Sage Open Aging

Estimated effects of comorbidities on risk of all-cause dementia in patients with mild cognitive impairment

Journal:	*Sage Open Aging*
Manuscript ID	GGM-24-0224.R1
Manuscript Type:	Original Manuscript
Keywords:	Alzheimer's/Dementia, Quantitative Methodology, Statistical analysis, Veterans, Comorbidity

Abstract:	INTRODUCTION: Estimating the effects of comorbidities on risk of all-cause dementia (ACD) could potentially better inform prevention strategies than more common post-hoc analyses from predictive modeling. METHODS: In a retrospective cohort study of patients with mild cognitive impairment (MCI) from US Veterans Affairs Medical Centers between 2009-2021, we used machine learning techniques from treatment effect estimation to estimate individualized effects of 25 comorbidities (e.g., hypertension) on ACD risk within 10 years. Age and healthcare utilization were adjusted for using exact matching. RESULTS: After matching, of 19,797 MCI patients, 6,767 (34.18%) experienced ACD onset. Three comorbidities had consistently non-zero average effects: dyslipidemia (percentage point increase of ACD risk range across techniques=0.009-0.044), hypertension (range=0.007-0.043), diabetes (range=0.007-0.191). DISCUSSION: Our findings suggest associations between dyslipidemia, hypertension, and diabetes that increase ACD risk in MCI patients. Early treatment for these comorbidities could delay ACD onset. The approaches used can also potentially identify novel risk factors.

Abstract

INTRODUCTION: Estimating the effects of comorbidities on risk of all-cause dementia (ACD) could potentially better inform prevention strategies and identify novel risk factors than more common post-hoc analyses from predictive modeling.

METHODS: In a retrospective cohort study of patients with mild cognitive impairment (MCI) from US Veterans Affairs Medical Centers between 2009-2021, we used machine learning techniques from treatment effect estimation to estimate individualized effects of 25 comorbidities (e.g., hypertension) on ACD risk within 10 years. Age and healthcare utilization were adjusted for using exact matching.

RESULTS: After matching, of 19,797 MCI patients, 6,767 (34.18%) experienced ACD onset. Dyslipidemia (percentage point increase of ACD risk range across techniques=0.009-0.044), hypertension (range=0.007-0.043), diabetes (range=0.007-0.191) consistently non-zero average effects.

DISCUSSION: Our findings suggest associations between dyslipidemia, hypertension, and diabetes that increase ACD risk in MCI patients and show the potential for these approaches to identify novel risk factors.

Key Words

Alzheimer's/Dementia, quantitative methodology, statistical analysis, veterans, comorbidity

1. Introduction

All-cause dementia (ACD) is a leading cause of death among individuals 65 years and older, and understanding what contributes to ACD onset in patients with mild cognitive impairment (MCI) could inform treatment and prevention (Alzheimer's Association, 2024). Past work has shown that intervening on modifiable lifestyle factors, such as diet and exercise, may slow cognitive decline (Rosenberg et al., 2018). Identifying factors related to ACD using machine learning (ML) can stimulate hypothesis generation, which can further aid in designing interventions to reduce ACD risk. However, current work using ML to identify risk factors for ACD onset focuses on post-hoc analyses from predictive modeling (Jo et al., 2019; Tjandra et al., 2020; Tang et al., 2024; Irwin et al., 2024). For example, past studies have used observational data, such as the Alzheimer's Disease Neuroimaging Initiative (Jo et al., 2019; Devarakonda et al., 2019) and electronic health records (EHRs) (Tjandra et al., 2020; Tang et al., 2024; Irwin et al., 2024), to predict onset of Alzheimer's disease and then identified risk factors as the features whose contribution to predictive performance was significant. In contrast, we aim to directly estimate the individual effects of common comorbidities on risk of ACD rather than measuring post-hoc feature importance. If the effect is strong, then treating the comorbidity could reduce ACD risk.

Ideally, we aim to identify comorbidities that cause ACD onset. However, verifying whether the relationship between a comorbidity (e.g., hypertension) and ACD onset is causal requires a randomized controlled trial (RCT), and for many potential risk factors,

an RCT is infeasible. In light of this, we investigate the applicability of ML approaches to an observational cohort to estimate the effects of comorbidities identified by the literature as risk factors for ACD onset (e.g., hypertension, and hearing loss). The approaches we use generally aim to quantify the effect of a feature (e.g., hypertension) on an outcome (e.g., ACD onset) from observational data. Under a set of standard assumptions, outlined below, these approaches can be used to estimate the effect of the feature on the outcome. Such knowledge can provide a focused set of hypotheses for future work in the clinical space to test. In past work, these approaches have been used in assessing treatment effects (Xu et al., 2023a). In the context of ACD onset, we aim to answer how much known risk factors change the risk of ACD onset over a 10-year horizon. Here, we measure the change in risk of ACD onset (i.e., the estimated effect) as the percentage point increase of having the risk factor compared to not having the risk factor. Hence, measuring the estimated effect allows us to come closer to a causal investigation than post-hoc analyses from predictive modeling, which aim to identify factors that are correlated with ACD onset.

Cardiovascular diseases, including hypertension and cerebrovascular disease, are among the most commonly studied risk factors for ACD onset (Alzheimer's Association, 2024). Since adequate heart health is required to deliver oxygen to the brain, researchers hypothesize that comorbidities adversely affecting the heart also adversely affect the brain (Mergenthaler et al., 2013; Kuzma et al., 2018). Similarly, factors like smoking can have a negative effect on heart health (Wells et al., 1994), thus affecting

ACD progression in similar ways. Though we cannot test these hypotheses directly, we can check whether our findings are consistent across approaches from the treatment effect literature.

In this paper, we identified a cohort of patients with MCI and used ML to estimate the effects of recognized risk factors on risk of conversion to ACD from MCI at a 10-year horizon as a proof of concept (Hulse et al., 2005; Newman et al., 2005; Beydoun et al., 2008; Tamura et al., 2011; Thomson et al., 2017; Choi et al., 2018; Stefanidis et al., 2018; Dunietz et al., 2021). Based on our results, we suggest potential mechanisms for how these factors could contribute to ACD conversion. Going forward, these approaches can potentially be used by researchers in dementia to guide clinical research by investigating the effects of novel risk factors on dementia risk in observational data, leading to novel hypotheses on ways to intervene that can then be verified in future work.

2. Methods

2.1. Study Cohort

We included patients with MCI, as defined by the VINCI (VA Informatics and Computing Infrastructure) CIPHER (Centralized Interactive Phenomics Resource) criteria (Honerlaw et al., 2023), from the Veterans Affairs' (VA) Cerner EHR instance (Cerner Corporation, North Kansas City, MO) (VINCI, 2008) who had an encounter with any of the 172 VA facilities in the United States between January 1, 2009 and December 31,

2021. Patient timelines were aligned at the first diagnosis of MCI (i.e., MCI onset). We excluded patients with an MCI or ACD diagnosis before 50 years of age, patients who converted to ACD less than six months after their MCI diagnosis, and patients with less than one year of historical data prior to MCI diagnosis. ACD diagnoses were also defined as described by the VINCI CIPHER criteria (Honerlaw et al., 2023) (more detail in **Appendix A1**), which identify diagnoses based on meeting specific diagnostic billing codes in the EHR. To control for the effects of age of MCI diagnosis and healthcare utilization on the risk of ACD conversion, we matched patients across each time to conversion (e.g., conversion after one year) and time of censoring (e.g., censored after one year) by age of MCI diagnosis and number of BMI measurements within the five years leading up to MCI diagnosis (more detail in **Appendix A2**). Here, the number of BMI measurements acted as a surrogate for healthcare utilization since we assumed that patients generally have their BMI measured during routine clinical encounters. We controlled for these so that our predictions would not be dominated by these factors (e.g., the model mainly uses age to predict ACD risk). This study was carried out between August 2023-August 2024 and was approved by the Institutional Review Board of the [REDACTED] Veterans Affairs Health Care Center (protocol [REDACTED]).

2.2. Risk Factors Considered

We considered comorbidities that were diagnosed before MCI onset. Since cardiovascular comorbidities have been identified by the literature as risk factors, we considered comorbidities like hypertension and cerebrovascular disease (Alzheimer's

Association, 2024). Similarly, since smoking can adversely effect heart health (Wells et al., 1994), we also considered it as a risk factor. Outside of cardiovascular comorbidities, TBI (traumatic brain injury) (Vincent et al., 2014; Logue et al., 2023) is often studied in the context of ACD onset, where TBI has been shown to be associated with increased risk (Alzheimer's Association, 2024). Additionally, comorbidities affecting mental health (e.g., anxiety and depression) and psychological trauma (e.g., PTSD [post-traumatic stress disorder]) (Yaffe et al., 2009; Logue et al., 2023; Prieto et al., 2023) have been suggested to be associated with increased dementia risk (Byers et al., 2011; Gardner et al., 2014; Kwak et al., 2017; Desmarais et al., 2020).

We limited our focus to mid- to late-life modifiable risk factors that can be identified in the EHR to highlight comorbidities that could guide future research for designing risk-reducing interventions. Thus, we did not consider factors like genetics, demographics, education, and socioeconomic status (SES) since they are either 1) not intervenable or 2) not observable in the EHR. However, since they are potential confounders to our study, we considered them as features during model training to account for their effects on the risk of ACD onset. Only demographics were observable in the EHR, so we relied on downstream variables to capture the effects of features like SES as described below.

We assumed that these risk factors are related to ACD onset as shown in **Figure 1**, where the risk factors we considered are highlighted in blue. **Figure 1** (described more in **Appendix A3**) was constructed based on our literature search of ACD risk factors

and aims to explicitly state which risk factors we are considering and what we assume the potential confounders are. Since we considered comorbidities identified with diagnostic billing codes, they are likely confounded by variables like genetics and SES. These variables, in turn, affect related vital sign measurements and laboratory test results. Thus, we included them as confounders. For unobservable variables like education, genetics, and SES, we indirectly accounted for them by relying on downstream variables to capture their effects, such as healthcare utilization, ZIP codes, vital sign measurements, and laboratory test results. If we assume that these relationships hold, then we can apply the approaches outlined below.

2.3. Data preprocessing

We extracted 114 covariates relating to the comorbidities mentioned above as well as potential confounders such as demographics, medications, vital signs, laboratory tests, and healthcare utilization from up to five years befure MCI onset (more detail in **Table 1** and **Appendix A4**).

2.4. Comorbidity Effect Estimation

Given the comorbidities and patient covariates, we used ML techniques from the treatment effect estimation literature to estimate the effect of each comorbidity on ACD onset. We considered the probability of ACD conversion within 10 years of MCI onset as our outcome, and we trained ML models to predict ACD onset at the time of MCI onset given patient covariates (Wang et al., 2019). To estimate effects, we estimated

the difference in probability of ACD onset within 10 years in the presence and absence of each comorbidity, averaged over all patients (more detail in **Appendix A5**) (Rubin, 2005). In summary, the average effect can be interpreted as the percentage point change in the probability of ACD onset within 10 years of MCI onset resulting from the comorbidity.

