a) multivariable main-effects boosted logistic regression (LogitBoost) [26]; (b) PSM and (c) IPW using either LogitBoost or random forests, with nearest-neighbor matching (distinct as well as one-to-many)[7]; (d) virtual twin [27] random forests, which allow for direct ITE calculation by predicting potential outcomes, flipping the treatment variable for each observation, upon which the odds ratio were estimated; (e) the counterfactual random forest [13], which also predicts potential outcomes by fitting two independent forests, one for the treated cases and one for the controls; (f) deep counterfactual network[14], a deep learning neural architecture trained in alternate batches with a so-called propensity dropout [27] akin to PSM.; and (g) the DAG/GAC approach.

The DAG/GAC approach was twofold. First, we draw a DAG using evidence from the literature on Alzheimer’s risk factors, mediators, and their relationships, using all covariates, without unmeasured confounding nodes. Then a DAG structure and parameters was learnt from the OneFlorida dataset using a hill-climbing algorithm [28]. On both DAGs, we derived all adjustment sets using the GAC (for total effect), and then estimated the odds ratio of statins for each adjustment set using multivariable logistic regression on the outcome of Alzheimer’s disease onset. Predictive performance was measured by means of the area under the receiver operating characteristics (AUROC) using bootstrapping (25 replicates) and out-of-bag estimates. The random forests’ number of trees were optimized by performing a grid search between 25–2,500 trees, choosing the value that minimized the out-of-bag error. The parameters of the deep counterfactual network were optimized using the Adam [29] method. The predicted values of the random forest and of deep counterfactual network models were re-calibrated using the population prevalence as target, and we also performed a quasibinomial regression on normalized probabilities to further address potential issues with calibration. Rather than pooling standard errors of odds ratios within and between

models, which would be biased by sample overlap from bootstrap runs [30], we estimated a maximum credibility range which corresponded to the model estimate closest to all other estimates (similar to maximum credibility clade calculation in trees) [31]. All analyses were carried out using R (<https://www.r-project.org/>) and PyTorch (<https://pytorch.org/>).

3 RESULTS

The study population comprised a total of 13,780 individuals seen between 2012 and 2020, of which 1,618 (11.7%) were exposed to statins before the index date, and 977 (7.09%) developed Alzheimer's disease after the index date. The crude odds ratio of developing Alzheimer's upon statin exposure was $(53/1,565)/(924/11,238) = 0.41$ (95% CI 0.31–0.55). Table 1 shows the population characteristics overall and stratified by statin exposure (note, some variables in the table were grouped to meet formatting/page display constraints, but not in the analyses).

Using the full, non-matched dataset, we estimated the predictive ability of the multivariable models with respect to diagnosis of Alzheimer's disease outcome. The AUROCs of all models indicated mild discrimination ability (0.69–0.72), as shown in 1 (panel A), with the DAG yielding an AUROC ~3% less than the other methods ($P=0.22$).

We then performed PSM for the statin exposure using LogitBoost, random forests (optimal number of trees = 1,300) and obtained the propensity dropout scores from the deep counterfactual network. All scores were highly correlated (Pearson's $\rho=0.92$ between LogitBoost and random forests, $\rho=0.88$ between LogitBoost and deep network, and $\rho=0.87$ between random forest and deep network, all $P<0.0001$). The predictive ability of propensity scores to correctly identify if a subject was treated with statin was high for all models (AUROC 0.875 for LogitBoost, 0.870 for random forest, and 0.91 for the deep network), as shown in 1 (panel B). After matching, the variable distributions of the non-treated were –as expected– much more similar to the statin-exposed, although a perfect prevalence match could not be obtained for all variables (e.g., heart disease). 2 provides PSM summary for most of study variables.

