

Novel approaches in causal inference and dimensionality reduction with an application to
Alzheimer's disease

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy (PhD)

Dimitris Karletsos

Health Policy and Management (HPAM)

Tulane University School of Public Health and Tropical Medicine

October 2023

Committee Chair

Dr. Charles Stoecker, Health Policy and Management

Committee Members

Dr. Charles Stoecker, Health Policy and Management

APPROVED , date 1/5/2024

Dr. Lizheng Shi, Health Policy and Management

APPROVED , date 12-06-2023

Dr. Yu-Ping Wang, Biomedical Engineering

APPROVED , date 12-13-2023

This work is a collection of three manuscripts that use novel methodological approaches in causal inference and dimensionality reduction on different classes of Alzheimer's disease data. The overarching aim of this dissertation is to investigate complementary perspectives that may improve the health of the population affected by Alzheimer's disease and related dementias, which is deemed to grow in number and public health relevance globally in the near future.

The first manuscript investigates the opportunity for drug repurposing of acetylcholinesterase inhibitors, a medication normally prescribed to Alzheimer's patients, in the treatment of certain cardiovascular disease. The second manuscript evaluates the effect of dual x-ray absorptiometry bone density scans on the likelihood of subsequent hip fracture in the Alzheimer's disease population, where osteoporosis is a highly incident comorbidity. The third manuscript presents an algorithm for dimensionality reduction and an application of set theory on pairwise classification problems to identify significant predictors of Alzheimer's disease progression.

The novelty of this work lies in the use of random treatment date generation, in combination with random sampling with replacement, to estimate the average treatment effect on the treated in the first two manuscripts. In the third manuscript, a novel algorithm is presented, which improves performance over the sparse-group lasso by adding a forward selection step on an external validation set of features.

Taken together, this work aims to contribute to the methodological advancement of statistical approaches for coefficient estimation in the context of causal inference, as well as to the empirical identification of elements that can be translated into actionable policies, on one hand, or utilized in clinical settings as part of diagnostic biomarkers, on the other.

I would like to extend my heartfelt gratitude to my advisor, Dr. Charles Stoecker, for his unwavering guidance, invaluable insights, and relentless support throughout the journey of this research. I wish to acknowledge the honest friendship that I found in my PhD program colleagues and the inspiration that I received from the Health Policy and Management department faculty at the Tulane School of Public Health and Tropical Medicine and all other departments where I have spent time attending seminars and lectures, including at the Tulane School of Medicine and the Tulane School of Engineering. Additionally, I am deeply appreciative of the contributions and encouragement from my colleagues and family, without whom this endeavor would not have been possible.

PAPER 1

Drug repurposing in Alzheimer's disease: the case of acetylcholinesterase inhibitors use for acute coronary syndrome prevention

Abstract

Alzheimer's disease (AD) is the most common cause of dementia, a neurodegenerative disease that affects cognitive functions, including memory, speech and mobility. In Louisiana, heart disease, including acute coronary syndrome (ACS), is often an incident comorbidity of AD patients. Exploring the opportunity for drug repurposing, we used Louisiana Medicare claims data to estimate the average treatment effect on the treated (ATT) of acetylcholinesterase inhibitors (ChEIs) prescription on the likelihood of subsequent ACS. We found that in our sample ChEIs decreased the likelihood of filing an ACS claim during the observation period by 0.56 percentage points (-.0056; CI: -.0091 to -.0021), or -6.18% over baseline value. Results are statistically significant and robust across all models, stratified over outpatient and inpatient claim types, and successfully passed placebo tests. This evidence should be considered in clinical decisions regarding ChEIs utilization and further assessed in patients with and without Alzheimer's disease and related dementia diagnoses.

Keywords: drug repurposing, acetylcholinesterase inhibitors, Alzheimer's disease, acute coronary syndrome

Introduction

Alzheimer's disease (AD) is the most common cause of dementia, a neurodegenerative disease that affects cognitive functions, including memory, speech and mobility (Kumar, 2022). It is estimated that in 2022 as many as 6.5 million Americans were living with Alzheimer's disease (CDC). Lower natality rates, extended lifespan, and the resulting ageing population will foster this trend in the future. Often, AD patients have comorbidities that may worsen their clinical condition further. Wang (2018a) found that multimorbidity is common in AD, and may be partially due to common risk factors, as depression, cardiovascular disease, osteoporosis, and diabetes mellitus. Also, AD and heart disease have common risk factors: Stampfer (2006) showed that cardiovascular risk factors are also risk factors for dementia, including hypertension, high LDL cholesterol, low HDL cholesterol and especially diabetes.

Although there still is no cure for AD, acetylcholinesterase inhibitors (ChEIs) are a drug commonly prescribed to Alzheimer's disease and other dementia patients to curb the symptoms of the cognitive decline. Acetylcholinesterase inhibitors inhibit acetylcholinesterase and hence increase cholinergic transmission. Acetylcholinesterase inhibitors function to decrease the breakdown of acetylcholine and increase both the levels and duration of actions of acetylcholine found in the central and peripheral nervous system (Colović, 2013). Three drugs belonging to this class have been approved by FDA to treat AD so far: donepezil (Aricept), rivastigmine (Exelon), and galantamine (Razadyne). They are also used off-label for other causes of dementia such as Lewy body and vascular dementia.

Drug repurposing describes the identification of new 'off-label' medications that can be used for a disease or medical condition that they are not formally approved to treat. In this regard, previous studies have associated ChEIs use with reduced risk of heart disease. Using claims data from the National Health Insurance Database in Taiwan, Wu and colleagues (2015) found use of ChEIs to be associated with a lower incidence of acute coronary syndrome (ACS) compared to the matched reference cohort, with an adjusted hazard ratio for ACS in AD patients treated with ChEIs equal to 0.836 (95% confidence interval, 0.750 to 0.933; $p < 0.001$). Isik's meta-analysis and systematic review (2018) found that ChEIs therapy may be associated with negative chronotropic and hypertensive effects but also with lower risk of cardiovascular events. In a sample of 6,070 patients with AD, Hsiao et al. (2021) found that ChEIs users had a significantly lower risk of cardiovascular events than nonusers (hazard

ratio: 0.57; 95% CI: 0.51 to 0.62). In a propensity score matching (PSM) study using data from the National Health Insurance Research Database of Taiwan, Hsieh et al. (2022) found that ChEIs users exhibited a significantly lower incidence of new-onset heart failure compared with untreated (HR 0.48; 95% CI 0.34–0.68, $p < 0.001$).

The objective of this study is to evaluate the association between ChEIs prescription and the likelihood of subsequent ACS in a sample of Louisiana Medicare beneficiaries. This is the first study of its kind to be conducted on the Louisiana Medicare population. The results of this study are relevant given the high incidence of both dementia and heart disease in the elderly population of Louisiana. The findings of this study add an important contribution to the literature investigating the effect of ChEIs on ACS risk.

Methods

We used propensity score methods for bias reduction to compare treatment group to a non-randomized control group, as described in Dehejia (2002) and D'Agostino (1998). We then used a weighted linear regression to estimate the impact of ChEIs on risk of ACS. Weighting by odds was used to estimate the average treatment effect on the treated (ATT). With weighting by the odds, treated individuals receive a weight equal to 1 and comparison individuals receive a weight equal to their propensity score (ρ_i), converted to the odds scale ($= \rho_i / (1 - \rho_i)$) (Hirano et al., 2003). This weighting effectively up-weights comparison individuals whose measured covariate values (propensity scores) best match those of the treated individuals and down-weights comparison individuals whose measured covariate values are dissimilar from treated individuals. One way of thinking about weighting by the odds is that the comparison individuals are first weighted to the entire population, using $1 / (1 - \rho_i)$, and are then weighted to the treatment group, using ρ_i . (Harder et al, 2010). Propensity scores were estimated using the `psmatch2` command in Stata 17. This command employs a logistic regression model, where the treatment assignment serves as the dependent variable. The relevant covariates, determined based on the examined literature, were included as independent variables. The estimated propensity scores represent the probability of receiving the treatment conditional on the observed covariates.

To assess the comparability of treated and control samples, we computed the standardized difference between treated and control groups across variables. Balance diagnostics were conducted using the `psbalance` command in Stata. The `psbalance` command examines the

distribution of covariates between the treated and control groups after weighting and provides statistical tests and graphical displays to assess balance. The balance diagnostics ensure that the weighted groups are comparable and reduce the potential for bias due to confounding. We assumed that an acceptable standardized difference would be lower than 10% (Austin, 2008). The functional form of the psmatch2 model was modified by including interaction terms and higher order polynomials until satisfactory standardized differences were obtained for all variables.

Finally, the ATT was estimated using the weighted sample. The differences in the outcome variable between the treated and control groups were computed and tested for statistical significance using a t-test. Sensitivity analyses were also conducted to examine the robustness of the results and assess the potential impact of unobserved confounders and individual-level fixed effects were included in the model to control for time-invariant unobservables.

To allow for an estimation of the likelihood of fracture occurring after the bone density test in both treated and untreated groups, the following algorithm was implemented. First, for each year in the observation period, the distribution of treatment dates in the treated sample was tabulated. Then, in the untreated sample, a numerical variable was created using a random number generator and the sample was sorted. The tabulated treatment dates distribution from the treated group was assigned to the sorted untreated sample. The process was repeated iteratively for all months included in the observation period. By doing so, treatment dates based on the distribution of treatment dates from the treated sample were assigned to the untreated group. Random assignment of treatment dates in the untreated was therefore constrained to the treatment dates distribution in the treated population. Finally, to ensure standard errors were adjusted for the random assignment of treatment dates, we implemented the procedure inside a bootstrap algorithm with replacement with 10,000 iterations over the entire analytical sample.

This method allowed for the estimation of the ATT, ensuring reliable causal inferences about the treatment's impact on the outcome variable.

To confirm the internal validity of our findings, we conducted a placebo test by running the same regressions on outcomes that were hypothesized to be unaffected by ChEIs use. To

perform this test, we sampled from the universe of Louisiana Medicare beneficiaries with AD and analysed all inpatient and outpatient claims where ICD9/10 codes were unrelated to heart disease, cardiovascular disease, or Alzheimer's disease. Finally, we estimated the probability of filing an inpatient or outpatient claim following ChEIs treatment.

Data

Our data is drawn from the Louisiana Medicare beneficiaries population. We obtained Medicare claims data from the Centers for Medicare & Medicaid Services (CMM) and we restricted our sample to beneficiaries who were continuously enrolled from 2014 to 2018, stratifying over inpatient and outpatient data. In total, 45,396 patients with AD were selected, based on the ICD-9/10 codes in Albrecht (2019). Then, we restricted our sample to those beneficiaries that filed at least one AD-related claim every year, from 2014 to 2018. A subset of 26,535 continuously enrolled beneficiaries was kept. Finally, patients were included in the final analytical sample if their estimated probability of receiving treatment were falling within the area of common support in the propensity score matching algorithm. The average age of the beneficiaries included in the study was 78.32 years in the treated group and 78.74 years in the untreated group. Females were 69.27% in the treated group and 61.65% in the untreated group. Black or African American beneficiaries were 26.02% in the treated group and 23.30% in the untreated group. As for comorbidities, in the treated group 7.93% ever had an acute myocardial infarction diagnosis at the time of ChEIs treatment, 82.58% ever had anemia, 17.19% asthma, 23.82% atrial fibrillation, 6.79% breast cancer, 4.06% colorectal cancer, 0.96% endometrial cancer, 1.58% lung cancer, 5.38% prostate cancer, 81.10% cataract, 50.78% congestive heart failure, 58.22% chronic kidney disease, 44.21% chronic obstructive pulmonary disease (COPD), 74.36% depression, 52.24% diabetes, 27.41% glaucoma, 13.07% hip or femur fracture, 88.81% hyperlipidemia, 20.25% hyperplasia, 96.14% hypertension, 41.93% hypothyroidism, 71.05% ischemic heart disease, 36.69% osteoporosis, 80.43% rheumatoid arthritis and 39.56% transient ischemic attack. In the treated group, at the time of treatment, 31.54% ever had a cholesterol medication prescription, 12.44% a diabetes medication prescription and 27.30% an anti-hypertension medication prescription. Finally, in the treated group, 4.53% of the overall sample had a pre-ChEIs ACS diagnosis.

Results

Table 1 shows the covariate means stratified by treatment status and weighting.

Standardized differences are included. In the crude, unweighted, sample, the only covariate with a standardized difference exceeding the 10% threshold was sex, as female accounted for 69.27% of the sample in the treated population and 61.65% in the untreated, with a standardized difference of 11.37%. However, after weighting by odds, the two groups became comparable. In the adjusted weighted sample, all demographics and comorbidities covariates had a standardized difference below 1%. Standardized difference in pre-ChEIs ACS was 1.99% in the overall sample, 1.29% in the inpatients sample and 2.26% in the outpatient sample.

Table 1. Bootstrap weighted means with standardized differences, by treatment status and weighting. ^a

Model	Crude			Adjusted		
	Untreated (N=10,902)	Treated (N=15,633)	Standardized difference	Untreated (N=10,901)	Treated (N=15,631)	Standardized difference
	Mean	Mean	Δ	Mean	Mean	Δ
<i>Demographics:</i>						
Age	78.32	78.74	.0322	78.66	78.62	-.0022
Age above 80	.4551	.4379	-.0245	.4370	.4379	.0012
Female	.6165	.6927	.1137	.6919	.6927	.0013
Black or African American	.2330	.2602	.0446	.2592	.2602	.0016
<i>Comorbidities:</i>						
Acute myocardial infarction	.0874	.0793	-.0207	.0794	.0793	-.0002
Anemia	.8212	.8258	.0085	.8244	.8258	.0027
Asthma	.1755	.1719	-.0066	.1718	.1719	.0003
Atrial fibrillation	.2650	.2382	-.0438	.2377	.2382	.0007
Cancer, breast	.0656	.0679	.0065	.0676	.0679	.0006
Cancer, colorectal	.0442	.0406	-.0126	.0404	.0406	.0007
Cancer, endometrial	.0084	.0096	.0086	.0096	.0096	.0000
Cancer, lung	.0233	.0158	-.0383	.0159	.0158	-.0007
Cancer, prostate	.0701	.0538	-.0478	.0541	.0538	-.0009
Cataract	.7956	.8110	.0274	.8107	.8110	.0007
Congestive heart failure	.5126	.5078	-.0068	.5063	.5078	.0020
Chronic kidney disease	.6002	.5822	-.0258	.5819	.5822	.0004
COPD	.4509	.4421	-.0125	.4418	.4421	.0005
Depression	.6847	.7436	.0923	.7438	.7436	-.0003
Diabetes	.5153	.5224	.0101	.5222	.5224	.0003
Glaucoma	.2700	.2741	.0066	.2746	.2741	-.0007
Hip fracture	.1387	.1307	-.0164	.1300	.1307	.0016
Hyperlipidemia	.8727	.8881	.0335	.8879	.8881	.0003
Hyperplasia	.2496	.2025	-.0797	.2029	.2025	-.0006

Hypertension	.9570	.9614	.0157	.9609	.9614	.0016
Hypothyroidism	.4108	.4193	.0123	.4201	.4193	-.0011
Ischemic heart disease	.7170	.7105	-.0102	.7107	.7105	-.0003
Osteoporosis	.3477	.3669	.0282	.3659	.3669	.0014
Rheumatoid arthritis	.7968	.8043	.0132	.8035	.8043	.0014
Transient ischemic attack	.4059	.3956	-.0149	.3961	.3956	-.0008
<i>Concomitant medications:</i>						
Cholesterol meds	.3132	.3154	.0032	.3149	.3154	.0007
Diabetes meds	.1303	.1244	-.0124	.1237	.1244	.0015
Hypertension meds	.2743	.2730	-.0020	.2738	.2730	-.0013
<i>Excluded covariates:</i>						
Pre-treatment ACS						
All	.0418	.0453	.0124	.0397	.0453	.0199
Inpatient	.0164	.0176	.0063	.0153	.0176	.0129
Outpatient	.0400	.0441	.0150	.0379	.0441	.0226

^a Variable means in unweighted (crude) and weighted (adjusted) models. Standardized differences were calculated as the difference between the means divided by the standard deviation of treated and untreated groups. Comorbidities refer to events occurring before the treatment date. Excluded covariates were not used in the calculation of the propensity scores.