2.4.1. Model training. For each comorbidity, we estimated the effects using common approaches from the treatment effect estimation literature, such as the X, R, and DR metalearners (see **Appendix A6**) (Funk et al., 2011; Xu et al., 2023a). To use these approaches, we make the following three assumptions, as is standard in the treatment effect estimation literature (VanderWeele, 2009; Xu et al., 2023a). The first is overlap: for a comorbidity of interest, the probability of any patient in the dataset having that comorbidity is non-zero. The second is unconfoundedness: the outcome (i.e., probability of ACD onset) is independent of whether the comorbidity is present, conditioned on patient covariates (i.e., all confounders are included in the covariate set). The third is consistency: a patient's observed outcome is the potential outcome, given their features and whether they have the comorbidity. In making these assumptions, we can use the metalearners to train survival analysis models using observational data and still recover the average effects. This is because the assumptions ensure that what the models learn from patients without the comorbidity, generalize to patients with the comorbidity and vice versa. We test for overlap as described in the preliminary analysis. Unconfoundedness holds if the assumed relationships outlined in **Figure 1** hold and

there are no additional confounders not shown in **Figure 1**. We cannot explicitly test for consistency, but it remains a reasonable assumption given our current understanding of the disease process.

2.4.2. Model evaluation. We first conducted a preliminary analysis to verify whether the overlap assumption holds and whether the approaches perform as expected in a controlled environment. Then, we conducted the main analysis, where we identified predictors of ACD onset using both a post-hoc analysis from predictive modeling and the metalearners described earlier. We compared the identified predictors from the post-hoc analysis to those from the metalearners.

The preliminary analysis is described in **Appendix A7**. Our main analysis consisted of two parts. In the first, we trained a standard model, reporting its discriminative performance using the time-dependent AUROC (area under the receiver operating characteristic curve) (Lambert et al., 2016). Potential predictors were identified with permutation importance (Breiman, 2001) on the standard model using the held-out test set. In the second part, we used the metalearners to measure the effects of the comorbidities on ACD onset. To evaluate the metalearners, we began by measuring the discriminative performance of all models trained for each metalearner using the time-dependent AUROC. Comorbidities whose 95% CI overlapped with 0.5 for at least one model were excluded from further analysis. For each comorbidity that remained, we measured the average effect of each comorbidity on ACD onset using each metalearner

by estimating the effect for all individuals in the test set and then taking the average. Here, we do not know what the ground truth effect of each comorbidity is so we can only evaluate whether the results among approaches are consistent (i.e., all of the approaches indicate that the comorbidity increases the risk of ACD onset). Inconsistencies among approaches would indicate that the signs of the predicted average effects are more likely to be false discoveries resulting from methodological differences among approaches (Xu et al., 2023b). For example, the X and R learners may be more sensitive to the quality of the estimated propensity scores. As such, for each comorbidity, we plotted the average effect for each approach and highlighted which comorbidities had consistent predictions, with error bars representing 95% CIs from 1,000 bootstrapped samples. Note that the features identified by consistent average effects and permutation importance are not guaranteed to be the same (see **Appendix A8** for more detail).

3. Results

3.1. Cohort characteristics. After applying our inclusion/exclusion criteria and matching, our cohort contained 19,797 MCI patients (**Figure 2**). 6,767 (34.18%) experienced ACD onset within 10 years, 1,320 (6.67%) did not experience ACD onset within 10 years, and the remaining were right censored (i.e., lost to follow-up less than 10 years after MCI onset without meeting the ACD criteria). The median age of MCI onset was 70 years [IQR (interquartile range) 65-78], 774 (3.91%) were female, 15,307 (77.32%) were White, and the median number of outpatient encounters prior to MCI

onset was 4 [IQR 2-6]. The most common comorbidity was hypertension, covering 57.72% of the cohort (11,427 patients). More details are in **Table 1**.

3.2. Preliminary analysis. We investigated the overlap assumption in **Appendix A9** by plotting the distributions of propensity scores for each comorbidity. The range of scores between positive and negative patients had a considerable amount of overlap for all comorbidities and the majority of scores were within the range [0.1, 0.9]. We investigated the approaches in a controled environment through a global null analysis in **Appendix A10**. The environment was controlled such that the average effects were known to be zero, and in the results, the estimated effects were close to zero.

3.3. Main analysis.

3.3.1. Part 1. Model performance, as measured by the time-dependent AUROC, was 0.61 (95% CI 0.59-0.64). The results from running permutation importance on a standard model using the held-out test set are shown in **Appendix A11**, where the only comorbidity identified as having a significant effect on performance was anxiety and related disorders, which resulted in a drop of 0.003 (95% CI 0.0003-0.007).

3.3.2. Part 2. The discriminative performance of the models trained is shown in **Appendix A12**. For each comorbidity, the performances of the models trained on the negative patients were all significantly better than random (time-dependent AUROC range=0.61-0.68). For the models trained on the positive patients for each comorbidity,

some were not significantly better than random (time-dependent AUROC range=0.41-0.61). In **Figure 3**, we show the estimated average effects for the comorbidities whose performance on all models trained for each approach were better than random. Dyslipidemia, hypertension, and diabetes were consistently identified as risk factors by all approaches.

4. Discussion

In our study, we identified hypertension, dyslipidemia, and diabetes as risk factors for ACD using approaches from the treatment effect estimation literature. While EHR data have been used by previous work to identify ACD risk factors, many focus on post-hoc analyses from predictive modeling instead of directly estimating the individual effect of each comorbidity (Jo et al., 2019; Tjandra et al., 2020; Tang et al., 2024; Irwin et al., 2024). For example, these studies identify potential risk factors by measuring feature importance post-hoc using approaches like permutation importance. While these approaches may indicate which features the model uses to make its predictions (i.e., which features may have a higher correlation with the outcome), they do not necessarily indicate which features could inform prevention. From our analysis using permutation importance, we found that the features identified differed from those identified by consistent average effects across metalearners. There are many reasons why this could occur. For example, the effects may not be large enough to significantly affect discriminative performance. This may be why hypertension, dyslipidemia, and diabetes were identified by the metalearners and not by permutation importance. It is also

possible that the effects among individuals within a comorbidity cancel each other out at the population level but can significantly change discriminative performance.

Identifying features that inform prevention requires identifying causal relationships. Verifying causal relationships requires an RCT. However, RCTs cannot be used to investigate the effect of comorbidities on the onset of ACD. We address this gap, in part, through retrospective analyses on observational data using ML techniques. With these techniques, we directly estimate how the presence of a comorbidity could change a patient's probability of ACD onset within some prediction horizon. While these analyses cannot replace RCTs, we have shown that they can identify risk factors that are consistent with the literature, and thus, have the potential to guide clinical research by suggesting avenues for future intervention. By observing similar trends across multiple approaches, we can strengthen our claim on whether the directions of our estimates (i.e., risk or protective) hold.

Approaches from the treatment effect estimation literature require a stricter set of assumptions, but these assumptions allow us to have greater confidence in the accuracy of the predictions under both potential outcomes. This is because these assumptions mean that the model can accurately learn the relationship between the comorbidity and outcome (due to consistency) while accounting for confounding (due to unconfoundedness), and that the learned relationship among patients with the comorbidity will generalize to patients without the comorbidity and vice versa (due to

overlap). As a result, we can estimate effects by taking the difference between the predictions of the two potential outcomes.

Our finding that dyslipidemia, hypertension, and diabetes are risk factors of ACD onset aligns with the literature (Biessels et al., 2006; Walker et al., 2017; Wee et al., 2023). Dyslipidemia increases the chance of cholesterol buildup in the arteries (Kopin et al., 2017), which could limit blood flow to the brain. Hypertension increases the chance of heart disease and stroke (Wajngarten et al., 2019), both of which can affect the heart's ability to supply oxygen to the brain. Insulin resistance from type 2 diabetes has been shown to lower insulin levels in the brain, which may contribute to cognitive decline (Gasparini et al., 2001). Further studies into the mechanism through which these comorbidities contribute to ACD progression could shed light on results from current work (Rosenberg et al., 2018) showing that medication or intervening on modifiable risk factors, such as diet and exercise, may slow the rate of cognitive decline. Notably, it has been proposed that drugs for diabetes and hypertension, two risk factors identified in our analyses, could potentially lower the risk of Alzheimer's disease, the most common form of dementia (Yasar et al., 2013; Michailidis et al., 2022).

In contrast to hypertension, dyslipidemia, and diabetes, some factors from the literature were not consistently identified by the metalearners. For example, smoking and cerebrovascular disease are associated with ACD onset in similar ways to hypertension, but the metalearners did not consistently indicate that they increased the probability of

ACD onset over 10 years. This may be because some of the comorbidities were prone to being affected by unobserved confounding. For example, comorbidities like smoking have been shown to be associated with SES (Hitchman et al., 2014). While we included ZIP codes in our feature set, they only serve as a proxy and may not fully capture SES. Despite hypertension, dyslipidemia, and diabetes also being associated with SES (Blok et al., 2022; Espirito et al., 2022), we hypothesize that including direct measures relating to vital signs and laboratory tests could more effectively capture the downstream effects of SES (e.g., blood pressure).

Our study has several limitations. First, we relied on EHR-based phenotyping tools to identify MCI, ACD, and the comorbidities we included, which may not always be accurate (Tjandra et al., 2020). Second, since the ground truth average effects are unknown, we could not quantitatively assess whether our predictions were correct. Third, we were unable to verify whether all of the assumptions required for the models held, such as unconfoundedness. Due to the complicated dynamics of ACD progression, it is likely that there are additional confounders that **Figure 1** did not include. Despite our inability to check the correctness of the main analysis, our results support well accepted, plausible hypotheses on what contributes to ACD onset. Fourth, our study only considered one comorbidity at a time and not how comorbidities act in combination. Additionally, the findings from our VA cohort should be validated in the general population. Past work has shown that veterans have a higher prevalence of mental health conditions, thus potentially putting them at higher risk for dementia

(Veitch et al., 2013). In addition, even among veterans, these conditions may affect male and female patients differently (Yaffe et al., 2019). In our cohort, the prevalence of PTSD and depression were 17% and 31% respectively, and patients were mostly male (96.09%). A previous study (Tjandra et al., 2022) showed that the performance of machine learning to predict AD onset using blood pressure trajectories trained using VA EHR data was similar when applied to EHR data from another institution even though the male/female demographic compositions are different, so generalizability to other demographics should be empirically established. Finally, when controlling for age of MCI onset and healthcare utilization, we used matching, which excluded many patients from our final cohort.

5. Conclusion

We demonstrated the potential of approaches for estimating treatment effects from observational data in the context of directly estimating the effects of comorbidities on risk of ACD onset. Results from analyses like ours can be used to inform future work in clinical research on identifying novel risk factors in settings where RCTs are infeasible.

References

Alzheimer's Association. 2024 Alzheimer's Disease Facts and Figures. Alzheimers Dement 2024;20(5).