After evaluating the predictive performance, we analyzed the adjusted odds ratio of statin exposure toward Alzheimer's onset for all the models, using the (a)-(f) approaches and the maximum credibility range as summary estimate. All approaches yielded odds ratios higher than the crude estimate, mostly between 0.5 and 1.0 as shown in 2, still suggesting that statins had a protective effect. The LogitBoost and PSM methods yielded the largest confidence intervals, while IPW methods had smaller ranges but were also less concordant in terms of point estimates. The virtual twin / counterfactual random forest also showed narrow confidence intervals, but their point estimates were very different depending on the cutoff used. With probability normalization, though, the cutoff optimization stabilized. Similar instability was observed for the deep counterfactual network, where we were able to calculate only the quasi-binomial estimate, as all cutoff threshold examined yielded division by zero in at least one of the denominators used for the odds ratio. Conversely, the odds ratio estimates for the DAG/GAC were very similar among all adjustment sets, and even between the data-mined and expert-drawn DAG topologies. Their confidence intervals were

also smaller than the PSM and LogitBoost methods. The maximum credibility range was supported by two models that had the same smallest distance value from all others, namely the data-mined DAG/GAC and the PSM (duplicated) random forest. The final estimate of the adjusted odds ratio was 0.82 (95% CI 0.57–1.19), which indicated a protective effect of statin, yet the 95% credibility interval included also the null effect, i.e. 1.0.

In order to clarify how the DAG/GAC selects confounders and excludes colliders/irrelevant variables, we show in 3 the nodes of the adjustment set for both the expert-drawn (panel **A**) and the data-mined (panel **B**) DAGs. Of note, not all nodes are necessarily used in an adjustment set at the same time. For instance, the adjustment sets for the data-mined DAG were: {anti-hypertensive medications, year}; {anti-hypertensive medications, insurance type, NSAID, smoking status}; {anti-hypertensive medications, BMI, CCI, hypertension, NSAID, smoking status}; {anti-hypertensive medications, BMI, diabetes, NSAID, smoking status}; {HbA1c, anti-hypertensive medications, BMI, NSAID, smoking status}; {anti-hypertensive medications, BMI, HDL, NSAID, smoking status}; {anti-hypertensive medications, NSAID, TGL}.

4 DISCUSSION AND CONCLUSION

Our study applied causal AI methods on a real-world observational dataset to investigate if statin treatment reduced the risk of Alzheimer's disease. Our findings indicate that there may be a moderately protective effect of statins, although the 95% credibility range did not exclude a null effect.

The DAG/GAC approach and the PSM yielded estimates all close to the maximum credibility range; however, the confidence intervals for the PSM methods were larger, likely due to the reduced sample size. The DAG/GAC estimates were similar among the expert-drawn and the data-mined structures, which is interesting given that automated learning usually cannot resolve the direction ambiguities in the DAGs. The counterfactual forest (and to some extent also the deep counterfactual network) yielded very unstable estimates upon bootstrapping and by changing the calibration cutoff. One possible explanation is that, for these methods, two models were fit separately; the one fitted on the treated included a much smaller sample size and even smaller number of events, making approximation of probabilities by votes as well as calibration more difficult. However, probability normalization helped to stabilize the output comparisons.

All the methods that we tested here in principle were robust with respect to (measured) confounding and in part to treatment stratification bias, but are prone to be affected by collider bias, except for the DAG/GAC approach, as the DAG/GAC approach explicitly eliminated collider bias rooted in its principle. Another issue is calculation of direct vs. total effects, i.e., how to account the contribution of mediators (here, for instance, LDL) to the causal effect estimation. The DAG/GAC approach helps eliminating colliders and allows calculation of direct and total effects. Nevertheless, the DAG/GAC approach is not free from misspecification of the causal structure and from unmeasured confounding. In fact, the mild predictive ability of all the models to identify Alzheimer's on-set suggests that we may have missed a part of the equation –there are other (or latent) factors that were not included in

our models. Although it is clear that a perfect prediction model is not achievable because of these missing factors, their impact on causal models and causal estimates raises curiosity. It is unclear whether our results that statins have moderately protective effect resemble a true causal effect, and whether the wide credibility range is influenced more by the small sample size or by the missing factors. Nevertheless, such problem is not unique to applying causal AI methods on observational data; for example, in RCTs, the assumption that randomization can automatically lead to strong ignorability also ignores the latent factors that are not measurable or apparent as they appear (e.g., patients who have no access to the RCTs would not even be able to enter the randomization process).