Table 2 shows the probability of ACS diagnosis in a cohort of Medicare beneficiaries with a diagnosis of Alzheimer's disease, continuously enrolled from 2014 to 2018. We found that in our sample ChEIs therapy was significantly associated with a reduced risk of ACS.

ChEIs therapy decreased the likelihood of filing an ACS claim during the observation period by 0.56 percentage points (-.0056; CI: -.0091 to -.0021). Restricting the sample to inpatient claims only, ChEIs therapy was associated with a 0.50 percentage points decrement in the probability of ACS diagnosis (-.0028; CI: -.0046 to -.0010). In outpatient claims, ChEIs treatment reduced the likelihood of ACS diagnoses by 0.69 percentage points (-.0022; CI: -.0036 to -.0008).

At baseline, the untreated group had a mean ACS diagnosis probability of 0.0908 (SD = .2840). This was equal to 0.0330 (SD = .1775) for inpatients and 0.0886 (SD = .2809) in outpatient claims. The largest percent decrease over baseline value was found in inpatient claims. The percent decrease in inpatient claims was 8.57% (-8.57; CI: -14.71 to -3.26) while the percent decrease in outpatient claims was 2.61% (-2.61; CI: -5.32 to -0.49). Finally, in the overall sample, the percent decrease in the probability of ACS diagnosis associated with ChEIs treatment was 6.18% (-6.18; CI: -11.60 to -1.97). Results are statistically significant and robust across three models, stratified over outpatient and inpatient claim types.

Table 2. Probabilities of acute coronary syndrome diagnosis in the observation period, by claim type. ^a

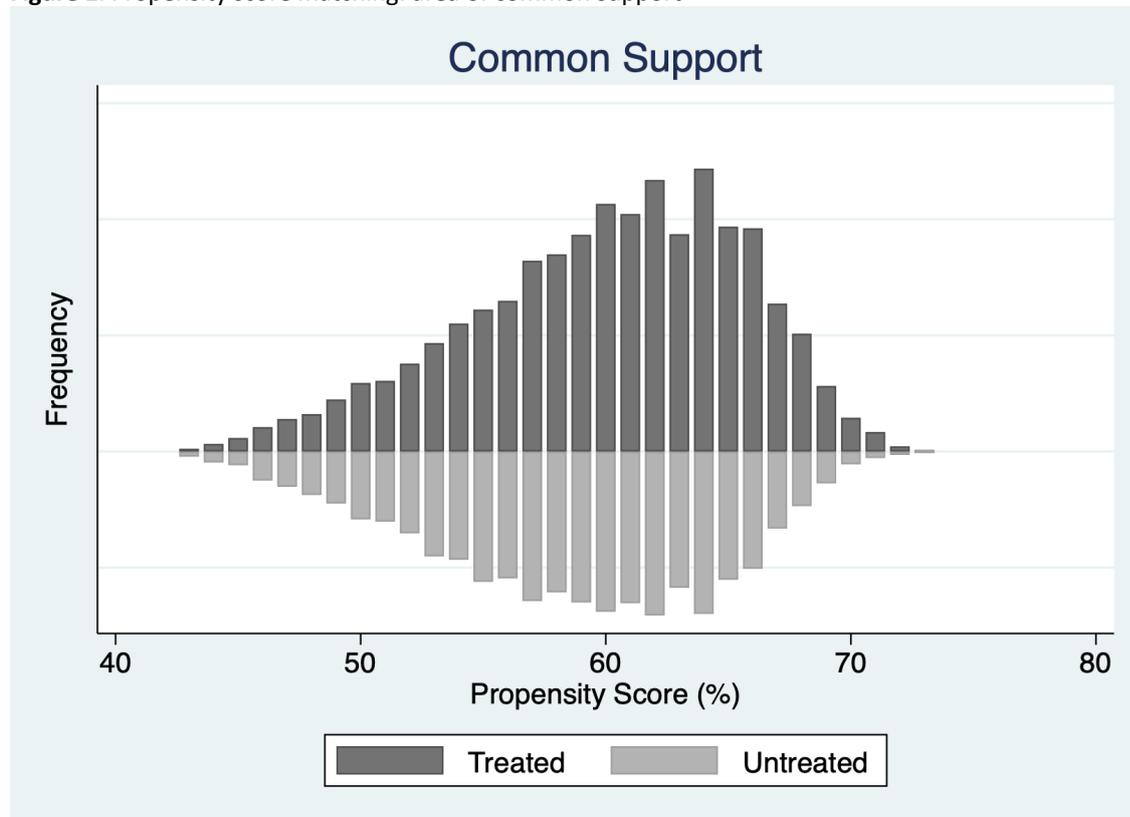
	ACS diagnosis		
	All	Inpatient	Outpatient
	b (95% CI)	b (95% CI)	b (95% CI)
ChEIs	-0.0056 (-.0091 to -.0021)	-0.0028 (-.0046 to -.0010)	-0.0022 (-.0036 to -.0008)
Baseline ^b	.0908	.0330	.0886
% increase over baseline	-6.18	-8.57	-2.61
person-months	1,591,920	1,591,740	1,591,920
N	26,532	26,529	26,532

^a Coefficients were calculated using linear regression models with weighting by odds. All regression models controlled for demographics, comorbidities, concomitant medications, time fixed-effects, and pre-treatment ACS admission variables, outlined in Table 1. Beneficiaries outside the area of common support were excluded from the analysis, hence the differences in N.

^b Mean ACS diagnosis probability in the untreated.

Figure 1 shows the area of common support by treatment status. The propensity score is displayed on the x-axis while the frequency is on the y-axis. The common support functions are smooth and the balancing property of the algorithm is satisfied. Beneficiaries outside of the area of common support were dropped in the linear regression models.

Figure 1. Propensity score matching: area of common support ^a



^a block # = 30. Propensity Score mean = .5891 sd = .0474 min = .4329 max = .6959 .

To further assess the robustness of our results, we conducted a placebo test using outcomes that are hypothesized to be unaffected by the treatment. As described above, we ran a linear regression with weighting by odds. This time, the outcome was probability of a non-heart-disease diagnosis. Results describe the probability of non-heart related diagnosis in the post-treatment observation period. Table 3 shows that none of the three models resulted in a significant beta coefficient of the treatment effect in this placebo test, in line with our expectations.

Table 3. Placebo test: probabilities of any non-heart related disease diagnosis in the observation period ^a

	Non-ACS diagnosis		
	All	Inpatient	Outpatient
	b (95% CI)	b (95% CI)	b (95% CI)
ChEIs	-0.003 (-.006 to .002)	-0.008 (-.015 to .008)	-0.006 (-.013 to .003)
Baseline ^b	.023	.028	.085
% increase over baseline	-12.54	-27.73	-7.45
person-months	1,591,980	1,591,320	1,591,860
N	26,533	26,522	26,531

^a Coefficients were calculated using linear regression models with weighting by odds. All regression models controlled for demographics, comorbidities, concomitant medications and pre-treatment ACS admission variables, outlined in Table 1. Beneficiaries outside the area of common support were excluded from the analysis, hence the differences in N.

^b Mean ACS diagnosis probability in untreated.

Discussion and Limitations

This study found a negative relationship between cholinesterase inhibitors therapy and acute coronary syndrome in the Medicare population of Louisiana from 2014 to 2018, in line with the findings emerging from previous published literature. Results are significant, in the hypothesized direction and in line with previous literature investigating the relationship between acetylcholinesterase inhibitors use and cardiac conditions.

In animal models, the relationship between cholinergic activity and heart disease has been previously investigated. Yang et al. (2019) show that ChEIs therapy regulates glucose metabolism and mitochondrial homeostasis to reduce myocardial vulnerability to injury in a mice model. Oikawa et al. (2021) describe reduced vagus nerve activity in genetically modified ChATKD mice possibly due to impaired cardiac nitric oxide production, which provides a possible biological explanation for the hypothesized recovery of the cholinergic system and non-neuronal cardiac cholinergic system (NNCCS) augmentation observed in donepezil users. A significant and negative relationship between ChEIs therapy and heart

disease is confirmed by other animal models such as in Durand et al. (2014), Sabino et al. (2013) and Li et al. (2013), where it is shown to improve cardiocirculatory function and long-term survival in rats with chronic heart failure.

Clinical trials and observational studies also investigated the relationship between ChEIs and heart conditions in humans. Androne et al. (2003) summarize the results of clinical trial NCT01415921, where heart rate recovery after completion of exercise was found to be significantly greater after the administration of ChEIs than placebo. Sangaletti et al. (2021) present the results of a clinical trial study observing that low-dose galantamine alleviates oxidative stress, alongside beneficial anti-inflammatory, and cardio-metabolic effects in subjects with metabolic syndrome. Dewland et al. (2007) find that, in a randomized controlled trial setting, sedentary adults ChEIs therapy decreased resting heart rate and increased postexercise heart rate recovery at one minute. In their review article, Roy et al. (2015) discuss relevant literature that implicate acetylcholine as a major regulator of cardiac remodeling and provide support for the notion that enhancing cholinergic signaling in human patients with cardiac disease can reduce morbidity and mortality. Serra et al. (2008) find that, in patients with chronic heart failure, ChEIs therapy was well tolerated, leading to improved hemodynamic profile during dynamic exercise. In a sample of sixty patients with ischemic heart disease and mild cognitive impairment, Wang et al. (2018b) show that donepezil administration resulted in significant reduction in mean heart rate and the lowest heart rate and prolongation of PR and RR intervals, whereas it had no significant effects on QRS duration. Khuanjing et al. (2020) discuss the possible mechanisms involved in the role of acetylcholinesterase inhibitors in acute myocardial infarction and heart failure remediation, suggesting that novel therapeutic approaches which moderate parasympathetic activities could be beneficial in the case of cardiovascular disease. Similarly, Olshansky et al. (2008) discuss the pathophysiology and potential implications for therapy of the parasympathetic nervous system and heart failure. Finally, Kaushik et al. (2018) report evidence from the literature that ChEIs treatment, prescribed for cognitive impairment, can reduce the load of medications in patients with AD by also addressing cardiovascular, gastrointestinal, and other comorbidities. Taken together, these findings from previous literature support the interpretation of our results.

Acute coronary syndrome is characterized, among others, by heart palpitations whereas cholinesterase inhibitors are not indicated in patients with low heartbeat. However, reverse

causality could still be a potential source of bias. We have taken specific measures to avoid it: first, we controlled for pre-treatment ACS episodes. Then, we used propensity scores with weighting by odds and ensured that pre-treatment comorbidities associated with heart conditions had a negligible standardized difference before proceeding with the regression model: these included pre-treatment atrial fibrillation, transient ischemic attack, ischemic heart disease and hypertension. Finally, we included only beneficiaries that were continuously enrolled for the whole observation period, and we captured the presence of comorbidities for at least one year prior to the treatment date.

In this study, we used a doubly-robust estimator to evaluate the association between ChEIS prescription and ACS risk and estimate the average treatment effect on the treated (ATT). To adjust for pre-treatment observable differences, we weighted by odds and we dropped participants whose distributions were lying outside of the area of common support. We included in the model all known comorbidities at the time of treatment, as well as demographic characteristics. This estimator is unbiased if at least one between the propensity score model and the outcome regression model is correctly specified. The propensity score captures the contributions of observables on the probability of receiving treatment. This estimator is superior to either simple linear regression or simple propensity score matching estimators. In this application, this approach helps mitigate potential selection bias arising from non-random treatment assignment. This ensures that the treated and untreated samples are balanced on observable characteristics. However, we could not control for any time-varying unobservable characteristics. Also, there might be omitted variables that would alter the propensity score algorithm results, which we are not aware of. Analysing claims data, we used drugs prescriptions as a proxy to model drugs consumption. However, we have no information on the proportion of medication prescriptions that translated into actual medication consumption. We had a 5-year observation period, ranging from 2014 to 2018. Considering this, the pre- and post-treatment observation period may be limited. However, we address this limitation by including in the study only those Medicare beneficiaries that were continuously enrolled for the whole length of the observation period. Another limitation of this study is that our results provide no information on the subject-specific effect of ChEIs on ACS risk: we estimated the average treatment effect, but further stratification based on genetic information or other clinical criteria is strongly warranted before providing individual counselling.

It is estimated that one in every four deaths in the US, about 659,000 people every year, can be attributed to heart disease. Heart disease costs the United States about \$363 billion each year from 2016 to 2017 (Virani, 2021). If off-label medications could help reduce the risk of heart disease, this could have important public health implications. The generalizability of our findings to different populations and other settings should be further assessed within a randomized, double-blind placebo controlled clinical trial, enrolling patients with relevant comorbidities including Alzheimer's disease and related dementia.

Conclusion

There is mounting evidence on the possible role of ChEIs in reducing risk of heart disease. In this cohort of Louisiana Medicare beneficiaries with AD, we found that ChEIs users were significantly less likely than non-users to develop acute coronary syndrome, a type of vascular-related heart condition. This evidence should be considered in clinical decisions regarding ChEIs utilization in AD patients, and comorbidity care should be integrated into current management for patients with AD. Further research should investigate whether prescription of ChEIs to prevent ACS risk could be justified in non-AD patients, and a thorough cost-effectiveness analysis should be conducted considering both known and potential adverse effects of the intervention.

References

- Albrecht J, Hanna M, Randall R, et al. An Algorithm to Characterize a Dementia Population by Disease Subtype. *Alzheimer Disease & Associated Disorders*. 2019; 33(2): 118–123.
- Androne AS, Hryniewicz K, Goldsmith R, Arwady A, Katz SD. Acetylcholinesterase inhibition with pyridostigmine improves heart rate recovery after maximal exercise in patients with chronic heart failure. *Heart*. 2003 Aug;89(8):854-8. doi: 10.1136/heart.89.8.854. PMID: 12860856; PMCID: PMC1767776.
- Austin G. The 'reversal of fortune' thesis and the compression of history: Perspectives from African and comparative economic history. *Journal of International Development: The Journal of the Development Studies Association*. 2008 Nov;20(8):996-1027.
- Colović MB, Krstić DZ, Lazarević-Pašti TD, Bondžić AM, Vasić VM. Acetylcholinesterase inhibitors: pharmacology and toxicology. *Curr Neuropharmacol*. 2013 May;11(3):315-35.
- D'Agostino RB (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statist. Med*. 1998; 17, 2265—2281.

Dehejia RH and Wahba S (2002). Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics*, February 2002, 84(1): 151–161.

Dewland TA, Androne AS, Lee FA, Lampert RJ, Katz SD. Effect of acetylcholinesterase inhibition with pyridostigmine on cardiac parasympathetic function in sedentary adults and trained athletes. *Am J Physiol Heart Circ Physiol*. 2007 Jul;293(1):H86-92. doi: 10.1152/ajpheart.01339.2006. Epub 2007 Feb 23. PMID: 17322413.