Beydoun MA, Beydoun HA, Wang Y. Obesity and central obesity as risk factors for incident dementia and its subtypes: a systematic review and meta-analysis. Obesity reviews. 2008 May;9(3):204-18.

Biessels GJ, Staekenborg S, Brunner E, Brayne C, Scheltens P. Risk of dementia in diabetes mellitus: a systematic review. The Lancet Neurology. 2006 Jan 1;5(1):64-74.

Blok S, Haggenburg S, Collard D, Van Der Linden EL, Galenkamp H, Van Charante EP, Agyemang C, Van Den Born BJ. The association between socioeconomic status and prevalence, awareness, treatment and control of hypertension in different ethnic groups: the Healthy Life in an Urban Setting study. Journal of hypertension. 2022 May 1;40(5):897-907.

Breiman L. Random forests. Machine learning. 2001 Oct;45:5-32.

Byers AL, Yaffe K. Depression and risk of developing dementia. Nature Reviews Neurology. 2011 Jun;7(6):323-31.

Choi D, Choi S, Park SM. Effect of smoking cessation on the risk of dementia: a longitudinal study. Annals of clinical and translational neurology. 2018 Oct;5(10):1192-9.

Desmarais P, Weidman D, Wassef A, Bruneau MA, Friedland J, Bajsarowicz P, Thibodeau MP, Herrmann N, Nguyen QD. The interplay between post-traumatic stress disorder and dementia: a systematic review. The American Journal of Geriatric Psychiatry. 2020 Jan 1;28(1):48-60.

Devarakonda ST, Wu JY, Fung YR, Fiterau M. FLARe: Forecasting by Learning Anticipated Representations. In Machine Learning for Healthcare Conference 2019 Oct 28 (pp. 53-65). PMLR.

Dunietz GL, Chervin RD, Burke JF, Conceicao AS, Braley TJ. Obstructive sleep apnea treatment and dementia risk in older adults. Sleep. 2021 Sep 1;44(9):zsab076.

Espírito Santo LR, Faria TO, Silva CS, Xavier LA, Reis VC, Mota GA, Silveira MF, Mill JG, Baldo MP. Socioeconomic status and education level are associated with dyslipidemia in adults not taking lipid-lowering medication: a population-based study. International health. 2022 Jul;14(4):346-53.

Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. American journal of epidemiology. 2011 Apr 1;173(7):761-7.

Gardner RC, Burke JF, Nettiksimmons J, Kaup A, Barnes DE, Yaffe K. Dementia risk after traumatic brain injury vs nonbrain trauma: the role of age and severity. JAMA neurology. 2014 Dec 1;71(12):1490-7.

Gasparini L, Gouras GK, Wang R, Gross RS, Beal MF, Greengard P, Xu H. Stimulation of β-amyloid precursor protein trafficking by insulin reduces intraneuronal β-amyloid and requires mitogen-activated protein kinase signaling. Journal of Neuroscience. 2001 Apr 15;21(8):2561-70.

Hitchman SC, Fong GT, Zanna MP, Thrasher JF, Chung-Hall J, Siahpush M. Socioeconomic status and smokers' number of smoking friends: findings from the International Tobacco Control (ITC) Four Country Survey. Drug and alcohol dependence. 2014 Oct 1;143:158-66.

Honerlaw J, Ho YL, Fontin F, Gosian J, Maripuri M, Murray M, Sangar R, Galloway A, Zimolzak AJ, Whitbourne SB, Casas JP. Framework of the Centralized

Interactive Phenomics Resource (CIPHER) standard for electronic health data-based phenomics knowledgebase. Journal of the American Medical Informatics Association. 2023 May 1;30(5):958-64.

Hulse GK, Lautenschlager NT, Tait RJ, Almeida OP. Dementia associated with alcohol and other drug use. International Psychogeriatrics. 2005 Sep;17(s1):S109-27.

Irwin C, Tjandra D, Hu C, Aggarwal V, Lienau A, Giordani B, Wiens J, Migrino RQ. Predicting 5-year dementia conversion in veterans with mild cognitive impairment. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2024 Jan;16(1):e12572.

Jo T, Nho K, Saykin AJ. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. Frontiers in aging neuroscience. 2019 Aug 20;11:220.

Kopin L, Lowenstein CJ. Dyslipidemia. Annals of internal medicine. 2017 Dec 5;167(11):ITC81-96.

Kuźma E, Lourida I, Moore SF, Levine DA, Ukoumunne OC, Llewellyn DJ. Stroke and dementia risk: a systematic review and meta-analysis. Alzheimer's & Dementia. 2018 Nov 1;14(11):1416-26.

Kwak YT, Yang Y, Koo MS. Anxiety in dementia. Dementia and neurocognitive disorders. 2017 Jun;16(2):33.

Lambert J, Chevret S. Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent ROC curves. Statistical methods in medical research. 2016 Oct;25(5):2088-102.

Logue MW, Miller MW, Sherva R, Zhang R, Harrington KM, Fonda JR, Merritt VC, Panizzon MS, Hauger RL, Wolf EJ, Neale Z. Alzheimer's disease and related dementias among aging veterans: Examining gene-by-environment interactions withpost-traumatic stress disorder and traumatic brain injury. Alzheimer's & Dementia. 2023 Jun;19(6):2549-59.

Mergenthaler P, Lindauer U, Dienel GA, Meisel A. Sugar for the brain: the role of glucose in physiological and pathological brain function. Trends in neurosciences. 2013 Oct 1;36(10):587-97.

Michailidis M, Tata DA, Moraitou D, Kavvadas D, Karachrysafi S, Papamitsou T, Vareltzis P, Papaliagkas V. Antidiabetic drugs in the treatment of Alzheimer's disease. International journal of molecular sciences. 2022 Apr 22;23(9):4641.

Newman AB, Fitzpatrick AL, Lopez O, Jackson S, Lyketsos C, Jagust W, Ives D, DeKosky ST, Kuller LH. Dementia and Alzheimer's disease incidence in relationship to cardiovascular disease in the Cardiovascular Health Study cohort. Journal of the American Geriatrics Society. 2005 Jul;53(7):1101-7.

Prieto S, Nolan KE, Moody JN, Hayes SM, Hayes JP, Department of Defense Alzheimer's Disease Neuroimaging Initiative. Posttraumatic stress symptom severity predicts cognitive decline beyond the effect of Alzheimer's disease biomarkers in Veterans. Translational Psychiatry. 2023 Mar 29;13(1):102.

Rosenberg A, Ngandu T, Rusanen M, Antikainen R, Bäckman L, Havulinna S, Hänninen T, Laatikainen T, Lehtisalo J, Levälahti E, Lindström J. Multidomain lifestyle intervention benefits a large elderly population at risk for cognitive

decline and dementia regardless of baseline characteristics: The FINGER trial. Alzheimer's & Dementia. 2018 Mar;14(3):263-70.

Rubin DB. Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association. 2005 Mar 1;100(469):322-31.

Stefanidis KB, Askew CD, Greaves K, Summers MJ. The effect of non-stroke cardiovascular disease states on risk for cognitive decline and dementia: a systematic and meta-analytic review. Neuropsychology review. 2018 Mar;28:1-5.

Tamura MK, Yaffe K. Dementia and cognitive impairment in ESRD: diagnostic and therapeutic strategies. Kidney international. 2011 Jan 1;79(1):14-22.

Tang AS, Rankin KP, Cerono G, Miramontes S, Mills H, Roger J, Zeng B, Nelson C, Soman K, Woldemariam S, Li Y. Leveraging electronic health records and knowledge networks for Alzheimer's disease prediction and sex-specific biological insights. Nature Aging. 2024 Mar;4(3):379-95.

Thomson RS, Auduong P, Miller AT, Gurgel RK. Hearing loss as a risk factor for dementia: a systematic review. Laryngoscope investigative otolaryngology. 2017 Apr;2(2):69-79.

Tjandra D, Migrino RQ, Giordani B, Wiens J. Cohort discovery and risk stratification for Alzheimer's disease: an electronic health record-based approach. Alzheimer's & Dementia: Translational Research & Clinical Interventions. 2020;6(1):e12035.

Tjandra D, Migrino RQ, Giordani B, Wiens J. Use of blood pressure measurements extracted from the electronic health record in predicting Alzheimer's disease: a

retrospective cohort study at two medical centers. Alzheimers Dement. 2022;18(11):2368-2372.

VA Informatics and Computing Infrastructure (VINCI), VA HSR RES 13-457, U.S. Department of Veterans Affairs. (2008). Retrieved August 24 2023, from https://vaww.VINCI.med.va.gov

VanderWeele TJ. Concerning the consistency assumption in causal inference. Epidemiology. 2009 Nov 1;20(6):880-3.

Veitch PD, Friedl KE, Weiner MW. Military risk factors for cognitive decline, dementia and Alzheimer's disease. Current Alzheimer Research. 2013 Nov 1;10(9):907-30.

Vincent AS, Roebuck-Spencer TM, Cernich A. Cognitive changes and dementia risk after traumatic brain injury: implications for aging military personnel. Alzheimer's & Dementia. 2014 Jun;10:S174-87.

Wajngarten M, Silva GS. Hypertension and stroke: update on treatment. European Cardiology Review. 2019 Jul;14(2):111.

Walker KA, Power MC, Gottesman RF. Defining the relationship between hypertension, cognitive decline, and dementia: a review. Current hypertension reports. 2017 Mar;19:1-6.

Wang P, Li Y, Reddy CK. Machine learning for survival analysis: A survey. ACM Computing Surveys (CSUR). 2019 Feb 27;51(6):1-36.

Wee J, Sukudom S, Bhat S, Marklund M, Peiris NJ, Hoyos CM, Patel S, Naismith SL, Dwivedi G, Misra A. The relationship between midlife dyslipidemia and lifetime incidence of dementia: A systematic review and meta-analysis of cohort studies.

Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2023 Jan;15(1):e12395.

Wells AJ. Passive smoking as a cause of heart disease. Journal of the American College of Cardiology. 1994 Aug 1;24(2):546-54.

Xu Y, Ignatiadis N, Sverdrup E, Fleming S, Wager S, Shah N. Treatment heterogeneity with survival outcomes. InHandbook of Matching and Weighting Adjustments for Causal Inference 2023 Apr 11 (pp. 445-482). Chapman and Hall/CRC.

Xu Y, Bechler K, Callahan A, Shah N. Principled estimation and evaluation of treatment effect heterogeneity: A case study application to dabigatran for patients with atrial fibrillation. Journal of biomedical informatics. 2023 Jul 1;143:104420.

Yaffe K, Vittinghoff E, Lindquist K, Barnes DE, Covinsky KE, Neylan T, Kluse M, Marmar C. Post-traumatic stress disorder and risk of dementia among US veterans. Alzheimer's & Dementia. 2009;4(5):P104.

Yaffe K, Lwi SJ, Hoang TD, Xia F, Barnes DE, Maguen S, Peltz CB. Military-related risk factors in female veterans and risk of dementia. Neurology. 2019 Jan 15;92(3):e205-11.