In regards to the clinical interpretation of the findings, the observed protective effect could be further reduced by identifying additional confounder and incorporating mediation. Hypertension, diabetes mellitus, and metabolic syndrome have been found associated with the development of dementia in middle-aged adults, whereas the effects of hypercholesterolemia, atrial fibrillation, and smoking have been less clear [32]. From a biological mechanism point of view, the apolipoprotein E gene, which has a role in transportation and modulation of cholesterol, seems involved in the development of Alzheimer's disease [33]; thus a role of statins, which lower cholesterol, might be expected [24]. Nevertheless, these are still speculations but raise questions on how to generate the necessary evidence to move the hypothesis that statins can be a preventative strategy for Alzheimer's disease to clinical practice. Is it time to conduct a large-scale RCT and wait for 10 years or is there sufficient evidence that can be generated from large-scale real-world observational data to convince the regulatory agencies? [34] The U.S. Food and Drug Administration (FDA) coined the term real-world data and real-world evidence recently (<https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>) and provided guidance on “*Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drugs and Biologics*” (<https://www.fda.gov/media/124795/download>) such as “*observational studies that generate RWE [Real-World Evidence] intended to help to support an efficacy supplement.*” Given the results from this study, we argue that further developments of statistical and machine learning methods are warranted, especially the causal AI approach demonstrated in this paper.

Our study has a number of limitations. First, we did not take into account the time to Alzheimer's onset, and censoring time for those who did not (yet) develop it at the end of their available follow up. Indeed, the median follow-up time –not used as a covariate– was quite different between statin and non-statin groups; this warrants further considerations on the analytical design, including the usage of survival models. Second, we grouped together three different statin drugs, and we did not analyze the differences in dosage and time of exposure of statins, whereas studies have reported that potency and the cumulative duration may play a critical role [22]. Third, DAG learning algorithms usually do not scale well with the number of covariates, and presence of noise/irrelevant variables does not help with structure learning. We foresee further improvements of our approach: one is to incorporate feature selection in the DAG structure learning process, or to inform the learning using prior knowledge; another, which applies to all methods, is to try to reduce bias already in the dataset creation, for instance using a target trial design approach [35, 36].

Unveiling the effect of statin exposure is difficult because this treatment is very common among middle-aged adults, and indicated for a variety of risks and comorbidities; therefore, confounding, mediation and collider bias are very likely to affect estimates from observational data, especially EHR. Causal AI methods, including the DAG/GAC framework presented here, can help in tackling the bias in observational studies and provide models for both prediction and intervention. These causal AI methods warrant further investigations, as real-world data and real-world evidence are playing increasingly important roles in the drug development process.

ACKNOWLEDGMENTS

Acknowledgments are placed before the references. Add information about grants, awards, work was supported in part by NIH grants UL1TR001427, R01CA246418, R21CA245858, R21AG068717, U18DP006512, and PCORI grant ME-2018C3-14754. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH and PCORI.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