Durand MT, Becari C, de Oliveira M, do Carmo JM, Silva CA, Prado CM, Fazan R Jr, Salgado HC. Pyridostigmine restores cardiac autonomic balance after small myocardial infarction in mice. *PLoS One*. 2014 Aug 18;9(8):e104476. doi: 10.1371/journal.pone.0104476. PMID: 25133392; PMCID: PMC4136726.

Harder VS, Stuart EA, Anthony JC. Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological methods*. 2010 Sep;15(3):234.

Hirano K, Imbens GW, Ridder G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*. 2003 Jul;71(4):1161-89.

Hsiao SH, Hwang TJ, Lin FJ, Sheu JJ, Wu CH. The Association Between the Use of Cholinesterase Inhibitors and Cardiovascular Events Among Older Patients With Alzheimer Disease. *Mayo Clin Proc*. 2021 Feb;96(2):350-362. doi: 10.1016/j.mayocp.2020.05.048. PMID: 33549256.

Hsieh MJ, Chen DY, Lee CH, Wu CL, Chen YJ, Huang YT, Chang SH (2022). Association Between Cholinesterase Inhibitors and New-Onset Heart Failure in Patients With Alzheimer's Disease: A Nationwide Propensity Score Matching Study. *Front Cardiovasc Med*. 2022 Mar 16;9:831730. doi: 10.3389/fcvm.2022.831730. PMID: 35369359; PMCID: PMC8966646.

Isik, A.T., Soysal, P., Stubbs, B., Solmi, M., Basso, C., Maggi, S., Schofield, P., Veronese, N. and Mueller, C. (2018), Cardiovascular Outcomes of Cholinesterase Inhibitors in Individuals with Dementia: A Meta-Analysis and Systematic Review. *J Am Geriatr Soc*, 66: 1805-1811.

Kaushik V, Smith ST, Mikobi E, Raji MA. Acetylcholinesterase Inhibitors: Beneficial Effects on Comorbidities in Patients With Alzheimer's Disease. *Am J Alzheimers Dis Other Demen*. 2018 Mar;33(2):73-85. doi: 10.1177/1533317517734352. Epub 2017 Oct 3. PMID: 28974110.

Khuanjing T, Palee S, Chattipakorn SC, Chattipakorn N. The effects of acetylcholinesterase inhibitors on the heart in acute myocardial infarction and heart failure: From cells to patient reports. *Acta Physiol (Oxf)*. 2020 Feb;228(2):e13396. doi: 10.1111/apha.13396. Epub 2019 Nov 6. PMID: 31595611.

Kumar A, Sidhu J, Goyal A, Tsao JW. Alzheimer Disease. 2021 Aug 11. In: *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan–. PMID: 29763097.

Li M, Zheng C, Kawada T, Inagaki M, Uemura K, Shishido T, Sugimachi M. Donepezil markedly improves long-term survival in rats with chronic heart failure after extensive myocardial infarction. *Circ J*. 2013;77(10):2519-25. doi: 10.1253/circj.cj-13-0476. Epub 2013 Jul 5. PMID: 23832513.

Oikawa S, Kai Y, Mano A, Ohata H, Kurabayashi A, Tsuda M, Kakinuma Y. Non-neuronal cardiac acetylcholine system playing indispensable roles in cardiac homeostasis confers resiliency to the heart. *J Physiol Sci*. 2021 Jan 18;71(1):2. doi: 10.1186/s12576-020-00787-6. PMID: 33461483.

Olshansky B, Sabbah HN, Hauptman PJ, Colucci WS. Parasympathetic nervous system and heart failure: pathophysiology and potential implications for therapy. *Circulation*. 2008 Aug 19;118(8):863-71. doi: 10.1161/CIRCULATIONAHA.107.760405. PMID: 18711023.

Roy A, Guatimosim S, Prado VF, Gros R, Prado MA. Cholinergic activity as a new target in diseases of the heart. *Mol Med*. 2015 Jan 26;20(1):527-37. doi: 10.2119/molmed.2014.00125. PMID: 25222914; PMCID: PMC4365064.

Sabino JP, da Silva CA, de Melo RF, Fazan R Jr, Salgado HC. The treatment with pyridostigmine improves the cardiocirculatory function in rats with chronic heart failure. *Auton Neurosci*. 2013 Jan;173(1-2):58-64. doi: 10.1016/j.autneu.2012.11.007. Epub 2012 Dec 4. PMID: 23218833.

Sangaletti CT, Katayama KY, De Anglis K, et al. (2021). The cholinergic drug galantamine alleviates oxidative stress alongside anti-inflammatory and cardio-metabolic effects in subjects with the metabolic syndrome in a randomized trial. *Front. Immunol.*, 11 March 2021 Sec. Inflammation.

Serra SM, Costa RV, Teixeira De Castro RR, Xavier SS, Nóbrega AC. Cholinergic stimulation improves autonomic and hemodynamic profile during dynamic exercise in patients with heart failure. *J Card Fail*. 2009 Mar;15(2):124-9. doi: 10.1016/j.cardfail.2008.10.018. Epub 2008 Dec 2. PMID: 19254671.

Stampfer MJ (2006). Cardiovascular disease and Alzheimer's disease: common links. *J Intern Med*. 2006 Sep;260(3):211-23. doi: 10.1111/j.1365-2796.2006.01687.x. PMID: 16918818.

Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. (2021). Heart disease and stroke statistics—2021 update: a report from the American Heart Associationexternal icon. *Circulation*. 2021;143:e254–e743.

Wang JH, Wu YJ, Tee BL, Lo RY (2018a). Medical Comorbidity in Alzheimer's Disease: A Nested Case-Control Study. *J Alzheimers Dis*. 2018;63(2):773-781.

Wang D, Wu Y, Wang A, Chen Y, Zhang T, Hu N (2018b). Electrocardiogram Changes of Donepezil Administration in Elderly Patients with Ischemic Heart Disease. *Cardiology Research and Practice*, vol. 2018, Article ID 9141320, 5 pages, 2018.

Wu, PH., Lin, YT., Hsu, PC. et al. (2015). Impact of acetylcholinesterase inhibitors on the occurrence of acute coronary syndrome in patients with dementia. *Sci Rep* 5, 15451 (2015).

Yang Y, Zhao M, Yu XJ, Liu LZ, He X, Deng J, Zang WJ. Pyridostigmine regulates glucose metabolism and mitochondrial homeostasis to reduce myocardial vulnerability to injury in diabetic mice. *Am J Physiol Endocrinol Metab*. 2019 Aug 1;317(2):E312-E326. doi: 10.1152/ajpendo.00569.2018. Epub 2019 Jun 18. PMID: 31211620.

PAPER 2

Bone density scan reduces risk of fracture among Medicare beneficiaries with Alzheimer's disease and related dementia

Abstract

Alzheimer's disease and related dementia (ADRD) is associated with an increased risk of fractures, due to cognitive decline, balance dysfunction and mobility limitations. Osteoporosis and other comorbidities associated with older age can increase the risk of fractures. Dual-energy X-ray absorptiometry (DXA) scan is a tool used to evaluate bone mineral density and enables clinicians to assess an individual's fracture risk. Using a linear regression model with bootstrap standard errors and weighting by odds, this study evaluates the average treatment effect on the treated (ATT) of DXA in reducing likelihood of subsequent hip and femur fracture in the ADRD Medicare population of Louisiana between 2014 and 2018. In the analytical sample, DXA was associated with a 30.09% reduction in the likelihood of fractures among ADRD beneficiaries and a 33.92% among non-ADRD beneficiaries. DXA was also significantly associated with osteoporosis medication prescription in both samples. Taken together, our study reinforces the significance of incorporating bone density testing into routine care for Medicare beneficiaries with dementia.

Keywords: Alzheimer's disease, osteoporosis, bone mineral density, DXA

Introduction

Fractures pose a significant health concern worldwide, leading to pain, disability, and increased mortality rates, particularly among the elderly population. Osteoporosis, a condition characterized by low bone mass and structural deterioration of bone tissue, is a primary risk factor for fractures. Patients with Alzheimer's disease and related dementia (ADRD) are at an increased risk of falls and fractures, which can be mainly attributed to osteoporosis and reduced bone mineral density (BMD) (Wang et al., 2014).

To address this pressing issue, numerous diagnostic tools have emerged, with bone density tests standing at the forefront. These tests, such as Dual-energy X-ray Absorptiometry (DXA), provide valuable information about bone mineral density (BMD) and enable clinicians to assess an individual's fracture risk. While the utility of bone density tests in diagnosing osteoporosis is well-established, their effectiveness in preventing fractures remains a topic of ongoing research and debate.

This manuscript aims to investigate the impact of bone density tests on fracture prevention among ADRD and non-ADRD Medicare beneficiaries, exploring their potential as a valuable tool in reducing fracture incidence, enhancing clinical decision-making, and formulating preventive strategies. Drawing from the existing literature and analyzing claims data obtained from the Centers for Medicare and Medicaid Services (CMS), this study seeks to shed light on the crucial role of bone density tests in the realm of fracture prevention, ultimately contributing to evidence-based healthcare practices and improved patient outcomes.

ADRD is an independent risk factor for fracture, hence the importance of investigating the occurrence of fracture in both ADRD and non-ADRD populations, and trace comparisons between the two groups. Several independent risk factors in ADRD patients have been associated with increased fracture risk. Cognitive decline was found to be independently associated with bone loss in a prospective population-based study employing mixed-effect models (Bliuc et al., 2021). In turn, bone loss, characterized by decreased bone density and strength, causes bones to be more susceptible to fractures (Keating et al., 2005).

Furthermore, there is a well-established association between mobility decline and worsening dementia in individuals with ADRD. As dementia progresses, it often leads to various cognitive impairments, affecting memory, thinking, and problem-solving abilities.

Simultaneously, physical abilities, including mobility and gait, tend to deteriorate (Tolea et al., 2016). As the cognitive and physical impairments progress in individuals with ADRD, balance dysfunction becomes more prevalent and can significantly impact their overall well-being and safety (Hill, 2009). These factors found in ADRD, as cognitive impairment, balance issues and mobility limitations, can all lead to increased fracture risk.

Beyond factors that are related to ADRD, there exist other independent risk factors, or comorbidities, that have been associated with increased fracture risk. These risk factors include rheumatoid arthritis (Xue et al., 2017), diabetes mellitus (Valderrábano et al., 2018), chronic pulmonary disease (Ionescu et al., 2003), bone cancer (Tsuzuki, 2017), lung cancer (Oliveira et al., 2018), cancer to the liver (Vestergaard et al., 2009), bladder cancer (Gupta et al., 2014), prostate cancer (Melton et al., 2012), cardiovascular disease (Veronese et al., 2017) and chronic kidney disease (Nickolas et al., 2008).

Bone density tests have an important role in the evaluation of individuals at risk of osteoporosis, and in helping clinicians advise patients about the appropriate use of osteoporosis treatment (Blake et al., 2007). Bone density tests have been confirmed as an effective tool for fracture prevention in selected populations (Suarez-Almazor et al., 2022) but no study to date has directly investigated the effect of bone density tests in effectively reducing hip fracture risk in the ADRD population.

Studying fracture occurrences in ADRD and non-ADRD populations not only sheds light on the vulnerability of dementia patients to fractures but also emphasizes the importance of targeted preventive interventions for this at-risk group. This study investigates the effect of bone density testing in reducing risk of hip fracture in the US Medicare population with a diagnosis of Alzheimer's disease and related dementia, taking into account observed comorbidities at the time of the intervention. We further include an analysis on the non-ADRD population to elucidate different causal patterns across different groups.

Methods

We used a weighted linear regression to estimate the impact of DXA on risk of hip and femur fracture. Weighting by odds was used to estimate the average treatment effect (ATT). With weighting by the odds, treated individuals receive a weight equal to 1 and comparison

individuals receive a weight equal to their propensity score (p_i), converted to the odds scale ($= p_i / (1 - p_i)$) (Hirano et al., 2003). This weighting effectively up-weights comparison individuals whose measured covariate values (propensity scores) best match those of the treated individuals and down-weights comparison individuals whose measured covariate values are dissimilar from treated individuals. One way of thinking about weighting by the odds is that the comparison individuals are first weighted to the entire population, using $1 / (1 - p_i)$, and are then weighted to the treatment group, using p_i . (Harder et al, 2010).

Propensity scores were estimated using the `psmatch2` command in Stata 17. This command employs a logistic regression model, where the treatment assignment serves as the dependent variable. The relevant covariates, determined based on the examined literature, were included as independent variables. The estimated propensity scores represent the probability of receiving the treatment conditional on the observed covariates.

To assess the comparability of treated and control samples, we computed the standardized difference between treated and control groups across variables. Balance diagnostics were conducted using the `psbalance` command in Stata. The `psbalance` command examines the distribution of covariates between the treated and control groups after weighting and provides statistical tests and graphical displays to assess balance. The balance diagnostics ensure that the weighted groups are comparable and reduce the potential for bias due to confounding. We assumed that an acceptable standardized difference would be lower than 10% (Austin, 2008). The functional form of the `psmatch2` model was modified by including interaction terms and higher order polynomials until satisfactory standardized differences were obtained for all variables.

Finally, the ATT was estimated using the weighted sample. The differences in the outcome variable between the treated and control groups were computed and tested for statistical significance using a t-test. Sensitivity analyses were also conducted to examine the robustness of the results and assess the potential impact of unobserved confounders and individual-level fixed effects were included in the model to control for time-invariant unobservables.

To allow for an estimation of the likelihood of fracture occurring after the bone density test in both treated and untreated groups, the following algorithm was implemented. First, for each year in the observation period, the distribution of treatment dates in the treated

sample was tabulated. Then, in the untreated sample, a numerical variable was created using a random number generator and the sample was sorted. The tabulated treatment dates distribution from the treated group was assigned to the sorted untreated sample. The process was repeated iteratively for all months included in the observation period. By doing so, treatment dates based on the distribution of treatment dates from the treated sample were assigned to the untreated group. Random assignment of treatment dates in the untreated was therefore constrained to the treatment dates distribution in the treated population. Finally, to ensure standard errors were adjusted for the random assignment of treatment dates, we implemented the procedure inside a bootstrap algorithm with replacement with 10,000 iterations over the entire analytical sample.

This method allowed for the estimation of the ATT, ensuring reliable causal inferences about the treatment's impact on the outcome variable.

Data

To identify the ADRD sample, we first considered all Medicare beneficiaries with at least one ADRD ICD-9/10 claim in the five-year period from 2014 to 2018, based on the ICD-9/10 codes in Albrecht (2019), resulting in 45,396 beneficiaries. Then, we restricted the sample to those that were continuously enrolled in the 5-year period, looking at whether they had at least one claim in each of the years considered. This resulted in a sample of 26,535 beneficiaries. Finally, to limit potential reverse causality bias, we excluded all beneficiaries that ever had an osteoporosis medication prescription before their first bone density test on records. The final ADRD sample consisted of 20,262 Medicare beneficiaries.

We used a similar approach to identify the non-ADRD analytical sample. We first considered all Medicare beneficiaries with at least one non-ADRD ICD-9/10 claim in the observation period, resulting in 1,197,810 beneficiaries. Then, we restricted the sample to those who had at least one claim in each of the five years in the observation period, resulting in 863,497 beneficiaries. Finally, we excluded those who had an osteoporosis medication prescription prior to their osteoporosis test. The final non-ADRD analytical sample consisted of 631,305 beneficiaries.