Yasar S, Xia J, Yao W, Furberg CD, Xue QL, Mercado CI, Fitzpatrick AL, Fried LP, Kawas CH, Sink KM, Williamson JD. Antihypertensive drugs decrease risk of Alzheimer disease: Ginkgo Evaluation of Memory Study. Neurology. 2013 Sep 3;81(10):896-903.

Figure Legends

Figure 1: Assumed feature relations. We assume that the conditions we examine are related to each other and ACD onset as shown in patients with MCI. Blue boxes: comorbidities assumed to contribute to ACD onset. Green boxes with dashed borders: observed confounders. Yellow boxes with dotted borders: potential unobserved confounders whose downstream effects were assumed to be captured by other variables. Dark gray boxes: feature groups. Arrows pointing at a dark gray box point to all features in the box. Arrows coming from a dark gray box are such that there is an arrow coming from each feature in the box.

Figure 2: Inclusion/exclusion criteria. We begin with all patients with an encounter at any VHA facility between January 1 2009 and December 2021. Numbers in each box correspond to the number of patients included or excluded.

Figure 3: Average estimated effects. We show the average of the estimated effects of each condition on ACD onset. Error bars represent bootstrapped 95% confidence intervals.

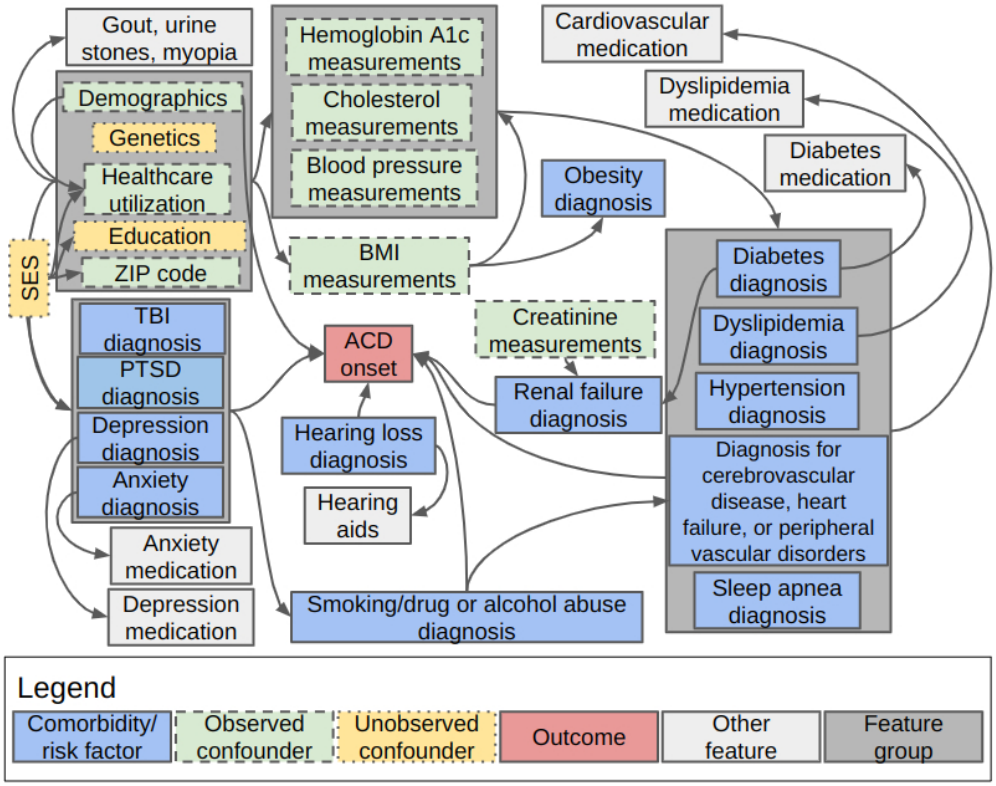


Figure 1: Assumed feature relations. We assume that the conditions we examine are related to each other and ACD onset as shown in patients with MCI. Blue boxes: comorbidities assumed to contribute to ACD onset. Green boxes with dashed borders: observed confounders. Yellow boxes with dotted borders: potential unobserved confounders whose downstream effects were assumed to be captured by other variables. Dark gray boxes: feature groups. Arrows pointing at a dark gray box point to all features in the box. Arrows coming from a dark gray box are such that there is an arrow coming from each feature in the box.

283x223mm (72 x 72 DPI)

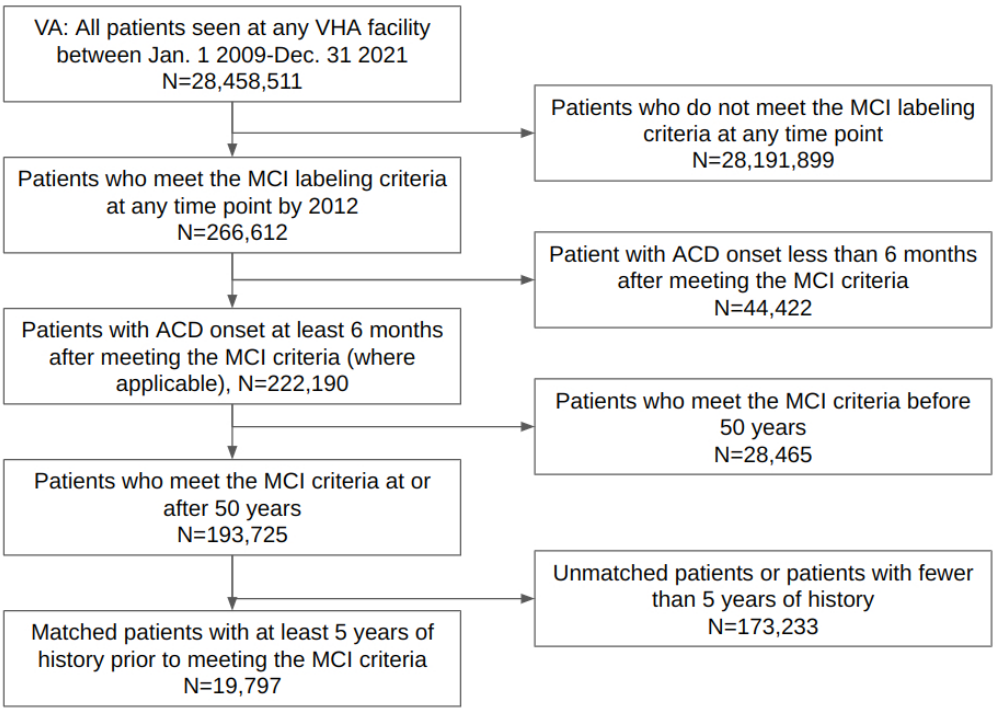


Figure 2: Inclusion/exclusion criteria. We begin with all patients with an encounter at any VHA facility between January 1 2009 and December 2021. Numbers in each box correspond to the number of patients included or excluded.

322x229mm (72 x 72 DPI)

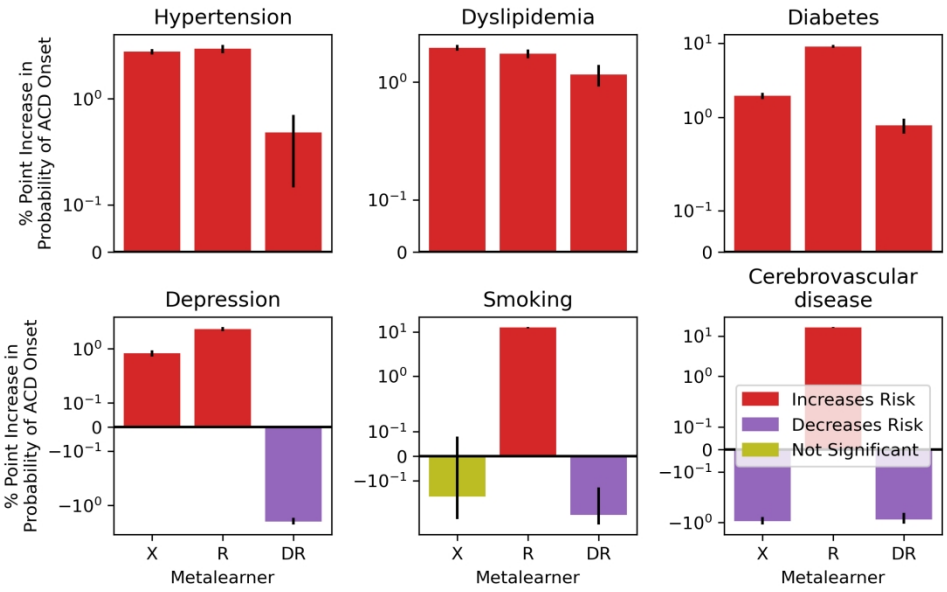


Figure 3: Average estimated effects. We show the average of the estimated effects of each condition on ACD onset. Error bars represent bootstrapped 95% confidence intervals.

516x322mm (118 x 118 DPI)

Table 1: Cohort characteristics and feature breakdown. We show the characteristics at alignment (i.e., MCI onset). Some patients did not have a race or ethnicity recorded. We report race and ethnicity categories as they are recorded. Abbreviations: N (number), IQR (interquartile range)

Characteristic: N(%) or Median[IQR]		N=19,797
Demographics (16 features)	Female	774(3.91%)
	Male	19,023(96.09%)
	White	15,307(77.32%)
	Black	2,843(14.36%)
	Declined to answer	612(3.09%)
	Two or more races	365(1.84%)
	Race unknown	186(0.94%)
	Hawaiian/Pacific Islander	171(0.86%)
	Asian	138(0.70%)
	American Indian/Alaskan	126(0.64%)
	White not of Hispanic Origin	27(0.14%)
	Non-Hispanic	17,976(90.80%)
	Hispanic	1,139(5.75%)
	Decline to answer	306(1.55%)
	Ethnicity unknown	250(1.26%)
	High risk ZIP code	2,224(11.23%)
Cardiovascul	Hypertension	11,427(57.72%)

ar comorbidities (6 features)	Cerebrovascular disease	2,833(14.31%)
	Peripheral vascular disorders	1,983(10.02%)
	Heart failure	918(4.64%)
	Myocardial infarction	761(3.84%)
	Pulmonary circulation disorders	349(1.76%)
Substance Abuse Comorbidities (3 features)	Smoking	3,725(18.82%)
	Alcohol abuse	2,099(10.60%)
	Drug abuse	1,314(6.64%)
Mental Health Comorbidities (4 features)	Depression	6,158(31.11%)
	Anxiety/related disorders	5,896(29.78%)
	Post-traumatic stress disorder	3,302(16.68%)
	Delirium	113(0.57%)
Other Comorbidities (14 features)	Dyslipidemia	11,421(57.69%)
	Diabetes	5,477(27.67%)
	Hearing loss	6,514(32.90%)
	Obesity	4,113(20.78%)
	Hearing aids	2,699(13.63%)
	Sleep apnea	1,819(9.19%)
	Renal failure	1,450(7.32%)
	Myopia	1,321(6.67%)