REFERENCES

- [1]. Prosperi M et al. 2020. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*. (2020). DOI:10.1038/s42256-020-0197-y.
- [2]. Sperrin M et al. 2019. Explicit causal reasoning is needed to prevent prognostic models being victims of their own success. *Journal of the American Medical Association*.
- [3]. Sackett DL. 1979. Bias in analytic research. *Journal of Chronic Diseases*. (1979). DOI:10.1016/0021-9681(79)90012-2.
- [4]. Fang G et al. 2019. Applying machine learning to predict real-world individual treatment effects: Insights from a virtual patient cohort. *Journal of the American Medical Informatics Association*. (2019). DOI:10.1093/jamia/ocz036.
- [5]. Hernán MA et al. 2019. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. *CHANCE*. (2019). DOI:10.1080/09332480.2019.1579578.
- [6]. Pushpakom S et al. 2018. Drug repurposing: Progress, challenges and recommendations. *Nature Reviews Drug Discovery*.
- [7]. Austin PC. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*. (2011). DOI:10.1080/00273171.2011.568786.
- [8]. Schneeweiss S et al. 2009. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. (2009). DOI:10.1097/EDE.0b013e3181a663cc.
- [9]. Elze MC et al. 2017. Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. *Journal of the American College of Cardiology*.
- [10]. Tian Y et al. 2018. Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *International Journal of Epidemiology*. (2018). DOI:10.1093/ije/dyy120.
- [11]. Pearl J. 2009. Remarks on the method of propensity score. *Statistics in Medicine*.
- [12]. Hill JL. 2011. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*. (2011). DOI:10.1198/jcgs.2010.08162.

- [13]. Lu M et al. 2018. Estimating Individual Treatment Effect in Observational Data Using Random Forest Methods. *Journal of Computational and Graphical Statistics*. (2018). DOI:10.1080/10618600.2017.1356325.
- [14]. Alaa AM et al. 2017. Deep Counterfactual Networks with Propensity-Dropout.
- [15]. Pearl J. 2009. *Causality: Models, Reasoning and Inference*. Cambridge University Press.
- [16]. Glymour C et al. 2019. Review of causal discovery methods based on graphical models. *Frontiers in Genetics*. (2019). DOI:10.3389/fgene.2019.00524.
- [17]. Perkovi E et al. 2015. A complete generalized adjustment criterion. *Uncertainty in Artificial Intelligence - Proceedings of the 31st Conference, UAI 2015* (2015).
- [18]. Hebert LE et al. 2013. Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. *Neurology*. (2013). DOI:10.1212/WNL.0b013e31828726f5.
- [19]. Chu C-S et al. 2018. Use of statins and the risk of dementia and mild cognitive impairment: A systematic review and meta-analysis. *Scientific Reports*. 8, 1 (2018), 5804. DOI:10.1038/s41598-018-24248-8. [PubMed: 29643479]
- [20]. Geifman N et al. 2017. Evidence for benefit of statins to modify cognitive decline and risk in Alzheimer's disease. *Alzheimer's Research & Therapy*. 9, 1 (2017), 10. DOI:10.1186/s13195-017-0237-y.
- [21]. Poly TN et al. 2020. Association between Use of Statin and Risk of Dementia: A Meta-Analysis of Observational Studies. *Neuroepidemiology*. 54, 3 (2020), 214–226. DOI:10.1159/000503105. [PubMed: 31574510]
- [22]. Zhang X et al. 2018. Statins use and risk of dementia: A dose–response meta analysis. *Medicine (United States)*.
- [23]. S. BG et al. 2018. The role of statins in both cognitive impairment and protection against dementia: A tale of two mechanisms. *Translational Neurodegeneration*. (2018). DOI:10.1186/s40035-018-0110-3
- [24]. Sano M et al. 2011. A randomized, double-blind, placebo-controlled trial of simvastatin to treat Alzheimer disease. *Neurology*. (2011). DOI:10.1212/WNL.0b013e318228bf11.
- [25]. Bykov K et al. 2017. Confounding of the association between statins and Parkinson disease: systematic review and meta-analysis. *Pharmacoepidemiology and Drug Safety*. (2017). DOI:10.1002/pds.4079.
- [26]. Friedman J et al. 2000. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics*. (2000). DOI:10.1214/aos/1016218223.
- [27]. Foster JC et al. 2011. Subgroup identification from randomized clinical trial data. *Statistics in Medicine*. (2011). DOI:10.1002/sim.4322.
- [28]. Gal Y and Ghahramani Z. 2016. Dropout as a Bayesian Approximation: Appendix. 33rd International Conference on Machine Learning, ICML 2016 (2016).
- [29]. Nagarajan R et al. 2013. *Bayesian Networks in R: with Applications in Systems Biology*.
- [30]. Kingma DP and Ba JL. 2015. Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings (2015).
- [31]. Nadeau C and Bengio Y. 2003. Inference for the generalization error. *Machine Learning*. 52, 3 (2003), 239–281. DOI:10.1023/A:1024068626366.
- [32]. Heled J and Bouckaert RR 2013. Looking for trees in the forest: summary tree from posterior samples. *BMC Evolutionary Biology*. 13, 1 (2013), 221. DOI:10.1186/1471-2148-13-221. [PubMed: 24093883]
- [33]. Duron E and Hanon O. 2008. Vascular risk factors, cognitive decline, and dementia. *Vascular Health and Risk Management*.
- [34]. Eichner JE et al. 2002. Apolipoprotein E polymorphism and cardiovascular disease: A HuGE review. *American Journal of Epidemiology*.
- [35]. Sherman RE et al. 2016. Real-world evidence - What is it and what can it tell us? *New England Journal of Medicine*. (2016). DOI:10.1056/NEJMs1609216.