Results

Table 1 shows the covariate means stratified by treatment status and weighted by odds. Standardized differences were used to determine whether the treated and control samples are comparable. Several covariates had a standardized difference higher than 10% in the crude unweighted sample. At the time of DXA, average age had a standardized difference of -12.11%. Sex had a standardized difference of -49.15%. Black or African American ethnicity had a standardized difference of -11.96%. Breast cancer and prostate cancer had standardized differences of 17.99% and -11.84%, respectively. Cataract diagnoses had a standardized difference of 17.43%. Congestive heart failure had a standardized difference of -11.19%. Hyperlipidemia had a standardized difference of 12.05%. Hyperplasia had a standardized difference of -34.64%. Osteoporosis had a standardized difference of 44.22%. Rheumatoid arthritis and osteoarthritis had a standardized difference of 16.08%. After weighting by odds, the two groups became comparable, as all standardized differences were below the threshold of 10%. Hip or femur fracture was the variable with the highest standardized difference after weighting, equal to -7.28%. The standardized differences suggest that treated and untreated groups were balanced after weighting by odds of treatment. The common support graph is shown in Figure A1 in the Appendix, providing a visual representation of the balance achieved after weighting by odds.

Table 1. Bootstrap weighted means, with standardized differences, ADRD beneficiaries. ^a

	Crude			Adjusted		
	Untreated (N=16,429)	Treated (N=3,833)	Standardized difference	Untreated (N=16,429)	Treated (N=3,832)	Standardized difference
	Mean	Mean	Δ	Mean	Mean	Δ
<i>Demographics:</i>						
Age	80.67	79.33	-.1211	78.92	79.31	.0327
Male	.4000	.1171	-.4915	.1146	.1172	.0056
Black or African American	.2600	.1944	-.1196	.1964	.1944	-.0035
<i>Comorbidities:</i>						
Acute myocardial infarction	.0890	.0681	-.0572	.0727	.0681	-.0128
Anemia	.8285	.8072	-.0391	.8172	.8072	-.0181
Asthma	.1667	.2069	.0731	.2110	.2069	-.0071
Atrial fibrillation	.2584	.2285	-.0492	.2284	.2286	.0003
Cancer, breast	.0534	.1255	.1799	.1282	.1255	-.0056

Cancer, colorectal	.0428	.0425	-.0009	.0414	.0425	.0039
Cancer, endometrial	.0075	.0125	.0358	.0114	.0125	.0076
Cancer, lung	.0190	.0198	.0043	.0203	.0198	-.0024
Cancer, prostate	.0697	.0329	-.1184	.0314	.0329	.0061
Cataract	.7937	.8836	.1743	.8808	.8836	.0063
Congestive heart failure	.5247	.4459	-.1119	.4556	.4460	-.0136
Chronic kidney disease	.6009	.5573	-.0626	.5692	.5572	-.0171
Chronic obstructive pulmonary disease	.4476	.4263	-.0303	.4335	.4264	-.0101
Depression	.7146	.7172	.0041	.7294	.7171	-.0194
Diabetes	.5313	.4720	-.0841	.4865	.4721	-.0204
Glaucoma	.2706	.3063	.0557	.3028	.3064	.0055
Hyperlipidemia	.8738	.9249	.1205	.9273	.9248	-.0067
Hyperplasia	.2661	.0851	-.3464	.0819	.0851	.0081
Hypertension	.9599	.9611	.0045	.9632	.9611	-.0076
Hypothyroidism	.4011	.4688	.0969	.4741	.4687	-.0077
Ischemic heart disease	.7213	.7096	-.0184	.7161	.7096	-.0102
Osteoporosis	.2698	.5640	.4422	.5770	.5639	-.0187
Rheumatoid arthritis and osteoarthritis	.7885	.8732	.1608	.8768	.8734	-.0072
Transient ischemic attack	.4080	.3741	-.0491	.3867	.3742	-.0183
<i>Excluded comorbidities</i>						
Hip or femur fracture	.1303	.1231	-.0151	.1590	.1232	-.0728

^a Variable means in unweighted (crude) and weighted (adjusted) models. Standardized differences were calculated as the difference between the means divided by the standard deviation of treated and untreated groups. Comorbidities refer to events occurring before the bone density test date. Excluded comorbidities were not used in the calculation of the propensity scores.

Table 2 shows the crude and the adjusted sample means with standardized differences for the non-ADRDR analytical sample. As with the ADRDR sample, after weighting by odds, the standardized differences decreased below the threshold of 10% in all instances. Hip fracture was not used in the calculation of propensity scores but, after including all other variables and weighting by odds of treatment, it maintained a standardized difference of 5.70%, below the threshold of 10%, thus allowing for comparison between untreated and treated groups. The common support graph is shown in Figure A2 in the Appendix. By comparing the information in Table 1 and Table 2, it can be noticed that treated ADRDR beneficiaries were on average older than non-ADRDR beneficiaries (79.33 vs 71.93) and had a significantly higher prevalence of pre-intervention hip and femur fracture (12.31% vs 2.34%).

Table 2. Bootstrapped weighted means, with standardized differences, non-ADRD beneficiaries ^a

	Crude			Adjusted		
	Untreated (N=530,834)	Treated (N=100,471)	Standardized difference	Untreated (N=530,834)	Treated (N=100,471)	Standardized difference
	Mean	Mean	Δ	Mean	Mean	Δ
<i>Demographics:</i>						
Age	68.96	71.93	.1919	72.7400	71.9298	-.0540
Male	.5293	.0926	-.7566	.0846	.0926	.0199
Black or African American	.2959	.1991	-.1596	.1950	.1991	.0074
<i>Comorbidities:</i>						
Acute myocardial infarction	.0366	.0361	-.0021	.0372	.0361	-.0041
Anemia	.3649	.5783	.3094	.6091	.5783	-.0443
Asthma	.0921	.1677	.1601	.1746	.1677	-.0130
Atrial fibrillation	.0928	.1360	.0962	.1447	.1360	-.0177
Cancer, breast	.0222	.1103	.2545	.1236	.1103	-.0293
Cancer, colorectal	.0186	.0266	.0380	.0285	.0266	-.0082
Cancer, endometrial	.0035	.0114	.0653	.0127	.0114	-.0087
Cancer, lung	.0097	.0148	.0330	.0161	.0148	-.0071
Cancer, prostate	.0420	.0193	-.0931	.0182	.0193	.0062
Cataract	.3894	.7537	.5599	.7706	.7537	-.0281
Congestive heart failure	.1931	.2543	.1041	.2712	.2543	-.0272
Chronic kidney disease	.2397	.3646	.1941	.3849	.3646	-.0296
Chronic obstructive pulmonary disease	.1988	.2902	.1512	.3089	.2902	-.0289
Depression	.2719	.4312	.2394	.4485	.4312	-.0246
Diabetes	.2981	.3822	.1259	.3975	.3822	-.0221
Glaucoma	.1365	.2462	.1991	.2581	.2462	-.0195
Hyperlipidemia	.5495	.8621	.5163	.8792	.8621	-.0361
Hyperplasia	.1524	.0517	-.2384	.0474	.0517	.0142
Hypertension	.6364	.8738	.4061	.8958	.8738	-.0486
Hypothyroidism	.1613	.3790	.3577	.4011	.3790	-.0320
Ischemic heart disease	.3576	.5034	.2104	.5219	.5034	-.0263
Osteoporosis	.0668	.4463	.6821	.4800	.4463	-.0479
Rheumatoid arthritis and osteoarthritis	.4001	.7347	.5075	.7607	.7347	-.0423
Transient ischemic attack	.1055	.1580	.1101	.1678	.1580	-.0189
<i>Excluded comorbidities</i>						
Hip or femur fracture	.0198	.0234	-.0411	.0164	.0136	-.0570

^a Variable means in unweighted (crude) and weighted (adjusted) models. Standardized differences were calculated as the difference between the means divided by the standard deviation of treated and untreated groups. Comorbidities refer to events occurring before the bone density test date. Excluded comorbidities were not used in the calculation of the propensity scores.

Table 3 shows the probability of hip fracture in a cohort of Medicare beneficiaries with a diagnosis of ADRD compared to those without ADRD. The table includes two sets of

probabilities: unweighted and adjusted treatment effects (ATT) for both ADRD and non-ADRD groups, continuously enrolled from 2014 to 2018. The probabilities are associated with bone density tests and are presented with their corresponding confidence intervals (CI). The baseline probabilities for both ADRD and non-ADRD groups are also provided. The coefficients in the table were calculated using linear regression models with bootstrap standard errors and weighted by odds. The models controlled for demographics, comorbidities, and pre-treatment outcome variables. Beneficiaries outside the area of common support were excluded from the analysis.

We found that in our sample having had a bone density test was associated with a reduced subsequent risk of hip fracture. After applying to the outcome a factor of one thousand, the intervention was associated with a reduction of .542 percentage points (-1.246 to -.312) in the probability of ever having had a fracture during the observation period for ADRD beneficiaries, or a 30.09 percent decrease over baseline. On the other hand, the intervention was associated with a reduction of .114 percentage points (-.214 to -.022) in the probability of ever having had a fracture during the observation period for non-ADRD beneficiaries, or a reduction of 33.92% over baseline value.

Table 3. Probabilities of hip fracture in the observation period ^a

	ADRD		Non-ADRD	
	Unweighted	ATT	Unweighted	ATT
Bone density test ^b (Normal-based 95% CI)	- .542 (-1.246 to -.312)	- .585 (-1.352 to -.144)	- .121 (-.415 to -.135)	- .114 (-.214 to -.022)
Baseline ^c	1.945	1.944	.336	.336
% increase over baseline	-27.86	-30.09	-36.01	-33.92
person-months	1,215,720	1,215,660	37,878,300	37,878,300
N	20,262	20,261	631,305	631,305

^a Coefficients were calculated using linear regression models with bootstrap standard errors and weighting by odds. All regression models controlled for demographics, time-fixed effects, comorbidities and pre-treatment outcome variables. Beneficiaries outside the area of common support were excluded from the analysis, hence the differences in N.

^b The outcome variable was rescaled by a factor of 1,000 to ease readability of coefficients.

^c Mean hip fracture probability in untreated

To shed light on the causal pathway linking bone density test and decreased likelihood of hip and femur fractures, we investigated the effect of the intervention on osteoporosis medications prescription. Table 4 presents the probabilities of osteoporosis medication prescription in the observation period for ADRD and non-ADRD beneficiaries. The table includes two sets of probabilities: unweighted and adjusted treatment effects (ATT) for both

ADRD and non-ADRD groups. The probabilities are associated with bone density tests and are presented with their corresponding confidence intervals (CI).

The coefficients in the table were calculated using linear regression models with bootstrap standard errors and weighted by odds. The regression models controlled for demographics, comorbidities, and pre-treatment outcome variables. Beneficiaries outside the area of common support were excluded from the analysis. The outcome variable, representing the probability of osteoporosis medication prescription, was rescaled by a factor of 1,000. Results are significant and in the expected direction for both the ADRD (.016 to .021) and the non-ADRD (.046 to .107) samples. The baseline probabilities for both ADRD and non-ADRD groups are set to zero since all beneficiaries with an osteoporosis medication prescription in the pre-period were excluded from the analysis.

Table 4. Probabilities of osteoporosis medication prescription in the observation period, ADRD beneficiaries ^a

	ADRD		Non-ADRD	
	Unweighted	ATT	Unweighted	ATT
Bone density test ^b (Normal-based 95% CI)	.031 (.029 to .033)	.020 (.016 to .021)	.065 (.042 to .093)	.077 (.046 to .107)
Baseline ^c	0	0	0	0
% increase over baseline person-months	- 1,215,720	- 1,215,660	- 37,878,300	- 37,878,300
N	20,262	20,261	631,305	631,305

^a Coefficients were calculated using linear regression models with bootstrap standard errors and weighting by odds. All regression models controlled for demographics, comorbidities and pre-treatment outcome variables. Beneficiaries outside the area of common support were excluded from the analysis, hence the differences in N.

^b The outcome variable was rescaled by a factor of 1,000 to ease readability of coefficients.

^c All beneficiaries with an osteoporosis medication prescription in the pre-period were excluded from the analysis.

Discussion

This study builds upon existing literature and provides robust evidence supporting the benefits of bone density tests in Medicare beneficiaries with dementia. Consistent with prior research, our findings demonstrate that bone density tests are associated with a significant reduction in hip fracture risk among individuals with dementia. This association underscores the importance of early identification of individuals at a higher risk of fractures, allowing for targeted preventive measures and interventions to mitigate the substantial burden of hip fractures in this vulnerable population.

Moreover, our study contributes to the growing body of evidence highlighting the positive impact of bone density tests on osteoporosis treatment implementation in individuals with

dementia. Our findings align with previous studies and suggest that bone density testing serves as a catalyst for appropriate management and treatment of osteoporosis in this population. By identifying those at risk and initiating timely interventions, bone density tests can potentially improve the healthcare outcomes of individuals with dementia and reduce the associated morbidity and mortality.

We used an inverse probability of treatment effects estimator with weighting by odds, combining outcome regression with weighing by propensity score, and we applied a bootstrap program with replacement to ensure that random assignment of treatment dates in the untreated group would not introduce estimation bias in the results. This approach is meant to produce robust estimates. However, there are several limitations that should be considered when evaluating the results. First, the propensity scores were calculated using observable variables which were available in the Centers for Medicare and Medicaid Services dataset. In the regression model, we were only able to control for observables. We limited potential bias in our estimates introduced by unobservables, by controlling for individual-level fixed effects, but we cannot exclude unobserved heterogeneity between beneficiaries that received the intervention and those who did not. If more advantaged patients were more likely to receive the bone density scan, and had lower propensity towards the likelihood of hip fracture in the post-treatment period, then our results are biased away from zero. Another limitation is that our observation period was limited to five years only, which may reduce our capability to explore any effects of the intervention occurring beyond this period. Finally, the limited observation period may have limited our ability to correctly identify and exclude from the sample all those beneficiaries that had ever received an osteoporosis medication prescription at the time of the treatment. However, we mitigated this potential bias by including pre-treatment osteoporosis diagnosis in the propensity scores calculation used in the weighted sample.

Conclusion

Taken together, our study reinforces the significance of incorporating bone density testing strategies into routine care for Medicare beneficiaries with dementia. The observed associations between bone density tests, reduced hip fracture risk, and increased prescription of osteoporosis medications highlight the potential of these tests to enhance

the overall healthcare management of this vulnerable population. Future research should focus on exploring the cost-effectiveness and long-term outcomes of implementing bone density testing strategies in individuals with dementia, further strengthening the evidence base for informed clinical decision-making and policy development.

References

Albrecht J, Hanna M, Randall R, et al. An Algorithm to Characterize a Dementia Population by Disease Subtype. *Alzheimer Disease & Associated Disorders*. 2019; 33(2): 118–123.

Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in medicine*. 2008 May 30;27(12):2037-49.

Blake GM, Fogelman I. The role of DXA bone density scans in the diagnosis and treatment of osteoporosis. *Postgraduate medical journal*. 2007 Aug;83(982):509-17.

Bliuc D, Tran T, Adachi JD, Atkins GJ, Berger C, van den Bergh J, Cappai R, Eisman JA, van Geel T, Geusens P, Goltzman D. Cognitive decline is associated with an accelerated rate of bone loss and increased fracture risk in women: a prospective study from the Canadian Multicentre Osteoporosis Study. *Journal of Bone and Mineral Research*. 2021 Nov;36(11):2106-15.

Gupta A, Atoria CL, Ehdaie B, Shariat SF, Rabbani F, Herr HW, Bochner BH, Elkin EB. Risk of fracture after radical cystectomy and urinary diversion for bladder cancer. *Journal of Clinical Oncology*. 2014 Oct 10;32(29):3291.

Harder VS, Stuart EA, Anthony JC. Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological methods*. 2010 Sep;15(3):234.

Hill KD, LoGiudice D, Lautenschlager NT, Said CM, Dodd KJ, Suttanon P. Effectiveness of balance training exercise in people with mild to moderate severity Alzheimer's disease: protocol for a randomised trial. *BMC geriatrics*. 2009 Dec;9(1):1-9.