	Gout	1,100(5.56%)
	Weight loss	864(4.36%)
	Traumatic brain injury	770(3.89%)
	Liver disease	762(3.85%)
	Urine stones	656(3.31%)
	Coagulopathy	481(2.43%)
Age and Healthcare Utilization (15 features)	Age of MCI onset	70[65-78]
	Number of outpatient encounters	4[2-6]
	Number of inpatient encounters	0[0-0]
Most Recent Vital Signs & Laboratory Tests (51 features)	Systolic blood pressure	130mmHg[120-140]
	Diastolic blood pressure	75mmHg[68-81]
	Body mass index	27.97kg/m^2[24.91-31.55]
	Creatinine	1.02mg/dL[0.90-1.22]
	Hemoglobin A1c	5.9%[5.5-6.7]
	Aspartate transaminase	22U/L[18-27]
	Alanine transaminase	23U/L[17-31]
Medications (5 features)	Cardiovascular medications	15,313(77.35%)
	Depression medications	7,079(35.76%)
	Diabetes medications	5,139(25.96%)
	Anxiety medications	3,639(18.38%)
	Dyslipidemia medications	865(4.37%)

Appendix A1: CIPHER definitions for MCI and ACD

Diagnosis of MCI was based on the patient having ICD-9 or ICD-10 classification of MCI made on two or more separate clinic visits, an entry criterion based on MVP Cog Working Group validated to have 95% specificity based on rigorous chart review (Logue et al., 2022). ACD was defined using the ICD-9 or ICD-10 codes from the VA Centralized Interactive Phenomics Resource (CIPHER) Phenotype 0083 validated to have 82% specificity based on rigorous chart review (Logue et al., 2022). Although these criteria are potentially less accurate than consensus-based diagnoses based on clinician chart review, we used them since chart review would have been infeasible given the size of our cohort.

MCI: For a patients to be diagnosed with MCI, at least one of the following ICD (international classification of diseases) codes must have been given at least twice:

- ICD9: 331.83
- ICD10: G31.84

The date at which either code was first given was taken as the date of diagnosis.

ACD: For patients to be diagnosed with ACD, at least one of the following ICD codes must have been given at least twice:

- ICD9: 290.0, 290.10 – 290.13, 290.20, 290.21, 290.3, 290.40 -290.43, 294.20, 294.21, 294.8, 331.0, 331.1, 331.19, 331.11, 331.2, 331.5, 331.82, 332., 333.4
- ICD10: A81.00, F01.50, F03.90, F03.91, F10.96, G10., G20., G30.0, G30.1, G30.8, G30.9, G31.0, G31.09, G31.1, G31.01, G31.83, G91.2

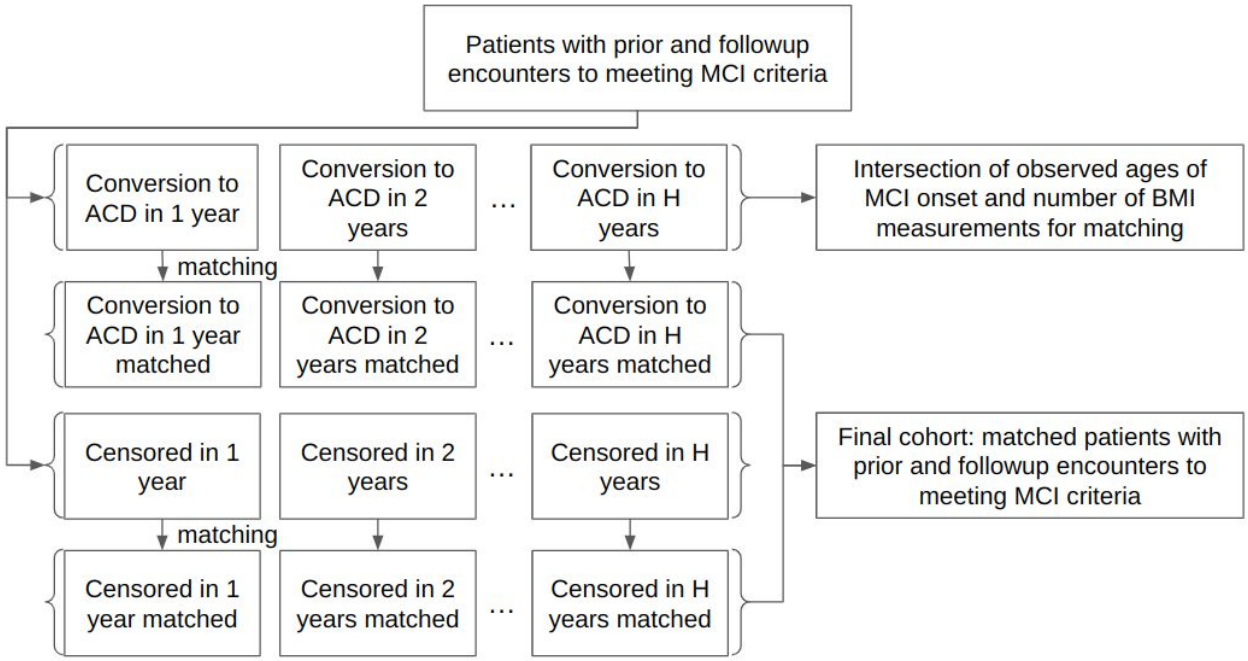
The date at which any of these codes was first given was taken as the date of diagnosis.

Appendix A2: Patient matching to control for age of MCI onset and healthcare utilization

To control for the effects of age and healthcare utilization on risk of ACD onset, we used the matching procedure in **Figure A1** (Lopez et al., 2017), matching on age of MCI onset and the number of BMI measurements. In summary, we begin with finding a group of patients with the same age of MCI onset and number of BMI measurements at the intersection of all conversion horizons. Then, at each conversion and censoring time point, we obtain the matched population by matching each patient in the intersection to a fixed number of patients within the conversion or censoring time point. Our matching constants are listed below, these were chosen to make the population as large as possible while keeping the distributions of age and number of BMI measurements between time points as similar as possible

- ACD conversion: [3, 6, 4, 3, 2, 2, 2, 2, 2, 1]
- Censoring: [5, 5, 5, 5, 5, 5, 5, 5, 5, 5]

Figure A1: Patient matching. We matched based on age of MCI onset and number of BMI measurements prior to MCI onset.



Appendix A3: Explanation of DAG

Here, the term “graph” refers to a set of nodes (i.e., the boxes) that are connected to each other through a set of edges (i.e., the arrows). The term “directed” refers to the arrows having a particular direction. Concretely, an arrow from “Box A” to “Box B” denotes the relationship that Box A causes Box B, although we note that Box A does not necessarily need to be the sole cause of Box B. For example, a patient’s hemoglobin A1c measurements can lead to a diagnosis of diabetes. The term “acyclic” means that there are no paths from a node to itself.

Appendix A4: Detailed feature breakdown

We show a more detailed breakdown of the features in **Table A1**. Summary statistics (i.e., average, most recent value, range of values, standard deviation, number of measurements) for features describing vital signs, laboratory tests, and healthcare utilization were calculated as in previous work (Tjandra et al., 2022). ZIP codes were formatted as an indicator variable, whose value was 1 if the patient’s ZIP code matched at least one in **Table A1** and 0 otherwise. We chose these ZIP codes based on past work (Dhana et al., 2023). We consulted the UDSv3 list (https://files.alz.washington.edu/documentation/uds3-tip-a4.pdf) in part for the medication list.

For categorical features (i.e., demographics, comorbidities, medications), we used a one-hot encoding. For numerical features (i.e., vital signs, laboratory tests, healthcare utilization), we binned the most recently recorded values into quintiles and used a one-hot encoding (Tang et al., 2020). Five years of historical data was chosen based on data availability.

Table A1: Feature definitions for vital signs, laboratory tests, healthcare utilization, diagnoses, medications, and ZIP codes

Feature Name/Category	Description
All vital signs, laboratory tests	Most recent value, indicator for missingness of value
Heart failure	ICD codes; '389.91', '402.11', '402.91', '404.11', '404.13', '404.91', '404.93', 'I09.0', 'I11.0', 'I13.0', 'I13.2', 'I42.5', 'I42.6', 'I42.7', 'I42.8', 'I42.9', 'P29.0', 'I43', 'I43.0', 'I43.1', 'I43.2', 'I43.8', 'I50', 'I50.1', 'I50.2', 'I50.20', 'I50.21', 'I50.22', 'I50.23', 'I50.3', 'I50.30', 'I50.31', 'I50.32', 'I50.33', 'I50.4', 'I50.40', 'I50.41', 'I50.42', 'I50.43', 'I50.8', 'I50.81', 'I50.82', 'I50.83', 'I50.84', 'I50.89', 'I50.9', 'I50.810', 'I50.811', 'I50.812', 'I50.813', 'I50.814', 'I25.5', 'I42.0', '428', '428.0', '428.1', '428.2', '428.3', '428.4', '428.9', '428.20', '428.21', '428.22', '428.23', '428.30', '428.31', '428.32', '428.33', '428.40', '428.41', '428.42', '428.43'
Diabetes	ICD codes: '250', '250.0', '250.00', '250.01', '250.02', '250.03', '250.1', '250.10', '250.11', '250.12', '250.13', '250.2', '250.20', '250.21', '250.22', '250.23', '250.3', '250.30', '250.31', '250.32', '250.33', '250.4', '250.40',