- [36]. Hernán MA and Robins JM. 2016. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *American Journal of Epidemiology*. (2016). DOI:10.1093/aje/kwv254.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

CCS CONCEPTS

- Computing methodologies; • Machine learning; • Machine learning algorithms;

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

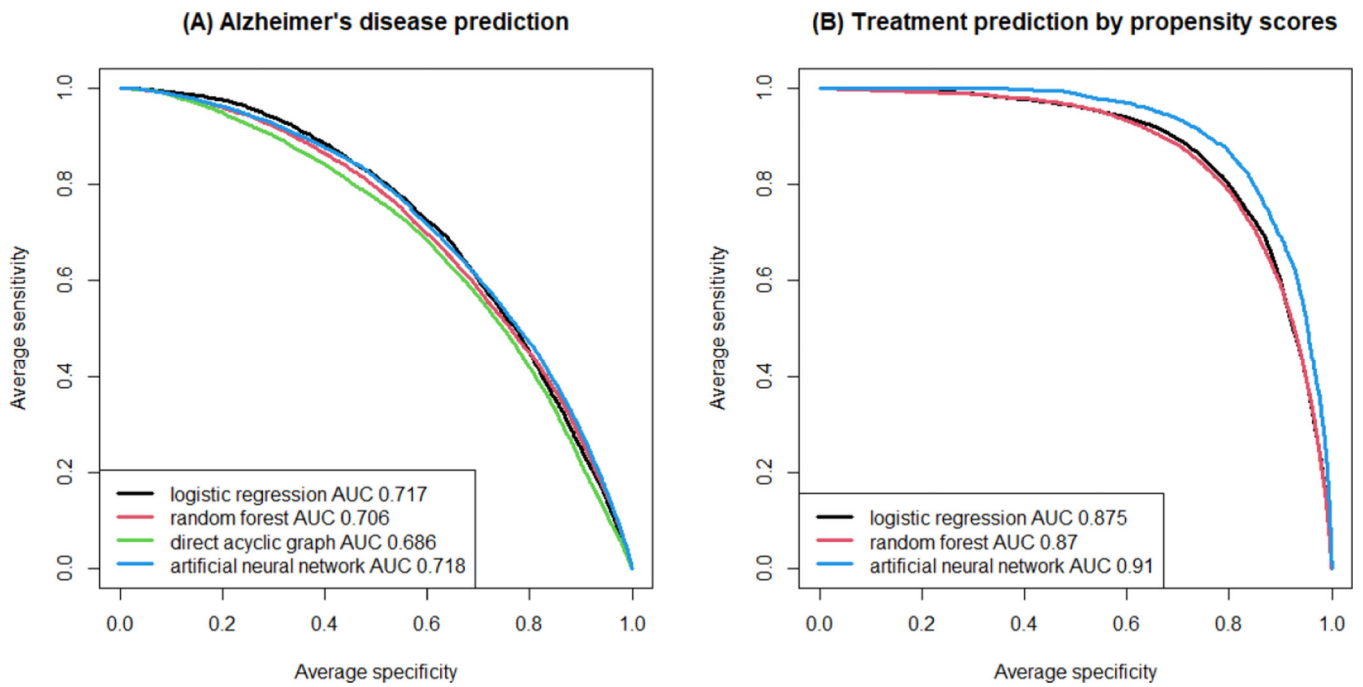


Figure 1: Discriminative performance –upon out-of-bag predictions– of (A) fully-adjusted models for diagnosis of Alzheimer’s disease after index date, and of (B) propensity scores for the probability of being prescribed statins.

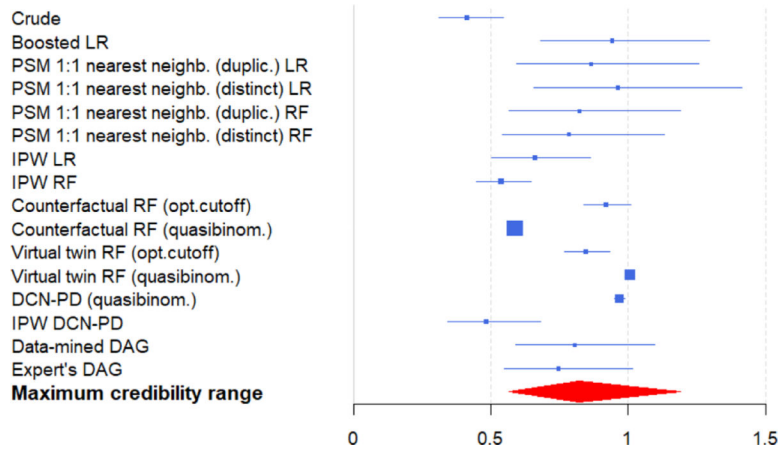


Figure 2: Adjusted odds ratio estimates of the effect of statin exposure toward development of Alzheimer’s disease using different approaches and their maximum credibility range