Hirano K, Imbens G, Ridder G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*. 2003;71:1161–1189.

Ionescu AA, Schoon E. Osteoporosis in chronic obstructive pulmonary disease. *European Respiratory Journal*. 2003 Nov 2;22(46 suppl):64s-75s.

Keating JF, Simpson AH, Robinson CM. The management of fractures with bone loss. *The Journal of Bone & Joint Surgery British Volume*. 2005 Feb 1;87(2):142-50.

McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack Jr CR, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC. The diagnosis of dementia due to Alzheimer's

disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia*. 2011 May 1;7(3):263-9.

Melton III LJ, Lieber MM, Atkinson EJ, Achenbach SJ, Zincke H, Therneau TM, Khosla S. Fracture risk in men with prostate cancer: a population-based study. *Journal of Bone and Mineral Research*. 2011 Aug;26(8):1808-15.

Nickolas TL, Leonard MB, Shane E. Chronic kidney disease and bone fracture: a growing concern. *Kidney international*. 2008 Sep 2;74(6):721-31.

Oliveira MB, Marques BD, Matos RA, Fontenelle CR, Mello FC, Paschoal ME. Pathological fractures due to bone metastases from lung cancer: risk factors and survival. *Acta ortopedica brasileira*. 2018 Dec 4;26:388-93.

Suarez-Almazor ME, Pundole X, Cabanillas G, Lei X, Zhao H, Elting LS, Lopez-Olivo MA, Giordano SH. Association of bone mineral density testing with risk of major osteoporotic fractures among older men receiving androgen deprivation therapy to treat localized or regional prostate cancer. *JAMA network open*. 2022 Apr 1;5(4):e225432-.

Tolea MI, Morris JC, Galvin JE. Trajectory of mobility decline by type of dementia. *Alzheimer disease and associated disorders*. 2016 Jan;30(1):60.

Tsuzuki S, Park SH, Eber MR, Peters CM, Shiozawa Y. Skeletal complications in cancer patients with bone metastases. *International Journal of Urology*. 2016 Oct;23(10):825-32.

Valderrábano RJ, Linares MI. Diabetes mellitus and bone health: epidemiology, etiology and implications for fracture risk stratification. *Clinical diabetes and endocrinology*. 2018 Dec;4:1-8.

Veronese N, Stubbs B, Crepaldi G, Solmi M, Cooper C, Harvey NC, Reginster JY, Rizzoli R, Civitelli R, Schofield P, Maggi S. Relationship between low bone mineral density and fractures with incident cardiovascular disease: a systematic review and meta-analysis. *Journal of Bone and Mineral Research*. 2017 May;32(5):1126-35.

Vestergaard P, Rejnmark L, Mosekilde L. Fracture risk in patients with different types of cancer. *Acta Oncologica*. 2009 Jan 1;48(1):105-15.

Wang HK, Hung CM, Lin SH, Tai YC, Lu K, Liliang PC, Lin CW, Lee YC, Fang PH, Chang LC, Li YC. Increased risk of hip fractures in patients with dementia: a nationwide population-based study. *BMC neurology*. 2014 Dec;14(1):1-8.

Xue AL, Wu SY, Jiang L, Feng AM, Guo HF, Zhao P. Bone fracture risk in patients with rheumatoid arthritis: a meta-analysis. *Medicine*. 2017 Sep;96(36).

Appendix

Figure A1. Common support graph in ADRD analytical sample

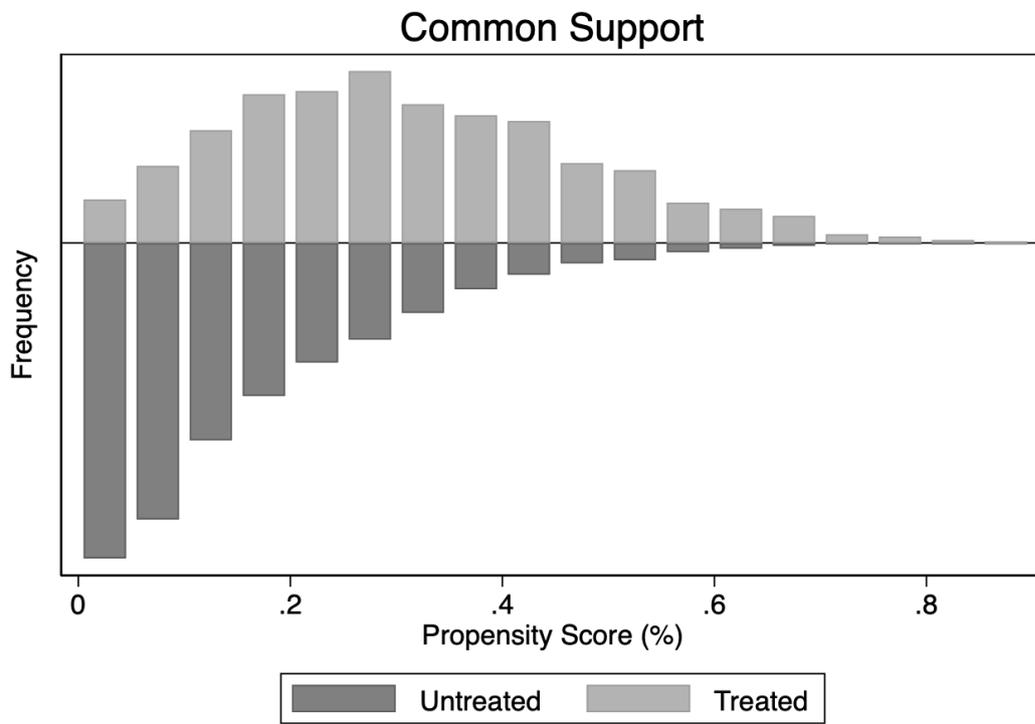
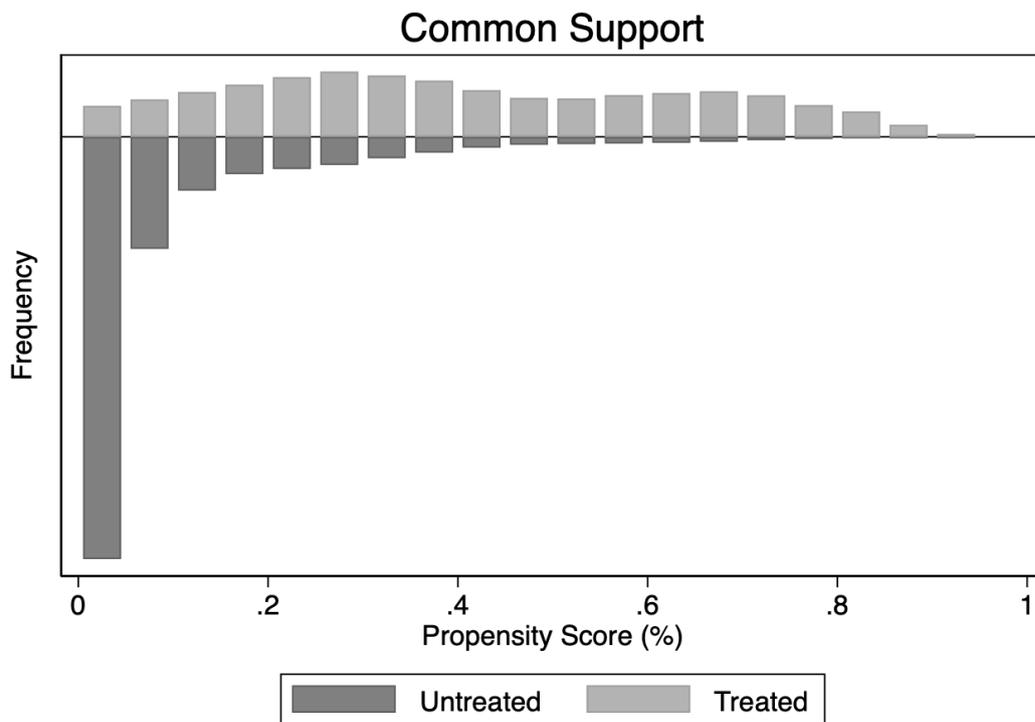


Figure A2. Common support graph in non-ADRD analytical sample



PAPER 3

Interpretable machine learning in Alzheimer's disease: a Forward Sparse-Group Lasso model

Abstract

Stochastic classification problems in Alzheimer's disease can be used to identify biomarkers of cognitive decline within a set of labelled brain regions of interest (ROIs). Using normalized fMRI data, we propose a novel classification algorithm designed as a linear combination of Sparse-Group Lasso and AUC-based forward selection models. The proposed approach, deemed forward sparse group lasso (FSGL), combines regularization methods with forward selection on a validation set, and uses set theory on joint distributions to identify predictors of disease staging. Statistical tests suggest that our FSGL algorithm improves prediction accuracy over the naïve SGL algorithm implementation. We demonstrated that in this application FSGL can achieve a significantly better performance than SGL. Consistent with previous literature, we further demonstrate that this approach could identify right hippocampal subiculum and presubiculum to be the most significant unique predictors of severe Alzheimer's disease progression.

Keywords: regularization, forward selection, Alzheimer's disease, hippocampal asymmetry

Introduction

Appropriate disease staging is crucial to inform timely therapy and adequate patient management in Alzheimer's disease. Identifying the most relevant disease progression predictors in cross-sectional data can improve diagnostic accuracy in case of high censoring or short observation period (Therriault et al., 2022). When machine learning is employed in the context of clinical decision making, interpretable algorithms are essential to discern crucial predictors, aiding comprehension of influential factors driving model decisions.

Interpretable machine learning holds a paramount significance in modern data-driven domains due to its pivotal role in deciphering the underlying insights within complex predictive models. In the realm of data science, where intricate algorithms are employed to analyse and predict outcomes, the ability to interpret these models is crucial for understanding the factors driving predictions. By unveiling the "black box" nature of advanced algorithms, interpretable machine learning provides valuable insights into the relationships between variables and the decision-making process of the model (Molnar, 2020). This transparency not only enhances our understanding of the model's inner workings but also fosters trust and acceptance among stakeholders, facilitating informed decision-making in clinical settings.

At the heart of interpretable machine learning lies the imperative to identify significant predictors accurately. In a multitude of applications, the goal is not solely to make predictions, but also to understand which variables contribute the most to those predictions. In the context of healthcare, pinpointing the influential factors enables clinicians to focus their attention on specific variables that have the most impact (Ahmad et al., 2018). This knowledge aids in allocating resources efficiently, designing targeted interventions, and tailoring strategies that leverage the most salient features. Without interpretability, the risk of relying on arbitrary or irrelevant predictors could lead to misguided decisions that fall short of capturing the true essence of the problem at hand.

Dimensionality reduction plays a significant role in enhancing the interpretability of machine learning models. In complex datasets with numerous features, it can be challenging to grasp the relationships and influences driving predictions. By reducing the number of features, these methods create a simplified representation that is easier to visualize and comprehend,

thus aiding in the interpretation of model behavior. Dimensionality reduction and features selection have been widely employed in several disease areas to predict likelihood of disease or disease progression. Li et al. (2016) illustrated an approach to signal processing and feature selection for atrial fibrillation detection in a noisy environment. Du et al. (2022) developed a feature selection model to assist clinicians in screening at risk patients who may benefit from early gestational diabetes prevention strategies. Jain et al. (2018) presented a comprehensive overview of various feature selection methods for chronic disease prediction. Polat et al. (2017) assessed wrapper and filter approaches for feature selection and dimensionality reduction in a chronic kidney disease dataset. Dimensionality reduction and features selection have found several other applications, including in the domains of genomics (Hauskrecht et al., 2007; Xing et al., 2001) and brain segmentation and classification (Kong et al., 2014; Zhang et al., 2011).

Several methods have been proposed to select the most important features in disease classification problems. Of these, regularization methods have gained popularity as some of the most reliable methods for features selection. Several regularization techniques have been proposed. These methods are commonly used in machine learning and statistics to handle high-dimensional data and perform feature selection by introducing regularization penalties. They are particularly useful when dealing with situations where the number of features is large compared to the number of observations.

Lasso stands for "Least Absolute Shrinkage and Selection Operator." It is a regularization technique that adds the absolute values of the coefficients as a penalty term to the loss function during model training (Tibshirani, 1996). In standard linear regression, the goal is to find the coefficients that minimize the sum of squared differences between predicted and actual values. However, when there are correlated predictors, the estimated coefficients can become sensitive to small changes in the data, leading to unstable and inaccurate predictions. Lasso regression introduces a regularization term to the traditional linear regression objective function. L1 regularization adds the absolute values of the coefficients to the loss function, encouraging sparsity in the feature selection, effectively shrinking some coefficients to exactly zero. Lasso is particularly useful for feature selection because it tends to select a subset of the most important features, making the resulting model more interpretable and reducing the risk of overfitting. (1)

$$\min_{\vec{\beta}} \|\vec{y} - X\vec{\beta}\|_2^2 + \lambda \|\vec{\beta}\|_1 \quad (1)$$

Where

$$\|\vec{\beta}\|_1 = |\beta_0| + |\beta_1|$$

Ridge regression, also known as Tikhonov regularization or L2 regularization, is a linear regression technique designed to address multicollinearity or high correlation between predictor variables, and prevent overfitting in predictive models (Hoerl et al., 2000). The regularization term is proportional to the square of the magnitude of the coefficients, effectively constraining their values to smaller but non-zero coefficients. By doing so, ridge regression encourages the model to not only fit the data but also to keep the coefficients small. This has the effect of shrinking the coefficients of less influential predictors towards zero, which can help mitigate the impact of multicollinearity and reduce model complexity. (2)

$$\min_{\vec{\beta}} \|\vec{y} - X\vec{\beta}\|_2^2 + \lambda \|\vec{\beta}\|_2^2 \quad (2)$$

Where

$$\|\vec{\beta}\|_2 = \sqrt{\beta_0^2 + \beta_1^2}$$

Elastic-Net combines both L1 (Lasso) and L2 (Ridge) regularization. It addresses some of the limitations of Lasso, such as selecting only one variable among highly correlated variables (Zou et al., 2005). Elastic-Net allows for both feature selection (sparse solutions) and handling multicollinearity by adding a combination of L1 and L2 penalties to the loss function. The balance between L1 and L2 penalties is controlled by a parameter, allowing for flexible tuning. (3)

$$\min_{\vec{\beta}} \|\vec{y} - X\vec{\beta}\|_2^2 + \lambda [(1 - \alpha) \|\vec{\beta}\|_2^2 + \alpha \|\vec{\beta}\|_1] \quad (3)$$

Group Lasso extends Lasso to incorporate grouping information among features. Instead of penalizing individual coefficients, it penalizes entire groups of coefficients together (Yuan et al, 2006). This is particularly useful when features are naturally grouped, such as in genomics, where genes often belong to the same pathways or biological processes. Group

Lasso encourages feature selection at the group level, effectively selecting entire groups of features, which can help capture higher-level relationships in the data. (4)

$$\min_{\vec{\beta}} \left\| \vec{y} - \sum_{l=1}^m X^{(l)} \vec{\beta}^{(l)} \right\|_2^2 + \lambda \sum_{l=1}^m \sqrt{p_l} \left\| \vec{\beta}^{(l)} \right\|_2 \quad (4)$$

Sparse Group Lasso combines the principles of both Group Lasso and Lasso. It extends the concept of group sparsity to include individual feature sparsity within each group (Simon et al., 2013). This allows for simultaneous selection of relevant groups and individual features within those groups. Sparse Group Lasso is beneficial when you want to capture both the overall structure of groups and the specific importance of individual features. (5)

$$\min_{\vec{\beta}} \left\| \vec{y} - \sum_{l=1}^m X^{(l)} \vec{\beta}^{(l)} \right\|_2^2 + (1 - \alpha) \lambda \sum_{l=1}^m \sqrt{p_l} \left\| \vec{\beta}^{(l)} \right\|_2 + \alpha \lambda \left\| \vec{\beta}^{(l)} \right\|_1 \quad (5)$$

Forward selection is a feature selection technique commonly used in machine learning and statistics. It involves building a predictive model iteratively by adding one feature at a time from a pool of available features (Sutter et al., 1993). The primary objective of forward selection is to incrementally identify the most relevant and informative features that contribute to the predictive power of the model while aiming to improve model performance.