	'250.41', '250.42', '250.43', '250.5', '250.50', '250.51', '250.52', '250.53', '250.6', '250.60', '250.61', '250.62', '250.63', '250.7', '250.70', '250.71', '250.72', '250.73', '250.9', '250.90', '250.91', '250.92', '250.93', 'E10.0', 'E10.1', 'E10.9', 'E11.0', 'E11.1', 'E11.9', 'E12.0', 'E12.1', 'E12.9', 'E13.0', 'E13.1', 'E13.9', 'E14.0', 'E14.1', 'E14.9', 'E10.2', 'E10.3', 'E10.4', 'E10.5', 'E10.6', 'E10.7', 'E10.8', 'E11.2', 'E11.3', 'E11.4', 'E11.5', 'E11.6', 'E11.7', 'E11.8', 'E12.2', 'E12.3', 'E12.4', 'E12.5', 'E12.6', 'E12.7', 'E12.8', 'E13.2', 'E13.3', 'E13.4', 'E13.5', 'E13.6', 'E13.7', 'E13.8', 'E14.2', 'E14.3', 'E14.4', 'E14.5', 'E14.6', 'E14.7', 'E14.8'
Hypertension	ICD codes: '401', '642.0', 'I10', '402', '403', '404', '405', '642.1', '642.2', '642.7', '642.9', 'I11', 'I12', 'I13', 'I15'
Peripheral vascular disorders	ICD codes: '440', '441', '442', '443.1', '443.2', '443.3', '443.4', '443.5', '443.6', '443.7', '443.8', '443.9', '447.1', '557.1', '557.9', 'V43.4', 'I70', 'I71', 'I73.1', 'I73.8', 'I73.9', 'I77.1', 'I79.0', 'I79.2', 'K55.1', 'K55.8', 'K55.9', 'Z95.8', 'Z95.9'
Pulmonary circulation disorders	ICD codes: '416', '417.9', 'I26', 'I27', 'I28.0', 'I28.8', 'I28.9', '415.0', '415.1', '417.0', '417.8'
Coagulopathy	ICD codes: '286', '287.1', '287.3', '287.5', 'D65', 'D66', 'D77', 'D68', 'D69.1', 'D69.3', 'D69.4', 'D69.5', 'D69.6'
Obesity	ICD codes: '278.0', 'E66'
Weight loss	ICD codes: '260', '261', '262', '263', '783.2', 'E40', 'E41', 'E42', 'E43', 'E44', 'E45', 'E46', 'R63.4', 'R64', '799.4'
Drug abuse	ICD codes: '202.0', '292.82', '292.83', '292.84', '292.85', '292.86', '292.87', '292.88', '292.89', '292.9', '304', '305.2', '305.3', '305.4', '305.5', '305.6', '305.7', '305.8', '305.9', '648.3', 'F11', 'F12', 'F13', 'F14', 'F15', 'F16', 'F18', 'F19', 'Z71.5', 'Z72.2', '292', 'V65.42'
Hearing loss	ICD codes: '389', 'H90', 'H91'
Renal failure	ICD codes: '403.11', '403.91', '404.12', '404.92', 'V42.0', 'V45.1', 'V56.0', 'V56.8', 'I12.0', 'I13.1', 'N25.0', 'Z94.0', 'Z99.2', '585.1', '585.2', '585.3', '585.4', '585.5', '585.6', '585.9', '586', 'N18.1', 'N18.2', 'N18.3', 'N18.30', 'N18.31', 'N18.32', 'N18.4', 'N18.5', 'N18.6', 'N18.9', 'N19', 'Z49.0', 'Z49.01', 'Z49.02'
Liver disease	ICD codes: '070.32', '070.33', '070.54', '456.0', '456.1', '456.2', '571.0', '572.3', '572.8', 'V42.7', 'I86.4', 'I98.2', 'K71.1', 'K71.1', 'K76.0', 'Z94.4', '571.2', '571.3', '571.4', '571.40', '571.41', '571.42', '571.49', '571.5', '571.6', '571.8', '571.9', 'B18.0', 'B18.1', 'B18.2', 'B18.8', 'B18.9', 'I85.0', 'I85.00', 'I85.01', 'I85.1', 'I85.10', 'I85.11', 'K70.0', 'K70.1', 'K70.10', 'K70.11', 'K70.2', 'K70.3', 'K70.30', 'K70.31', 'K70.4', 'K70.40', 'K70.41', 'K70.9', 'K71.3', 'K71.4', 'K71.5', 'K71.50', 'K71.51', 'K72.0', 'K72.00', 'K72.01', 'K72.1', 'K72.10', 'K72.11', 'K72.9', 'K72.90', 'K72.91', 'K73.0', 'K73.1', 'K73.2', 'K73.8', 'K73.9', 'K74.0', 'K74.00', 'K74.01', 'K74.02',

	'K74.1', 'K74.2', 'K74.3', 'K74.4', 'K74.5', 'K74.6', 'K74.60', 'K74.69', 'K76.2', 'K76.3', 'K76.4', 'K76.5', 'K76.6', 'K76.7', 'K76.8', 'K76.81', 'K76.89', 'K76.9'
Depression	ICD codes: '300.4', '301.12', '309.0', '309.1', '311', 'F20.4', 'F34.1', 'F41.2', 'F43.2', 'F31.3', 'F31.4', 'F31.5', 'F32', 'F32.0', 'F32.1', 'F32.2', 'F32.3', 'F32.8', 'F32.9', 'F33', 'F33.0', 'F33.1', 'F33.2', 'F33.3', 'F33.4', 'F33.8', 'F33.9'
Alcohol abuse	ICD codes: '291.1', '291.2', '303.9', '305.0', 'V11.3', '291.5', '291.8', '291.81', '291.82', '291.89', '291.9', 'F10', 'E52', 'G62.1', 'I42.6', 'K29.2', 'K70.0', 'K70.3', 'K70.9', 'Z50.2', 'Z71.4', 'Z72.1', 'T51', 'T51.0', 'T51.1', 'T51.2', 'T51.3', 'T51.8', 'T51.9'
PTSD	ICD codes: 'F43.1', '309.81'
TBI	ICD codes: 'S02.0', 'S02.1', 'S06.2', 'S06.3', 'S06.8', 'S06.A', 'S06.0', 'S06.1', 'S09', '850', '851', '852', '853', '854', 'V15.52', 'Z87.8', 'S02', 'S04', 'S06', 'S07', 'S09', '800', '801', '803', '804', '907', 'S06.4', 'S06.5', 'S06.6', 'S06.9'
Cerebrovascular disease	ICD codes: 'I60', 'I61', 'I62', 'I63', 'I64', 'I65', 'I66', 'I67', 'I68', 'I69', '430', '431', '432', '433', '434', '435', '436', '437', '438'
Delirium	ICD codes: '293.0', 'T81.89', 'F05'
Hearing aids	ICD codes: 'Z97.4', 'V53.2'
Smoking	ICD codes: '305.1', 'V15.82', 'F17', 'Z87.891'
Dyslipidemia	ICD codes: '272', 'E78'
Myocardial infarction	ICD codes: ‘410’, ‘412’, ‘I21’, ‘I22’, ‘I25.2’
Sleep apnea	ICD codes: 'G47.3', '327.2'
Myopia	ICD codes: ‘H52.1’, ‘367.1’
Urine stones	ICD codes: ‘N20’, ‘592’
Gout	ICD codes: ‘M10’, ‘274
Cardiovascular medication	'lisinopril', 'ramipril', 'losartan', 'amiodarone', 'warfarin', 'aspirin', 'bisoprolol', 'amlodipine', 'simvastatin', 'digoxin', 'bendroflumethiazide', 'atorvastatin', 'fluvastatin', 'rosuvastatin', ‘dabigatran’, ‘rivaroxaban’, ‘apixaban’, ‘edoxaban’, ‘betrixaban’, ‘valsartan’, ‘nitroglycerin’, ‘nifedipine’, ‘niacin’, ‘metoprolol’, ‘lovastatin’, ‘hydrochlorothiazide’, ‘furosemide’, ‘enalapril’, ‘diltiazem’, ‘clopidogrel’, ‘carvedilol’, ‘benazepril’, ‘atenolol’
Dyslipidemia	'fenofibrate', 'gemfibrozil'

medication	
Diabetes medication	'insulin', 'metformin', 'acarbose', 'miglitol', 'bromocriptine', 'gliclazide', 'glipizide', 'glimepiride', 'tolbutamide'
Depression medication	'fluoxetine', 'paroxetine', 'fluvoxamine', 'citalopram', 'escitalopram', 'sertraline', 'desvenlafaxine', 'duloxetine', 'levomilnacipran', 'milnacipran', 'venlafaxine', 'nefazodone', 'trazodone', 'vilazodone', 'vortioxetine', 'esketamine', 'moclobemide', 'isocarboxazid', 'phenelzine', 'tranylcypromine', 'quetiapine', 'mirtazapine', 'bupropion'
Anxiety medication	'Clonazepam', 'alprazolam', 'lorazepam', 'bromazepam', 'oxazepam', 'chlordiazepoxide', 'diazepam', 'clorazepate'
ZIP codes	Any beginning with: 330, 331, 212, 104, 206, 207, 390, 391, 392, 701, 317, 291, 922, 798, 799, 885, 990, 901, 902, 903, 904, 905, 906, 907, 908, 910, 911, 912, 913, 914, 915, 916, 917, 918, 935, 600, 601, 602, 603, 604, 605, 606, 607, 608, 850, 851, 852, 853, 770, 772, 773, 774, 775, 919, 920, 921, 906, 907, 928, 927, 112, 111, 334

Appendix A5: Explanation of potential outomes framework

In our study, we used the potential outcomes framework. In the potential outcomes framework, for a given individual and comorbidity, we define the potential outcomes as the outcome in the presence and absence of the comorbidity. For example, with hypertension, the potential outcomes we aim to model are the probability of ACD onset within 10 years in the presence of hypertension and the probability of ACD onset within 10 years in the absence of hypertension. Using our predictions for the potential outcomes, we can then calculate the average effect as the signed difference of the outcome in the presence of the comorbidity and the outcome in the absence of the comorbidity, averaged over all individuals in the population. In this context, the average effect can be interpreted as the percentage point change in the probability of ACD onset within 10 years of MCI onset resulting from the comorbidity.

Appendix A6: Description of treatment effect estimation approaches

We summarize the approaches below.

• The X learner (Kunzel et al., 2019; Xu et al., 2023) first learns two models, one to predict the outcome in the presence of the comorbidity and one to predict the outcome in the absence of the comorbidity. It then learns to predict the comorbidity's effect by regressing on the difference between the observed outcome and potential outcome under the opposite comorbidity assignment. After this initial learning step, it further refines its predictions by incorporating propensity score adjustment, where the propensity score is the probability of whether the patient has the comorbidity.

• The R learner (Nie et al., 2021; Xu et al., 2023) is similar to the X learner in that it uses propensity score adjustment, but differs in that it uses the propensity score to estimate a weighted average of the potential outcomes, and then uses the difference between the observed outcome and the weighted average to estimate the effect.

• The DR (doubly robust) learner (Funk et al., 2011) builds on the X learner in that it further adjusts the models' predictions during training to improve accuracy so that it is more robust to inaccuracies during intermediate steps (e.g., estimating propensity scores).

For all approaches, the outcome prediction models were trained as random survival forests (Ishwaran et al., 2008). The propensity model to predict propensity scores was trained as a random forest (Breiman 2001). In addition, we also used the inverse probability of censorship to weight patients during some of the intermediate steps of training (Robins et al., 2000) and thus trained a random forest to predict the probability of censorship. For all models, we used the 6)implementations from scikit-survival (Polsterl 2020) and scikit-learn (Kramer 2016). We randomly split the cohort into an 80%/20% training/test set split and report all results on the held-out test set. All code will be made publicly available upon publication.

We focus on these approaches since the X and R metalearners have been shown to be more robust to class balance with respect to the comorbidities than other approaches like the S and T metalearners (Xu et al., 2023). The DR metalearner has been shown to be more robust to inaccuracies in intermediate modeling steps (Funk et al., 2011). Note that these approaches do not directly estimate the average effect, but rather the conditional average effect, which is the average effect conditioned on individual features. While there are analogs to these metalearners that estimate the average effect directly, using them would have required us to exclude patients with right censored outcomes, greatly reducing the sample size and potentially biasing the results. The metalearners we use, in contrast, allow us to train survival analysis models while including patients with right censored outcomes.

With 1) the absence of overlap, 2) the presence of unobserved confounding, or 3) a mismatch between the observed and potential outcomes, the models will no longer generalize, and conclusions on the estimated effects will no longer hold.