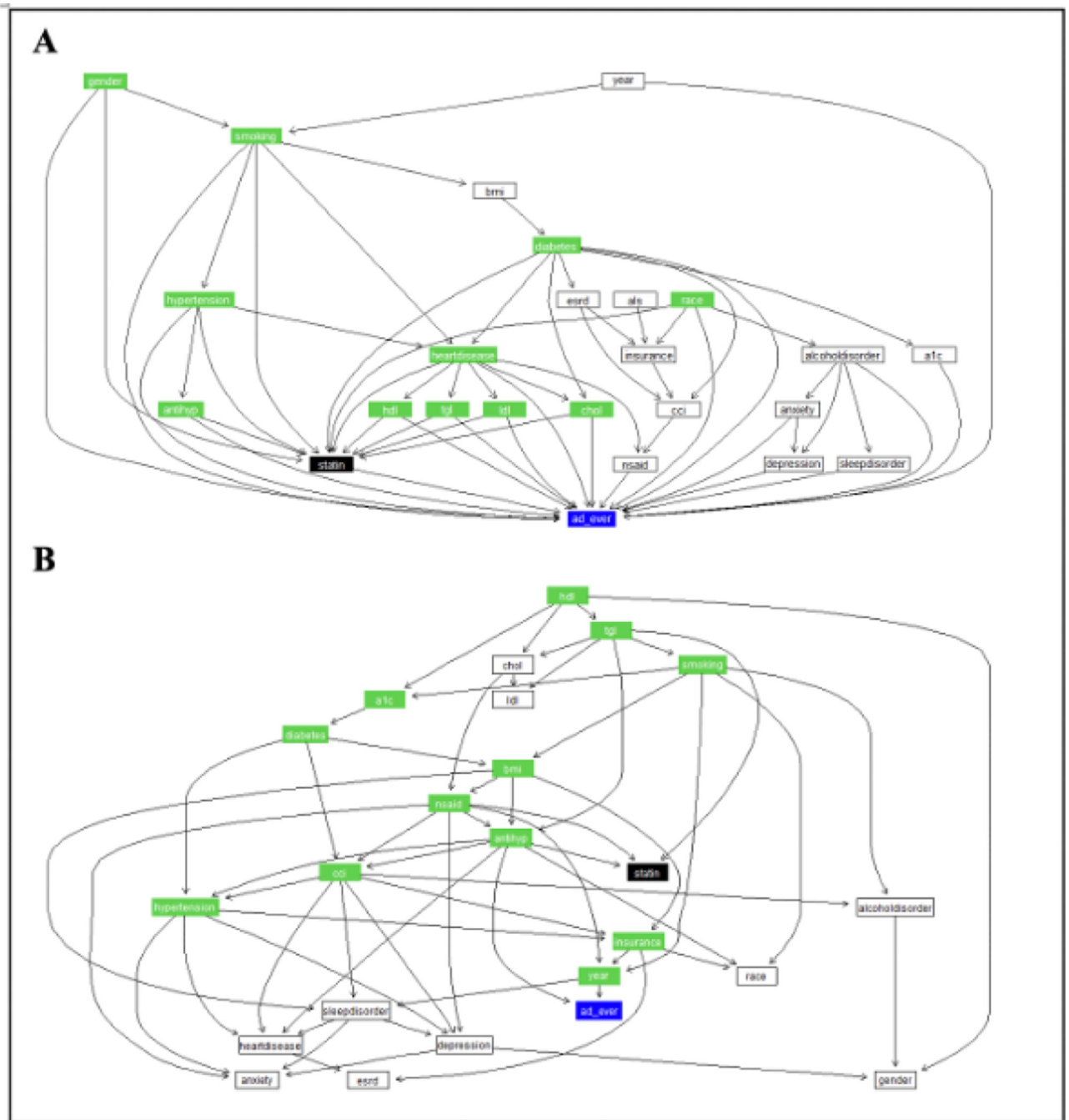


Figure 3: Expert-drawn (A) and data-mined (B) directed acyclic graphs for estimating the effect of statin exposure with respect to Alzheimer’s onset. The query treatment (statin) is indicated in black and the outcome (Alzheimer’s) in blue, while the nodes included in adjustment sets are displayed in green.

Table 1:

Characteristics of the study population, stratified by statin exposure.

Variable	All patients (13,780)	No statins administered(12,162; 88.3%)	Statin-treated(1,618; 11.7%)
Sex			
Male	5542 (40.2%)	4824 (39.7%)	718 (44.4%)
Female	8238 (59.8%)	7338 (60.3%)	900 (55.6%)
Race/Ethnicity			
Hispanic	2982 (21.6%)	2716 (22.3%)	266 (16.4%)
Non-Hispanic Black	2992 (21.7%)	2502 (20.6%)	490 (30.3%)
Non-Hispanic White	7206 (52.3%)	6398 (52.6%)	808 (49.9%)
Other	412 (3%)	373 (3.1%)	39 (2.4%)
Unknown	188 (1.4%)	173 (1.4%)	15 (0.9%)