The main steps of forward selection can be described as follows. (i) Initialization: the process starts with an empty set of selected features. (ii) candidate feature selection: all available features that have not yet been selected are considered as candidates. The algorithm evaluates the potential of each candidate feature to improve the model's performance. (iii) Feature evaluation: for each candidate feature, a model is trained using the currently selected features along with the candidate feature. The model's performance is evaluated using a predefined metric such as accuracy, mean squared error, or another appropriate metric depending on the problem. (iv) Feature selection: the candidate feature that leads to the most significant improvement in model performance is selected and added to the set of selected features. This step is crucial, as it focuses on identifying features that contribute the most to the model's predictive power. (v) Iteration: steps 3 and 4 are repeated iteratively, with the algorithm evaluating the remaining candidate features and selecting the one that provides the most improvement in performance. This process continues until a stopping

criterion is met, such as reaching a predefined number of selected features or observing a decline in performance improvement. (vi) Final model: once the iteration is complete, the selected set of features forms the final feature subset for the model. This subset is used to build the final predictive model, which can be evaluated on unseen data.

In this paper, we propose a method to identify clinically relevant features for Alzheimer's disease staging using pairwise classification models on disjoint cross-sectional sets and we provide clinical interpretation and validation. The objective of this work is to describe a method to retain the minimum number of clinically relevant Alzheimer's disease predictors while preserving model accuracy. The method we propose, deemed forward sparse-group lasso (FSGL), is a linear combination of sparse-group lasso and forward selection, and it makes use of an external validation set in the iteration algorithm. We demonstrate that FSGL is superior to the naïve sparse-group lasso (SGL) implementation in this application.

Univariate feature selection has several limitations when used on its own. In this work, however, we show its potential when used in combination with state-of-the-art regularization methods and an external validation set. We discuss the findings and provide an interpretation of the underlying biological mechanisms. Finally, we highlight the limitations of the proposed approach in the context of the identification of significant ROIs in Alzheimer's disease progression.

Methods

Sparse group lasso inherently encourages sparsity in feature selection, but combining it with a forward selection step allows for a more granular selection of features. This stepwise approach can help fine-tune and select specific subsets of features, potentially leading to a more interpretable final model.

In Sparse Group Lasso, if correlated features belong to the same group, the penalty imposed by the group lasso encourages sparsity within these groups, effectively selecting only a subset of features from each group. By penalizing entire groups, it indirectly addresses correlation within those groups. However, there is no way to instruct SGL to select a certain feature over another highly correlated feature. The process of selecting, within a set of

highly correlated features, the most important ones to be retained in the model is inherently random and can possibly exclude relevant features for clinical evaluation and diagnosis.

The method we propose is a linear combination of sparse-group lasso (SGL) and a univariate forward selection loop on an external validation set, deemed forward sparse-group lasso (FSGL). This method allows to restore potential clinically relevant features excluded by sparsity in the SGL selection process.

The minimization problem is the one of a naïve SGL with a further condition. The set of selected features F in the FSGL model must include r rows and q^* columns, whereas the naïve SGL implementation includes the same rows and q columns. The FSGL set of features F^{FSGL} is a subset of the SGL feature set F^{SGL} .

The F^{FSGL} features set is initialized to be identical to the F^{SGL} features set and the iteration counter k is initialized to be equal to 0. At every new iteration, an additional feature q^+ is added until the value of the object function J becomes lower than or equal to the value of the same function in the previous iteration. (6)

$$\min_{\vec{\beta}} \left\| \vec{y} - \sum_{l=1}^m X^{(l)} \vec{\beta}^{(l)} \right\|_2^2 + (1 - \alpha) \lambda - \sum_{l=1}^m \sqrt{p_l} \|\vec{\beta}^{(l)}\|_2 + \alpha \lambda \|\vec{\beta}^{(l)}\|_1$$

$$s. t. \quad F_{r \times q^*}^{FSGL}, F_{r \times q}^{SGL}, q \in q^*$$

Where:

$$F_{r \times q^*}^{FSGL} \equiv \left[\begin{array}{l} \text{Initiate } F_0 = F_{r \times q}^{SGL}; k = 0 \\ \quad q^+ = \arg \max_{q \in F_k} [J(F_k + q)] \\ \\ \text{where:} \\ \quad J(F_k + q) = \arg \min \left(\sum_{i=1}^n \log(1 + \exp(-Y_i W^T X_i)) \right) \\ \\ \text{until } J(F_k + q) \leq J(F_k): \\ \quad \text{update } F_{k+1} = F_k + q^+; \\ \quad k = k + 1 \end{array} \right]$$

(6)

We chose logistic regression as a putative classifier in the base model to avoid possible confounding effects due to hyperparameters settings, which might be present in other classifiers (e.g. SVM). To test the value of the object function in iteration $k + 1$ against the

same in iteration k , we use the area under the precision-recall curve (AUC-PR), because it is an appropriate metric to identify relevant features in that it is insensitive to class imbalance (Saito et al., 2015).

The forward sparse group lasso algorithm, therefore, is a procedure that aims to identify the locally optimal selection of features q^* including the subset of sparse group lasso features, plus additional features that marginally improve the AUC metric. The dataset is split into train, validation and test. Then, each of the components is normalized by subtracting the mean and dividing by the standard deviation. The naïve results are calculated by fitting the model using all available features. Iterations are run over the sparse group lasso (SGL) penalty parameter and the SGL algorithm is used for features selection. The best SGL selection of features is identified using a validation set. Results for SGL are calculated on the test set. Finally, additional features are iteratively added on to the SGL selection until the AUC value in a given iteration k is higher than the AUC value in the previous iteration $k-1$ in the validation set. Results for the forward sparse group lasso (FSGL) are calculated on the test set using the identified selection of features. Algorithm 1 describes the procedure to obtain the locally optimal set of features denoted as $F_{r \times q}^{FSGL}$.

Algorithm 1 Forward Sparse-Group Lasso Algorithm

```

1: procedure  $F_{r \times q}^{FSGL}$ 
2:   for random_state in random(1, 1000000, size = 50000) {
3:     # dataset split: train/validation/test
4:     x_train, x_validation, x_test = split(0.8, 0.1, 0.1, random_state)
5:     # normalize data
6:     subtract.mean(train, validation, test)
7:     divide.std(train, validation, test)
8:     # calculate naïve results (training vs test, all features)
9:     model.fit(x_train, y_train)
10:    model.predict(x_test)  $\rightarrow$  Naïve_AUC (test set)
11:    # select SGL features
12:    SGL_AUC_list = []
13:    for penalty in random(0, 1, size = 100) {
14:      x_train_SGL_selected = SGL(x_train, penalty)
15:      model.fit(x_train_SGL_selected, y_train)
16:      model.predict(x_validation)  $\rightarrow$  SGL_AUC (validation set)
17:      SGL_AUC_list = SGL_AUC_list.append(SGL_AUC)
18:    }
19:    sorted_SGL_AUC_list = sort(SGL_AUC_list, ascending = FALSE)
20:     $F_k = x\_train\_SGL\_selected.sorted\_SGL\_AUC\_list[[1]]$ 
21:    # calculate SGL results (training vs test, SGL features)
22:    model.fit( $F_k$ , y_train)
23:    model.predict(x_test)  $\rightarrow$  SGL_AUC (test set)
24:    # forward selection loop
25:    FSGL_AUC_list = []
26:    k = 0
27:    while  $F_k.AUC > F_{k-1}.AUC$ 
28:      for each  $q$  in colnames(x_train) if  $q$  not in colnames( $F_k$ ) {
29:         $F_k = F_k.append(q)$ 

```

```

30:     model.fit(Fk, y_train)
31:     model.predict(x_validation) → FSGL_AUCk (validation set)
32:     FSGL_AUC_list = FSGL_AUC_list.append(FSGL_AUC)
33: }
34: sorted_FSGL_AUC_list = sort(FSGL_AUC_list, ascending = FALSE)
35: Fk = Fk.sorted_FSGL_AUC_list[[1]]
36: if FSGL_AUCk < FSGL_AUCk-1 {
37:     return(Fk)
38: }
39: k = k + 1
40: # calculate FSGL results (training vs test, FSGL features)
41: model.fit(Fk, y_train)
42: model.predict(x_test) → FSGL_AUC (test set)
43: Fr×qFSGL = Fk
44: }
45: end of procedure
46: return(Fr×qFSGL)

```

Data

We used the UCSF - Cross-Sectional FreeSurfer (5.1) dataset available from ADNI. The dataset consists of 4,896 observations and 341 brain imaging features from ADNI1, ADNI GO and ADNI2. We restricted our sample to observations from ADNI2 only, obtaining a sample of 4,202 observations. Next, we dropped 269 observations that failed the overall quality check, resulting in 3,933 observations. We then merged the dataset with the ADNIMERGE dataset to obtain the current diagnosis at the time of the brain scan. We dropped 837 observations that could not be matched or that resulted in missing diagnosis, obtaining a sample of 3,096 observations. We sorted the observations by patient ID and exam date, and kept only the first observation for each subject, which resulted in 935 unique subjects. We dropped 5 subjects that were missing multiple variables. Finally, we dropped 11 variables that were missing multiple subjects. Our final dataset is comprised of 930 unique subjects and 330 brain imaging features and 1 labels outcome variable. The diagnoses were as follows: 319 cognitively normal (CN), 459 mild cognitive impairment (MCI) and 152 dementia (AD).

All 330 brain imaging features were compartmented into six groups, selected based on ROIs anatomical regions on the UCSFFSX51_DICT_08_01_14 data dictionary from ADNI: frontal lobe, temporal lobe, parietal lobe, occipital lobe, limbic system and subcortical areas, insular cortex.

We then split the dataset in two parts, holding 20% of observations from each class in a separate dataset used for external validation. The final dataset consists of 255 CN, 367 MCI and 122 AD subjects. The external validation dataset includes 64 CN, 92 MCI and 30 AD subjects.

Table 1 describes the analytical sample. Average age was 73.04 in the CN sample, 71.90 in the MCI sample and 74.07 in the AD sample. Female patients were 54.71% in the CN sample, 43.42 in the MCI sample and 38.09 in the AD sample. In the CN sample, 6.28% of the patients were Black or African American, 2.19% in the MCI sample and 2.72% in the AD sample. Those who had never gotten married were 5.66% in the CN sample, 3.07% in the MCI sample and 0.68% in the AD sample. Patients in the CN sample had 16.58 years of education, those in the MCI sample 16.18 years and those in the AD sample 15.75 years. The ADAS13 average score was 8.83 in the CN sample, 14.76 in the MCI sample and 30.01 in the AD sample. The number of patients with one APOE ϵ 4 alleles were 27.35% in the CN sample, 36.84% in the MCI sample and 49.65% in the AD sample, while the number of patients with two APOE ϵ 4 alleles were 2.20% in the CN sample, 9.86% in the MCI sample and 20.40% in the AD sample.

Table 1. Analytical sample descriptive statistics

	Diagnosis		
	CN	MCI	AD
Age	73.04	71.90	74.07
Female	0.5471	0.4342	0.3809
Black or African American	0.0628	0.0219	0.0272
Never married	0.0566	0.0307	0.0068
Education (years)	16.58	16.18	15.75
ADAS13 score	8.83	14.76	30.01
Number of APOE ϵ 4 alleles:			
Zero	0.7044	0.5328	0.2993
One	0.2735	0.3684	0.4965
Two	0.0220	0.0986	0.2040
N	319	459	152

Results

Table 1 presents the mean number of features and AUC over 50,000 model iterations for three distinct pairwise comparisons. In the HC vs AD comparison, the SGL algorithm identified 26.05 (± 16.47) significant features achieving an average AUC of 93.06% ($\pm 4.07\%$), while the FSGL model identified 28.08 (± 16.79) significant features, achieving an average AUC equal to 93.38% ($\pm 3.97\%$). In the HC vs MCI comparison, the SGL algorithm identified 25.97 (± 26.35) significant features achieving an average AUC of 63.92% ($\pm 6.31\%$), while the FSGL model identified 29.89 (± 26.88) significant features, achieving an average AUC equal to 64.10% ($\pm 6.84\%$). In the MCI vs AD comparison, the SGL algorithm identified 24.89 (± 21.86) significant features achieving an average AUC of 79.82% ($\pm 6.41\%$), while the FSGL model identified 28.68 (± 22.36) significant features, achieving an average AUC equal to 80.06% ($\pm 5.87\%$). On a Welch's t-test, all three models had a p-value lower than 0.0001, rejecting the null hypothesis of FSGL AUC mean being equal or lower than SGL AUC mean.

In the HC vs AD classification algorithm, the most frequently added features in the FSGL model were right parahippocampal surface area, right parahippocampal cortical volume and right paracentral thickness average. In the HC vs MCI classification algorithm, the most frequently added features in the FSGL model were third ventricle subcortical volume, right choroid plexus subcortical volume and right superior parietal cortical volume. Finally, in the MCI vs AD classification algorithm, the most frequently added features in the FSGL was right parahippocampal cortical volume, right cerebellum subcortical volume and left transverse temporal thickness average.

Table 1. Mean number of features, AUC, standard deviation, and t-test results of FSGL and reference models

	HC vs AD			HC vs MCI			MCI vs AD		
	NAÏVE	SGL	FSGL	NAÏVE	SGL	FSGL	NAÏVE	SGL	FSGL
Number of Features (std. dev.)	330 (0.00)	26.05 (16.47)	28.08 (16.79)	330 (0.00)	25.97 (26.35)	29.89 (26.88)	330 (0.00)	24.89 (21.86)	28.68 (22.36)
AUC	0.9401 (0.0376)	0.9306 (0.0407)	0.9338 (0.0397)	0.6045 (0.0727)	0.6392 (0.0631)	0.6410 (0.0684)	0.8110 (0.0657)	0.7982 (0.0641)	0.8006 (0.0587)
	p<0.0001			p<0.0001			p<0.0001		
Most frequently added feature	'ST103SA' (4.76%)			'ST127SV' (2.52%)			'ST103CV' (2.51%)		
Second most frequently added feature	'ST103CV' (2.88%)			'ST80SV' (1.92%)			'ST77SV' (2.14%)		
Third most frequently added feature	'ST102TA' (2.80%)			'ST116CV' (1.67%)			'ST62TA' (2.06%)		

Notes: AUC=Area under the ROC curve; standard deviation in brackets. Statistical t-tests conducted on 50,000 iterations results. Frequency of additional features utilization in FSGL model in brackets. 'ST103SA' - Surface Area (aparc.stats) of RightParahippocampal, 'ST103CV' - Cortical Volume (aparc.stats) of RightParahippocampal, 'ST102TA' - Thickness Average (aparc.stats) of RightParacentral, 'ST127SV' - Subcortical Volume (aseg.stats) of ThirdVentricle, 'ST80SV' - Subcortical Volume (aseg.stats) of RightChoroidPlexus, 'ST116CV' - Cortical Volume (aparc.stats) of RightSuperiorParietal, 'ST77SV' - Subcortical Volume (aseg.stats) of RightCerebellum, 'ST62TA' - Thickness Average (aparc.stats) of LeftTransverseTemporal

Table 2 shows the probability of ranking among the top three most important features in the HC vs AD classification over 50,000 runs of the model. Left presubiculum ranked first in 47.84% of cases, second in 48.24% of cases and third in 3.92% of occurrences. Left subiculum ranked third in 12.24% of cases. Right presubiculum ranked second in 0.96% of cases and third in 10.00% of cases. Left hippocampus ranked first in 2.96% of cases, second in 15.92% of cases and third in 64.00% of cases. Right hippocampus ranked third in 2.96% of times. Right entorhinal ranked first in 49.20% of times, second in 34.88% of times, and third in 6.88% of times.