Appendix A7: Description of preliminary analyses

In the preliminary analysis, we first qualitatively evaluated whether the overlap assumption holds by plotting the propensity scores among patients for each comorbidity who do and do not have the comorbidity and comparing the distributions. We checked 1) whether the distributions overlapped with each other, or 2) whether the majority of values fell within the range [0.1, 0.9], which is sometimes used as a rule of thumb (Crump et al, 2006). After, we performed a global null analysis on all metalearners (Xu et al., 2023). In this experiment, we tested a setting where we synthetically created a random "comorbidity" such that its ground truth effect on ACD onset is known to be 0. Then, for each real comorbidity and metalearner, we estimated the effect of the synthetic random comorbidity within the patients who do and do not have the real

comorbidity. We then evaluated the estimated effects using the mean squared error (MSE), with error bars representing 95% confidence intervals (CIs) from 1,000 bootstrapped samples, and checked whether the MSEs were close to zero.

Appendix A8: Identifying features through average effects and permutation importance

Here, we explore whether an association-based feature identification approach like permutation importance would identify the same features as average effects would, even when the standard causal assumptions hold, in the context of our study. We first review how we use permutation importance in our study, where we measure whether permuting the values within a feature significantly worsens discriminative performance. Then, we describe some cases where features identified by permutation importance are not always the same as features identified by average effects. Note that, since we consider average effects in our study, we are interested in the effect of the features at a population level. Studying causal effects at the individual level is an active field of study but is beyond the scope of this paper.

In survival analysis, discriminative performance is measured through a set of comparisons, where each comparison considers an individual who experienced AD onset and an individual who did not experience ACD onset. In order for the comparison to be a correct ranking, the probability of ACD onset must be higher for the individual who experienced onset than the individual who did not experience onset at the time of onset for the individual who experienced onset. In permutation importance, permuting the values within a feature worsens discriminative performance when pairs that were correctly ranked become incorrectly ranked. Furthermore, when the standard causal assumptions hold, changing the value of the feature during permutation importance gives the value of the potential outcome under the opposite value. However, whether the values of the potential outcome under the opposite feature value significantly change discriminative performance and whether the average effect is non-zero are not always equivalent. Below, we describe why this is.

Consider two patients, A and B. Patient A has hypertension and experiences ACD onset 10 years after MCI onset, while patient B does not have hypertension and does not experience ACD onset 10 years after MCI onset. Therefore, the probability of ACD onset at 10 years should be higher in patient A than patient B if the patients are correctly ranked. In order for the ranking to become incorrect, the diagnosis of hypertension would have to change in at least one of the patients, since otherwise the ranking would not change. Suppose patient A loses their hypertension diagnosis during permutation importance. In order to make the ranking incorrect, the probability of ACD onset for patient A would have to become lower than that for patient B. This means that patient A's causal effect would indicate that hypertension is a risk factor, since not having hypertension would result in a lower probability of ACD onset. However, note that the causal effect for patient A must be sufficiently large in order to make the ranking incorrect. Therefore, if the causal effect is small, the ranking may still remain correct. Thus, at a population level, if the causal effects are not large enough to significantly worsen the rankings among pairs

of individuals, permutation importance might not always identify features with non-zero average effects.

Now consider two additional patients, C and D. Both have hypertension, but patient C experiences ACD onset 10 years after MCI onset while patient D does not. For patients C and D to be correctly ranked, the probability of ACD onset at 10 years should be higher for patient C. To make the ranking incorrect, at least one patient must lose their hypertension diagnosis during permutation importance. Suppose patient D loses the diagnosis. To make the ranking incorrect, the probability of ACD onset for patient D must increase so that it is greater than that of patient C. Therefore, in patient D, hypertension is a protective factor since losing the diagnosis increased their risk of ACD onset. However, for patient A above, hypertension was a risk factor. As a result, it is possible that heterogeneous effects of hypertension on different patients may significantly degrade the discriminative performance but result in an average effect that is not significantly different from 0. Thus, at a population, features identified by permutation importance might not always have a non-zero average effect.

We have demonstrated that, even when the causal assumptions are met and toggling the value of the feature of interest of the S learner can generate the true causal effect, features identified by the average effect do not necessarily align with those identified by permutation importance. For a more concrete example, see below. This is not a fault of either approach, but rather a highlight of the different ways important features can be identified. Note that the differences between approaches also depend on the evaluation metric as well as the outcome definition. Thus, the examples above are specific to our study and may not be universally true across all possible outcomes and all possible evaluation metrics.

Example: Consider a binary feature (x) with two potential outcomes and a horizon of 5 with time points 1, 2, 3, 4, 5. The values of the survival curve at each time point under the observed feature value is shown in bold for each patient. Assume that the model learns the correct potential survival curve for each patient

Patient with observed feature value	Event Status	Survival Curve Points	
		Potential survival curve if x=0	Potential survival curve if x=1
1 (x=1)	Event at time 1	0.50, 0.40, 0.30, 0.20, 0.20	**0.30, 0.25, 0.20, 0.15, 0.10**
2 (x=0)	Censored at time 2	**0.45, 0.30, 0.25, 0.15, 0.15**	0.35, 0.30, 0.25, 0.20, 0.15
3 (x=1)	Censored at time 3	0.40, 0.35, 0.25, 0.20, 0.15	**0.40, 0.25, 0.20, 0.20, 0.15**
4 (x=0)	Event at time 4	**0.65, 0.60, 0.55, 0.45, 0.20**	0.70, 0.60, 0.50, 0.35, 0.25

5 (x=1)	No event by time 5	0.70, 0.60, 0.55, 0.50, 0.40	**0.70, 0.65, 0.60, 0.50, 0.45**

- Average effect = -0.1 + 0 + 0 + 0.05 + 0.05 = 0 (measured by how much probability of survival by end of horizon changes when x changes from 0 to 1)

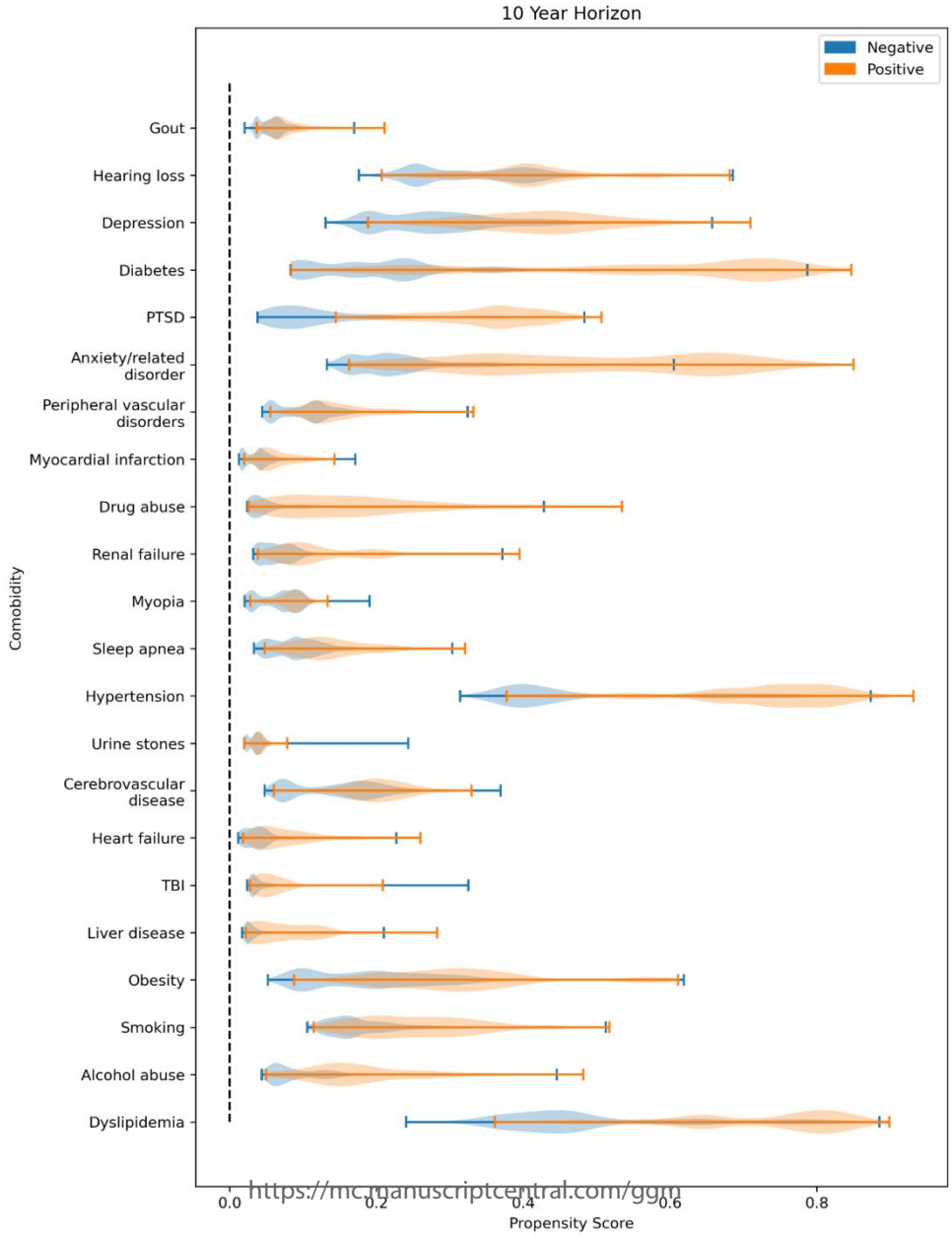
- Ranking: The comparable pairs are patients 1-2, 1-3, 1-4, 1-5, 4-5. All pairs are correctly ranked using the survival curves for the observed potential outcome. If the values of x are permuted such that the values become 0, 1, 0, 1, 1, for the five patients, respectively, then only 0.6 of the pairs are correctly ranked since 1-2 and 1-3 are no longer correct (demonstrated below). Note that for each pair, we compare the probabilities of the corresponding survival curves at the time of the earlier time-to-event For a comparison to correspond to a correct ranking, the quantity on the left must be less than the quantity on the right since this would indicate that the patient with the earlier event time has a lower probability of survival

Pair	Unpermuted Comparison	Permuted Comparison
1-2	0.30 < 0.45	0.50 > 0.35
1-3	0.30 < 0.40	0.50 > 0.40
1-4	0.30 < 0.65	0.50 < 0.45
1-5	0.30 < 0.75	0.50 < 0.70
4-5	0.45 < 0.50	0.35 < 0.50

Appendix A9: Checking for the overlap assumption.

We plot the propensity scores for each condition among all patients who were positive and negative for the comorbidity/condition and overlay them to visualize whether the ranges of values overlap in **Figure A2**. The overlap assumption requires that all patients have a propensity score strictly greater than 0 and strictly less than 1, although rule of thumb prefers values to be in the range [0.1, 0.9]. We notice that the range of propensity scores has considerable overlap between the positive and negative patients within each condition even if the distributions do not always overlap with each other perfectly. Although some conditions had more propensity scores outside the range [0.1, 0.9], there was more distributional overlap.