Outcome			
Alzheimer's disease onset	977 (7.1%)	924 (7.6%)	53 (3.3%)
Socioeconomic factors			
Private health insurance	3283 (23.8%)	2897 (23.8%)	386 (23.9%)
Medicare/Medicaid/Government ins.	4667 (33.9%)	4190 (34.5%)	477 (29.5%)
Other/unknown health insurance	5830 (42.3%)	5075 (41.7%)	755 (46.7%)
Behavioral factors			
Current smoker	1149 (8.3%)	929 (7.6%)	220 (13.6%)
Former smoker	2119 (15.4%)	1662 (13.7%)	457 (28.2%)
Never smoker	2978 (21.6%)	2474 (20.3%)	504 (31.1%)
Unknown smoking status	7534 (54.7%)	7097 (58.4%)	437 (27%)
Alcohol disorders	693 (5%)	561 (4.6%)	132 (8.2%)
Vitals			
Obesity/overweight	7797 (56.6%)	6489 (53.4%)	1308 (80.8%)
Comorbidities			
Charlson's Comorbidity Index >5	3487 (25.3%)	2770 (22.8%)	717 (44.3%)
Heart disease	2885 (20.9%)	2244 (18.5%)	641 (39.6%)
Diabetes	5649 (41%)	4600 (37.8%)	1049 (64.8%)
Hypertension	9355 (67.9%)	7916 (65.1%)	1439 (88.9%)