Table 2. Probability (%) of ranking among the top three most important features in FSGL HC vs AD

ROIs	classification model		
	HC vs AD		
	First	Second	Third
ST136HS Presubiculum L	47.84	48.24	3.92
ST137HS Subiculum L	-	-	12.24
ST144HS Presubiculum R	-	0.96	10.00
ST29SV Hippocampus L	2.96	15.92	64.00
ST88SV Hippocampus R	-	-	2.96
ST83TA Entorhinal R	49.20	34.88	6.88
TOTAL	100.00	100.00	100.00

Table 3 shows the probability of ranking among the top three most important features in the HC vs MCI classification over 50,000 runs of the model. Left lateral occipital ranked second 0.76% of cases, left superior frontal ranked third in 2.48% of cases, right lateral occipital ranked third in 0.76% of cases, right superior parietal ranked third in 1.24% of cases, left presubiculum ranked third in 2.76% of cases. Left subiculum ranked first in 12.76% of cases, second in 42.48% and third in 15.80% of cases. Left tail ranked first in 1.24% of cases. Left posterior cingulate ranked third in 1.24% of cases. Right precuneus ranked third in 1.24% of cases. Left hippocampus ranked first in 56.48% of cases, second in 31.72% of cases, and third in 11.80% of cases. Left inferior lateral ventricle ranked third in 0.76% of cases. Right hippocampus ranked third in 4.00% of cases. Right inferior lateral ventricle ranked first in 23.04% of cases, second in 11.04% of cases and third in 33.52% of cases. Right thalamus ranked second in 1.24% of cases. Left postcentral ranked third in 0.76% of cases. Right entorhinal ranked first in 6.48% of cases, second in 12.76% of cases and third in 14.96% of cases.

Table 3. Probability (%) of ranking among the top three most important features in FSGL HC vs MCI

ROIs	classification model		
	HC vs MCI		
	First	Second	Third
ST35CV LateralOccipital L	-	0.76	-
ST56CV SuperiorFrontal L	-	-	2.48
ST94CV LateralOccipital R	-	-	0.76
ST116CV SuperiorParietal R	-	-	1.24
ST136HS Presubiculum L	-	-	2.76
ST137HS Subiculum L	12.76	42.48	15.80
ST138HS Tail L	1.24	-	8.68
ST50SA PosteriorCingulate L	-	-	1.24
ST111SA Precuneus R	-	-	1.24
ST29SV Hippocampus L	56.48	31.72	11.80
ST30SV InfLateralVentricle L	-	-	0.76
ST88SV Hippocampus R	-	-	4.00
ST89SV InfLateralVentricle R	23.04	11.04	33.52
ST120SV Thalamus R	-	1.24	-
ST49TA Postcentral L	-	-	0.76
ST83TA Entorhinal R	6.48	12.76	14.96
TOTAL	100.00	100.00	100.00

Table 4 shows the probability of ranking among the top three most important features in the MCI vs AD classification over 50,000 runs of the model. Left presubiculum ranked first in 49.16% of runs, second in 33.36% of runs and third in 14.64% of runs. Right presubiculum ranked first in 25.36% of cases, second in 25.80% of cases, third in 8.84% of cases. Right subiculum ranked third in 1.68% of cases. Left amygdala ranked second in 2.00% of cases and third in 6.52% of cases. Right cerebellum ranked third in 2.32% of cases. Right temporal pole ranked third in 2.00% of cases. Left fusiform ranked second in 1.16% of cases. Left middle temporal ranked first in 9.16% of cases, second in 8.20% of cases, third in 22.32% of cases. Right entorhinal ranked first in 16.32% of cases, second in 28.32% of cases, and third in 40.84% of cases. Right inferior parietal ranked second in 1.16% of cases. Left banks of the superior temporal sulcus ranked third in 0.84% of cases.

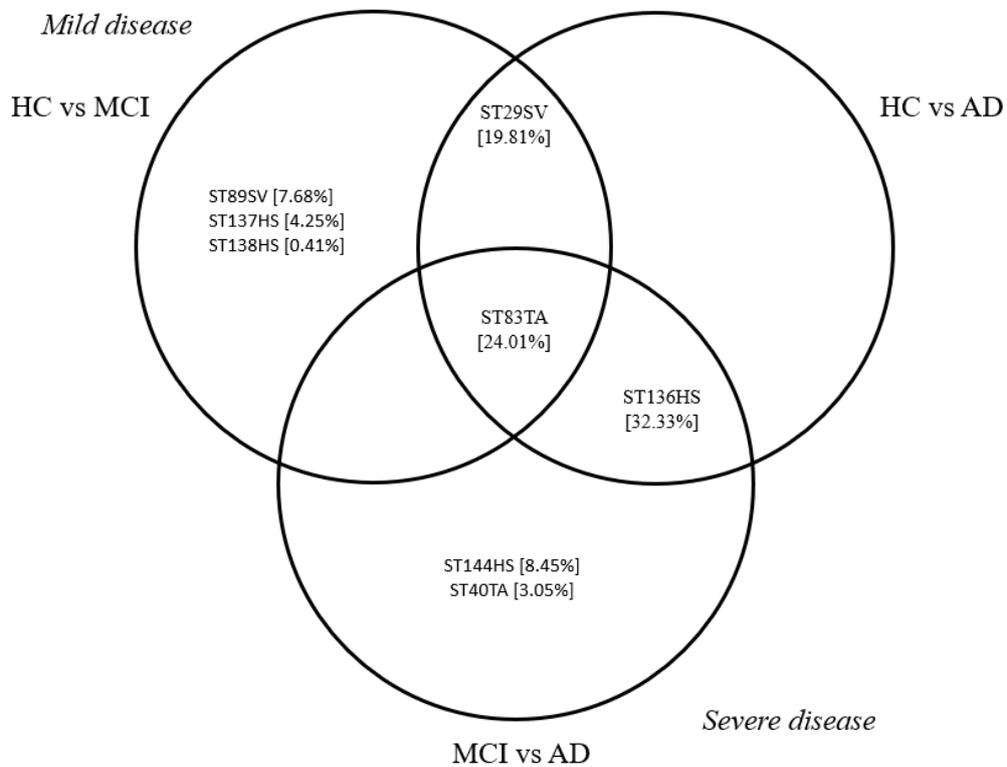
Table 4. Probability (%) of ranking among the top three most important features in FSGI MCI vs AD

ROIs	MCI vs AD		
	First	Second	Third
ST136HS Presubiculum L	49.16	33.36	14.64
ST144HS Presubiculum R	25.36	25.80	8.84
ST145HS Subiculum R	-	-	1.68
ST12SV Amygdala L	-	2.00	6.52
ST77SV Cerebellum R	-	-	2.32
ST119TA TemporalPole R	-	-	2.00
ST26TA Fusiform L	-	1.16	-
ST40TA MiddleTemporal L	9.16	8.20	22.32
ST83TA Entorhinal R	16.32	28.32	40.84
ST90TA InferiorParietal R	-	1.16	-
ST13TS Bankssts L	-	-	0.84
TOTAL	100.00	100.00	100.00

Figure 1 represents the joint probabilities of ranking first over 150,000 algorithm runs, in three pairwise comparisons. Left hippocampal subcortical volume (7.68%), hippocampal subfields volume of left subiculum (4.25%) and hippocampal subfields volume of left tail (0.41%) ranked first in the HC vs MCI classification model only. Subcortical volume of left hippocampus (19.81%) ranked first in the HC vs MCI and the HC vs AD models. Hippocampal subfields volume of right presubiculum (8.45%) and thickness average of left middle temporal (3.05%) ranked first in the MCI vs AD classification model only. Hippocampal subfields volume of left presubiculum (32.33%) ranked first in both MCI vs AD and HC vs AD set. Thickness average of right entorhinal (24.01%) ranked first in all three models. Therefore, the four predictors of mild disease are identified in the HC vs MCI set not included in the MCI vs AD set: left hippocampal subcortical volume, hippocampal subfields volume of left subiculum, hippocampal subfields volume of left tail and subcortical volume of left hippocampus. Similarly, the three predictors of severe disease are identified in the MCI vs AD set not included in the HC vs MCI set: hippocampal subfields volume of right presubiculum, thickness average of left middle temporal and hippocampal subfields volume of left presubiculum. Brain networks representing these relationships are represented in Figure 2: yellow dots represent predictors associated with milder disease, red dots represent predictors associates with more severe disease, whereas orange dots are shared nodes in the network. Figure S1-S6 in the appendix provide similar information in relation to the joint probabilities of ranking second and third in 150,000 model iterations: similar patterns are

highlighted, with right hippocampus volume variation being uniquely and significantly associated with more severe disease.

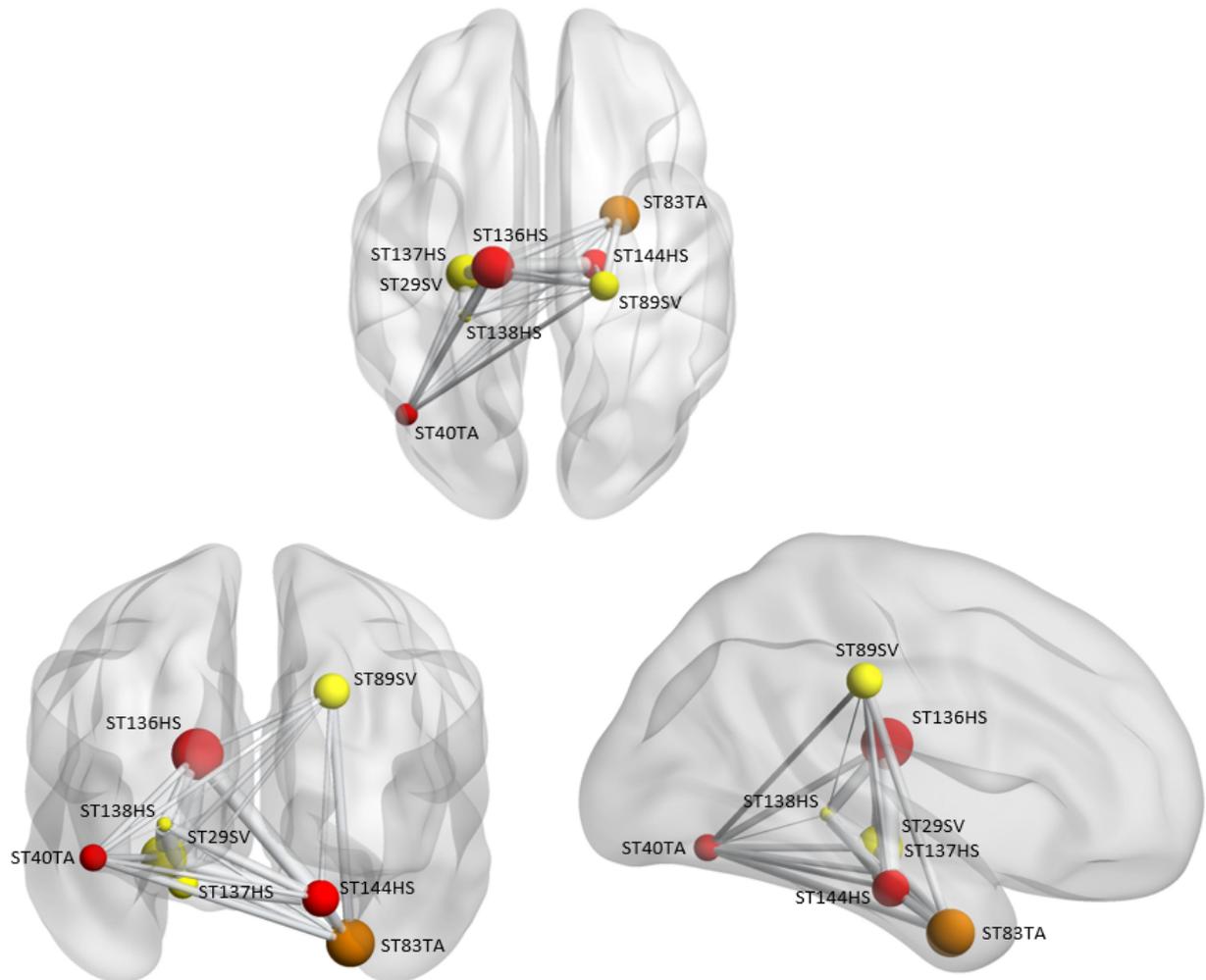
Figure 1. Joint probabilities of ranking first over 150,000 iterations



Notes:

ST29SV - Subcortical Volume (aseg.stats) of LeftHippocampus. -25.03 -20.74 -10.13 [Song, 2019]
 ST40TA - Thickness Average (aparc.stats) of LeftMiddleTemporal. -44.00, -67.00, -19.00 [Crone, 2011]
 ST83TA - Thickness Average (aparc.stats) of RightEntorhinal. 25.00, -2.00, -33.00 [Konishi, 2018]
 ST89SV - Subcortical Volume (aseg.stats) of RightInferiorLateralVentricle. 20, -25, 35 [Rorden, 2007 *]
 ST136HS - Hippocampal Subfields Volume of LeftPresubiculum. -16.00, -19.00, 18.50 [Palomero-Gallagher, 2020 *]
 ST137HS - Hippocampal Subfields Volume of LeftSubiculum. -20.00, -19.00, -17.00 [Palomero-Gallagher, 2020 *]
 ST138HS - Hippocampal Subfields Volume of LeftTail. -25.00, -35.00, -1.00 [Rorden, 2007 *]
 ST144HS - Hippocampal Subfields Volume of RightPresubiculum. 17.00, -18.00, -20.50 [Palomero-Gallagher, 2020]
 The asterisk "*" indicates further authors' elaboration on referenced publication.

Figure 2. First-ranking ROIs distribution in Axial, Sagittal and Coronal views



Notes. Clockwise from top to bottom left: Axial, Sagittal and Coronal views. The brain networks were generated using BrainNet Viewer (<http://www.nitrc.org/projects/bnv/>) (Xia et al., 2013). Red dots represent ROIs associated with more severe disease. Yellow dots represent ROIs associates with milder disease.