Figure A2: Testing the overlap assumption. We show the distribution of propensity scores as violin plots. To satisfy the overlap assumption, the probability of having the condition/comorbidity tested in the causal analysis (i.e., the propensity score) must be strictly greater than 0 and strictly less than 1 for all patients in the cohort.

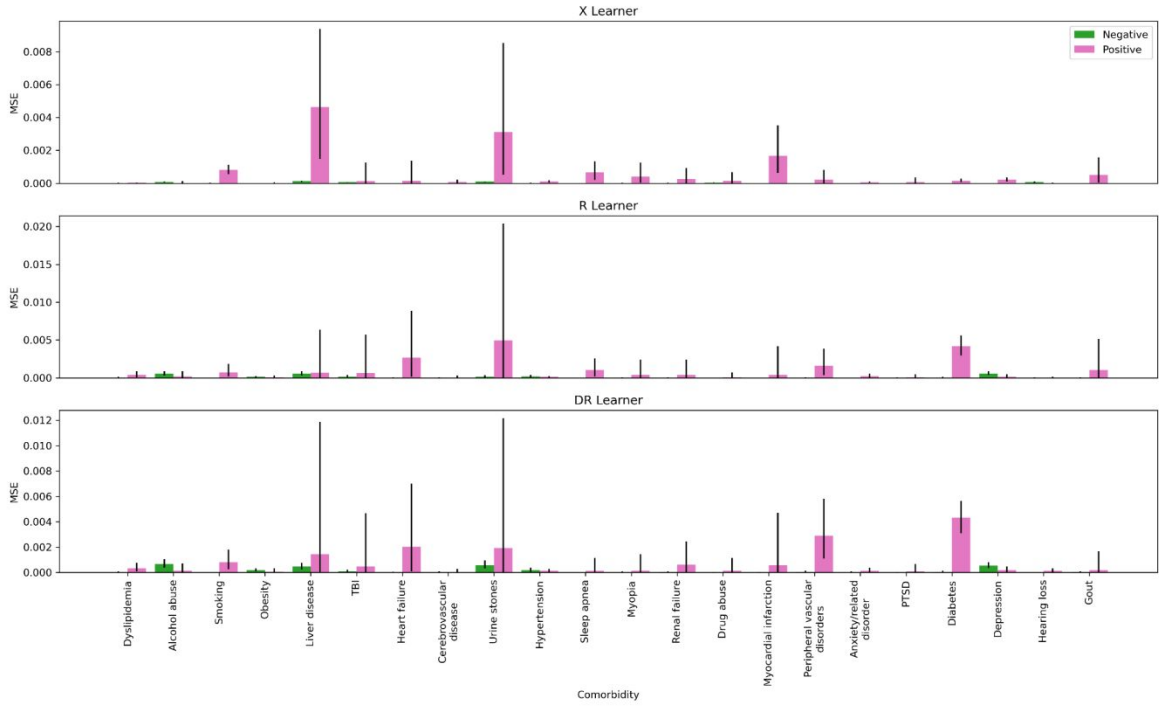


Appendix A10: Global null analysis

We include a control experiment in a semi-synthetic setting (**Figure A3**) where we synthetically create a random condition such that its ground truth effect on ACD onset is 0. For each condition we aim to test and for each approach, we conduct this experiment within the positive and negative individuals separately so that the synthetic condition is independent of the features and time of ACD onset and also that the time ACD onset is not confounded by whether the condition we aim to test is present.

The MSEs for the X, R, and DR learners were as follows: X<0.003, R<0.06, DR<0.03, where shorthand such as X<0.003 means that the MSE for all comorbidities among positive and negative patients for the comorbidity was less than 0.003 for the X learner. Smaller values are better, since the average effects in this experiment are known to be zero.

Figure A3: Global null analysis (control experiment). We show the mean squared error (MSE) for each condition and approach. Error bars represent bootstrapped 95% confidence intervals.



Appendix A11: Permutation Importance Results

Here, we show the results of running permutation importance on a standard survival prediction model (**Table A2**). Only anxiety and related disorders significantly decreased performance

Table A2: Results from permutation importance on a standard survival analysis model, implemented as a random forest. We show the median drop in the time-dependent AUROC with error bars representing 95% bootstrapped CIs from 100 permutations. Here, we show features whose median drop in performance was significantly greater than 0.001.

Feature	Drop in Time-Dependent AUROC (95% CI)
Anxiety/related disorders	0.003 (0.0003-0.007)

Appendix A12: Discriminative performance of base models for metalearners

We show the performance (**Table A3**) of a standard predictive model, the predictive model on positive patients from the X learner, the predictive model on negative patients from the X learner, the propensity model from the X learner, and the censorship model from the X learner. Note that we trained predictive models on the positive and negative patients as an intermediate step in learning the R, and DR learners. However, we only include results for the X learner to reduce redundancy. Similarly, we do not include results for the propensity and censorship models for the DR learner to reduce redundancy. In addition, since the predictive model (overall) and censoring model did not depend on the comorbidity, results were the same across comorbidity. We excluded comorbidities whose discriminative performance of the predictive model (positive) was not significantly better than random (95% CI included 0.5) and whose AUROC for the propensity model was above 0.9 from further analyses. Poorer performance on these models is likely due to small sample sizes, and a high propensity AUROC is indicative that the overlap assumption is less likely to hold.

Table A3: Performance for the models required for each metalearner. We show the performance of the causal models with respect to the following metrics: 'predictive' models: time-varying AUROC; 'propensity' and 'censorship' models: AUROC. Error bars represent bootstrapped 95% confidence intervals.

Comorbidity	Discriminative Performance (95% CI)				
	Predictive model (overall)	Predictive model (negative)	Predictive model (positive)	Propensity model	Censoring model
Dyslipidemia	0.61 (0.59-0.63)	0.68 (0.66-0.71)	0.54 (0.51-0.56)	0.83 (0.81-0.84)	0.74 (0.73-0.75)

Alcohol abuse		0.61 (0.59-0.64)	0.55 (0.48-0.60)	0.85 (0.83-0.87)	
Smoking		0.63 (0.60-0.65)	0.56 (0.51-0.60)	0.78 (0.76-0.80)	
Obesity		0.63 (0.61-0.65)	0.53 (0.48-0.58)	0.81 (0.79-0.82)	
Liver disease		0.62 (0.60-0.64)	0.55 (0.46-0.64)	0.79 (0.75-0.83)	
TBI		0.61 (0.59-0.63)	0.60 (0.50-0.69)	0.74 (0.70-0.78)	
Heart failure		0.61 (0.59-0.64)	0.41 (0.32-0.52)	0.78 (0.74-0.81)	
Cerebrovascular disease		0.63 (0.61-0.65)	0.58 (0.52-0.63)	0.75 (0.74-0.77)	
Urine stones		0.62 (0.60-0.64)	0.58 (0.46-0.71)	0.67 (0.63-0.71)	
Hypertension		0.66 (0.62-0.69)	0.55 (0.52-0.58)	0.84 (0.83-0.85)	
Sleep apnea		0.63 (0.61-0.65)	0.43 (0.37-0.49)	0.79 (0.76-0.81)	
Myopia		0.62 (0.59-0.64)	0.50 (0.41-0.59)	0.67 (0.64-0.70)	
Renal failure		0.62 (0.60-0.64)	0.49 (0.41-0.57)	0.82 (0.79-0.84)	
Drug abuse		0.62 (0.60-0.64)	0.56 (0.48-0.63)	0.88 (0.86-0.90)	
Peripheral vascular disorders		0.62 (0.60-0.65)	0.50 (0.43-0.57)	0.74 (0.72-0.76)	
Anxiety/related disorders		0.62 (0.59-0.64)	0.58 (0.54-0.62)	0.91 (0.90-0.92)	
PTSD		0.61 (0.59-0.64)	0.61 (0.56-0.65)	0.93 (0.92-0.94)	
Diabetes		0.62 (0.59-0.64)	0.57 (0.53-0.61)	0.83 (0.82-0.85)	

Depression		0.63 (0.61-0.66)	0.57 (0.53-0.61)	0.84 (0.83-0.86)	
Hearing loss		0.62 (0.59-0.64)	0.54 (0.50-0.57)	0.81 (0.80-0.82)	
Gout		0.61 (0.59-0.63)	0.47 (0.39-0.57)	0.69 (0.66-0.73)	
Myocardial infarction		0.62 (0.59-0.64)	0.55 (0.46-0.65)	0.78 (0.76-0.82)	

Appendix References

Breiman L. Random forests. Machine learning. 2001 Oct;45:5-32.

Crump RK, Hotz VJ, Imbens GW, Mitnik OA. Moving the Goalposts: Addressing Limited Overlap in the Estimation of Average Treatment Effects by Changing the Estimand. Technical Report 330, National Bureau of Economic Research, Cambridge, MA. 2006.

Dhana K, Beck T, Desai P, Wilson RS, Evans DA, Rajan KB. Prevalence of Alzheimer's disease dementia in the 50 US states and 3142 counties: A population estimate using the 2020 bridged-race postcensal from the National Center for Health Statistics. Alzheimer's & Dementia. 2023 Oct;19(10):4388-95.

Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. American journal of epidemiology. 2011 Apr 1;173(7):761-7.

Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random Survival Forests. The Annals of Applied Statistics. 2008 Sep 1:841-60.

Kramer O. Scikit-learn. Machine learning for evolution strategies. 2016:45-53.

Künzel SR, Sekhon JS, Bickel PJ, Yu B. Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the national academy of sciences. 2019 5;116(10):4156-65.

Logue MW, Miller MW, Sherva R, et al. Alzheimer's disease and related dementias among aging veterans: examining gene-by-environment interactions with post-traumatic stress disorder and traumatic brain injury. Alzheimers Dement. 2022.

Lopez MJ, Gutman R. Estimation of causal effects with multiple treatments: a review and new ideas. Statistical Science. 2017 Aug 1:432-54.

Nie X, Wager S. Quasi-oracle estimation of heterogeneous treatment effects. Biometrika. 2021 1;108(2):299-319.

Pölsterl S. scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn. The Journal of Machine Learning Research. 2020 Jan 1;21(1):8747-52.

Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000 Sep;56(3):779-88.

Tang S, Davarmanesh P, Song Y, Koutra D, Sjoding MW, Wiens J. Democratizing EHR analyses with FIDDLE: a flexible data-driven preprocessing pipeline for structured

clinical data. Journal of the American Medical Informatics Association. 2020 Dec;27(12):1921-34.

Tjandra D, Migrino RQ, Giordani B, Wiens J. Use of blood pressure measurements extracted from the electronic health record in predicting Alzheimer's disease: A retrospective cohort study at two medical centers. Alzheimer's & Dementia. 2022 Nov;18(11):2368-72.

Xu Y, Ignatiadis N, Sverdrup E, Fleming S, Wager S, Shah N. Treatment heterogeneity with survival outcomes. InHandbook of Matching and Weighting Adjustments for Causal Inference 2023 Apr 11 (pp. 445-482). Chapman and Hall/CRC.