Variable	All patients (13,780)	No statins administered(12,162; 88.3%)	Statin-treated(1,618; 11.7%)
Sleep disorders / depression / anxiety	9903 (71.9%)	4902 (40.3%)	990 (61.2%)
Lab tests			
Cholesterol >200	957 (6.9%)	729 (6%)	228 (14.1%)
High-density Lipoprotein <60	2618 (19%)	1816 (14.9%)	802 (49.6%)
Triglycerides >200	512 (3.7%)	319 (2.6%)	193 (11.9%)
Low-Density Lipoprotein >100	423 (3.1%)	123 (1%)	36 (2.2%)
Hemoglobin A1C >5.7%	2097 (15.2%)	1478 (12.2%)	619 (38.3%)
Other medications			
Nonsteroidal Anti-inflammatory Drugs	4304 (31.2%)	3145 (25.9%)	1159 (71.6%)
Anti-hypertensive medications	4222 (30.6%)	2871 (23.6%)	1351 (83.5%)
Time period			
Follow-up time >3 years	5464 (39.7%)	5083 (41.8%)	381 (23.5%)
Calendar year >2016	5642 (40.9%)	4648 (38.2%)	994 (61.4%)

Table 2: Characteristics of the control population (not exposed to statin) after propensity score matching (n=1,618)

Variable	Statin-treated	Matched controls	LogitBoost controls	Random forest
Sex Female	900 (55.6%)	925 (57.2%)	953 (58.9%)	
Race/Ethnicity Hispanic	266 (16.4%)	267 (16.5%)	294 (18.2%)	
Non-Hispanic Black	490 (30.3%)	475 (29.4%)	483 (29.9%)	
Non-Hispanic White	808 (49.9%)	794 (49.1%)	771 (47.7%)	
Behavioral factors Current smoker	220 (13.6%)	205 (12.7%)	212 (13.1%)	
Alcohol disorder	132 (8.2%)	155 (9.6%)	141 (8.7%)	
Vitals Obesity/overweight	1308 (80.8%)	1340 (82.8%)	1207 (74.6%)	
Comorbidities Charlson's Comorbidity Index >5	717 (44.3%)	728 (45%)	706 (43.6%)	
Heart disease	641 (39.6%)	555 (34.3%)	593 (36.7%)	
Diabetes	1049 (64.8%)	1034 (63.9%)	1011 (62.5%)	
Hypertension	1439 (88.9%)	1439 (88.9%)	1408 (87%)	
Sleep disorders / depression / anxiety	990 (61.2%)	719 (44.4%)	781 (48.3%)	
Lab tests Cholesterol >200	228 (14.1%)	212 (13.1%)	259 (16%)	
Other medications Nonsteroidal Anti-inflammatory Drugs	1159 (71.6%)	1150 (71.1%)	1092 (67.5%)	
Anti-hypertensive medications	1351 (83.5%)	1361 (84.1%)	1229 (76%)	