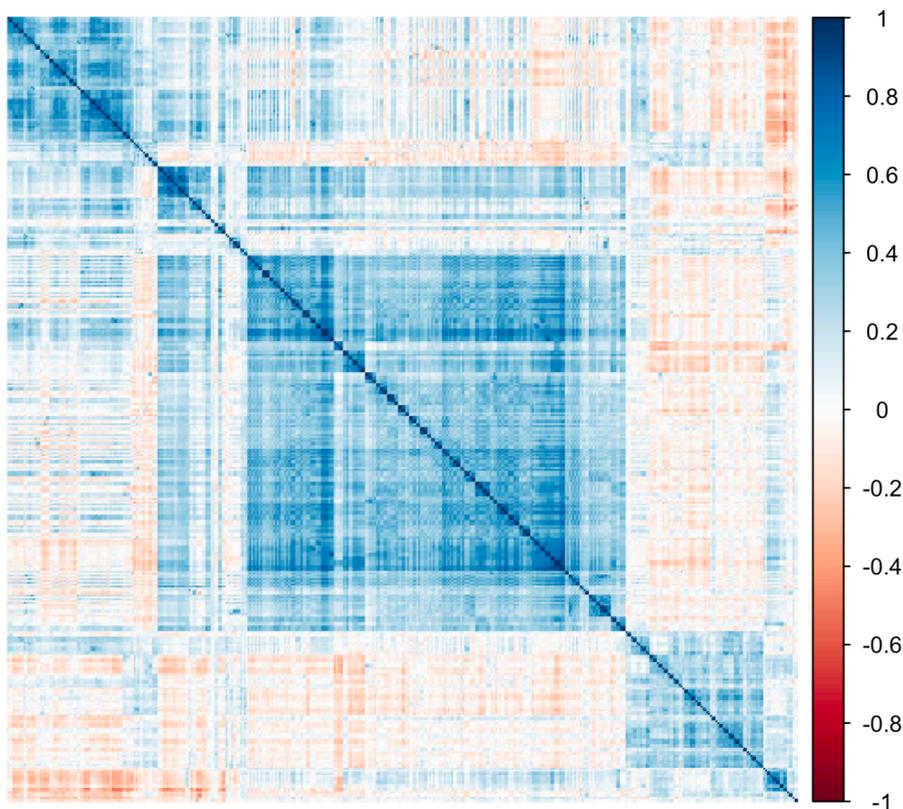
Simulated Data Experiment

Finally, a simulated experiment was executed to replicate the feature selection process within a dataset containing a blend of clinically significant variables exhibiting strong correlations with other statistically significant variables.

A dataset comprising 10000 observations and 400 variables was created. The variables were synthesized to emulate ADNI fMRI data ranges and distributions.

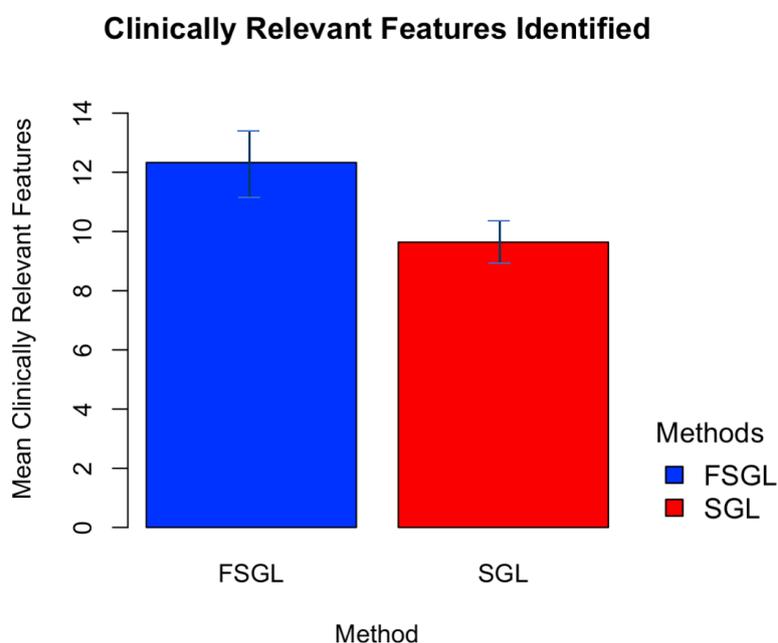
Within this dataset, 100 variables were chosen at random to have a highly correlated paired feature labelled as clinically significant. As a result, specific variables were identified as clinically significant, while the remaining variables retained their initial randomized nature. Figure 3 shows the correlation matrix of all 430 features. It displays the correlation coefficients between variables. Each cell in the matrix represents the correlation between two variables. The correlation coefficient quantifies the strength and direction of the linear relationship between two variables. It ranges between -1 and 1, where 1 indicates a perfect positive linear relationship (as one variable increases, the other also increases proportionally). -1 indicates a perfect negative linear relationship (as one variable increases, the other decreases proportionally). 0 indicates no linear relationship between the variables. The diagonal contains perfect correlations (always 1) since it compares the variables with themselves and with their 'clinically significant' copies.

Figure 3. Synthetic Data Correlation Matrix



SGL algorithm was applied to introduce sparsity and identify relevant features. Similarly, FSGL was used to identify the significant features, using SGL and a combination of univariate forward selection steps, as described above. Iterations were set to be equal to 1,000. Within the pool of selected features, those labelled 'clinically significant' were counted and standard error computed. SGL algorithm identified 9.64 (CI: 7.81 to 11.46) clinically relevant features while FSGL could identify 12.33 (CI: 10.74 to 13.91) clinically relevant features.

Figure 4. Clinically relevant features identified by FSGL and SGL



Discussion and Limitations

In this study, we introduced a novel algorithm, designed as a linear combination of the regularization method sparse group lasso (SGL) and a sequential forward selection step on an external validation set, and deemed it Forward Sparse Group Lasso (FSGL). We showed that this method improves the prediction accuracy of the model significantly across the three models considered. We then utilized the results of pairwise diagnoses classification models within a set theory framework to identify predictors that are exclusively associated with either milder or more severe disease. Further, we illustrated in a synthetic dataset the ability of FSGL to identify a greater number of clinically relevant features with respect to SGL.

This suggests that the approach we proposed is particularly useful when there exist clinically significant features that are highly correlated with other statistically significant features, and hence may be discarded by the SGL regularization algorithm. The introduction of a forward selection step addresses this issue by restoring clinically significant features in the pool of variables used for clinical assessment and disease staging.

In this sample of 930 ADNI participants, including 152 AD patients, 459 MCI and 319 cognitively normal individuals, left hippocampus atrophy was linked to both mild and severe cognitive decline: we identified both left and right hippocampi volume to be strong predictors of disease. However, right hippocampus atrophy was the most frequent unique predictor of progression from MCI to AD: we identified right hippocampus volume, in the subiculum and presubiculum subfields, to be uniquely associated with full-blown Alzheimer's disease. This is consistent with the clinical observation of the fact that deficits in verbal memory are usually seen in the MCI stage, associated with left hippocampal volume (Bonner-Jackson et al., 2015), whereas deficits in visual memory, associated with right hippocampal volume (Huang et al., 2022), usually present themselves at later stages.

These findings are supported by other peer-reviewed studies. In a stepwise discriminant analysis predicting progression to Alzheimer's disease, Galton et al. (2005) found that left hippocampus atrophy was associated with a 63.6% sensitivity while right hippocampus atrophy was associated with a 90.9% sensitivity. In a study on MCI to AD progression, Herukka (2008) found that the volume of right hippocampus exhibited sensitivity of 87.5%, while the volume of the left hippocampus had a sensitivity of 75.0%. In a magnetic resonance study analyzing MCI progression of 220 subjects, right presubiculum subfield volume was not significantly different between the negative controls and the non-converting MCI group, but it was significantly different between the negative controls and the converting MCI group (Guo et al., 2020).

In Bottino et al. (2002), left hippocampus volume could significantly differentiate between MCI and HC and also between MCI and AD; right hippocampus volume could identify AD versus MCI, but was not significant in predicting HC versus MCI. Bozzali et al. (2006) found a reduction in gray matter density in the right hippocampus, but not in the left hippocampus,

in MCI converters as opposed to nonconverters. Similarly, Chételat et al. (2005) found the right posterior hippocampus to be a significant predictor of MCI progressors versus MCI stable patients. Hämmäläinen et al. (2007) found the right hippocampus to be a predictor of conversion in MCI patients, but a similar pattern was not identified with respect to the left hippocampus. In Petrella et al. (2006), the authors found decreased activity in the left hippocampus of MCI participants as compared with HC, but not in the right hippocampus.

The proposed approach combines regularization methods with forward selection on a validation set, and uses set theory on joint distributions to identify predictors of disease staging. The proposed approach can be computationally efficient for small to medium-sized feature sets and can provide insights into the relative importance of different features. However, this approach also has several limitations: it might lead to suboptimal solutions if the order in which features are added affects the final selection. It does not consider the interactions between features, which can limit its effectiveness in capturing complex relationships.

Other limitations of this approach include the fact that FSGL requires more computational resources and it is slower than SGL, although this could become more sustainable with the increasing computational power and decreasing costs of HPC machines. Also, forward selection is best applied only when the initial number of features is reduced, e.g. using PCA or using ROIs instead of voxel-level brain data. Another important limitation resides in the fact that SGL, on which FSGL is built, randomly chooses which variable(s) to drop among a group of correlated variables, which could have potentially unpredictable implications for the clinical validity of the findings and the generalizability of the method in clinical settings. However, as a proof of concept, although computationally intensive, this method can achieve a better performance than the state-of-the-art approach.

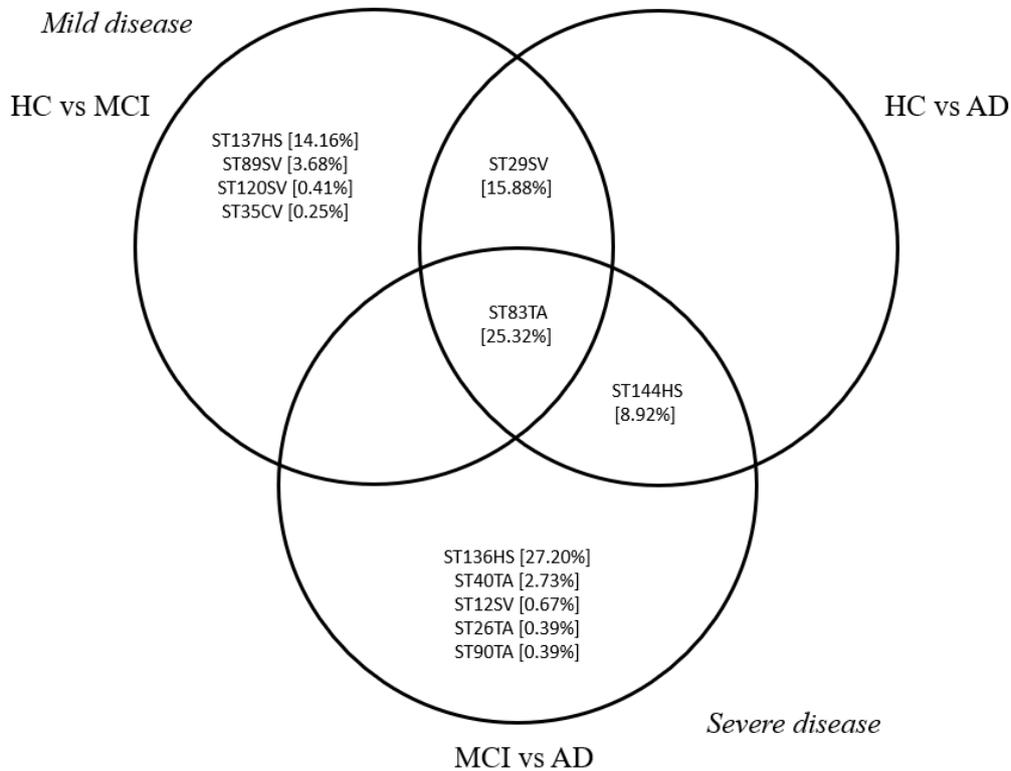
Conclusion

In this work, we have used a novel algorithm, deemed forward sparse group lasso (FSGL), as a linear combination of sparse group lasso (SGL) and forward selection on a validation set, to

then identify predictors of disease staging within a set theory framework. We demonstrated that in this application FSGL can achieve a significantly better performance than SGL. Consistent with previous literature, we further demonstrated that this approach was able to identify right hippocampal subiculum and presubiculum to be the most significant and unique predictors of severe Alzheimer's disease progression.

Appendix

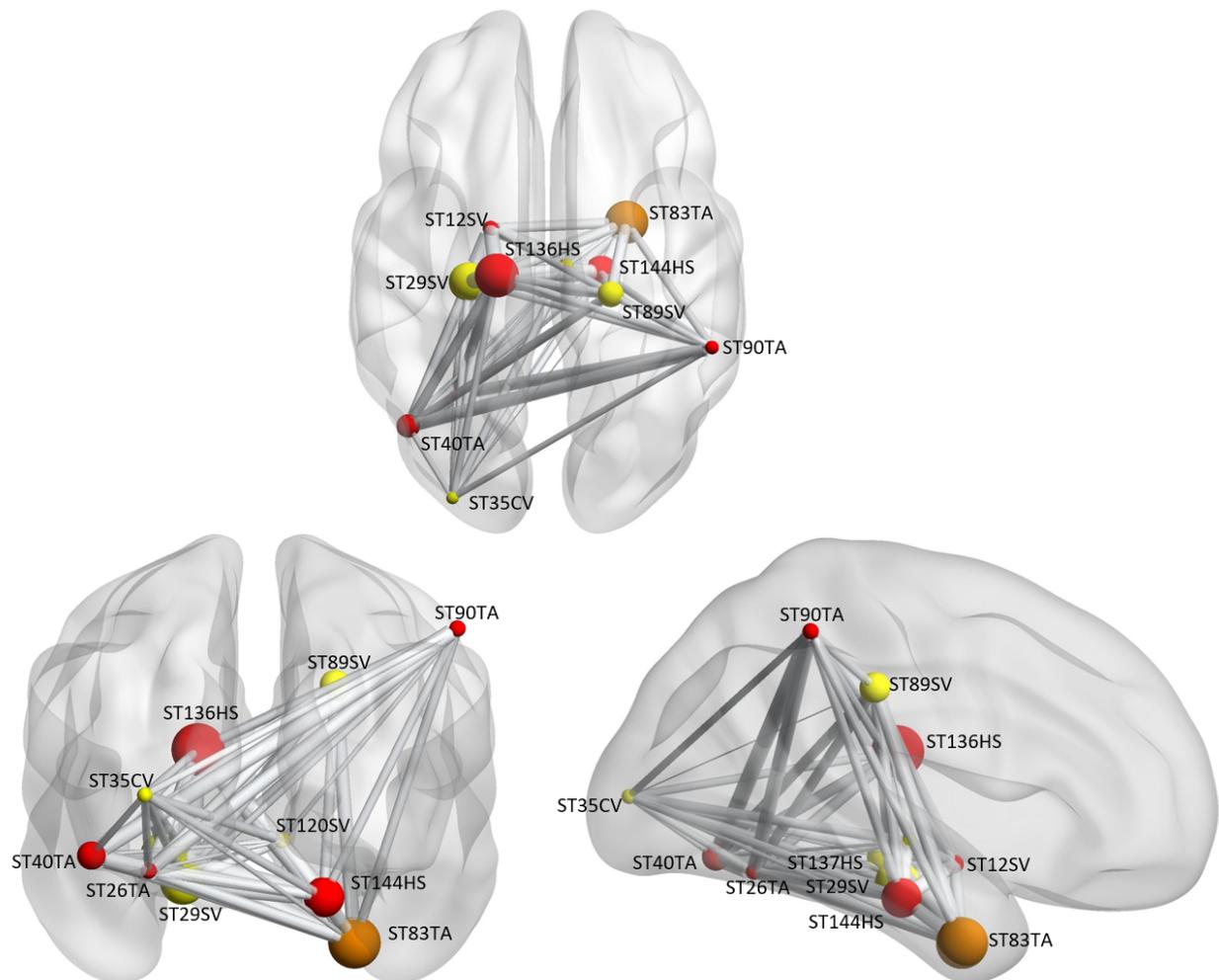
Figure S1. Most important features: probability of ranking second over 150,000 iterations



Notes:

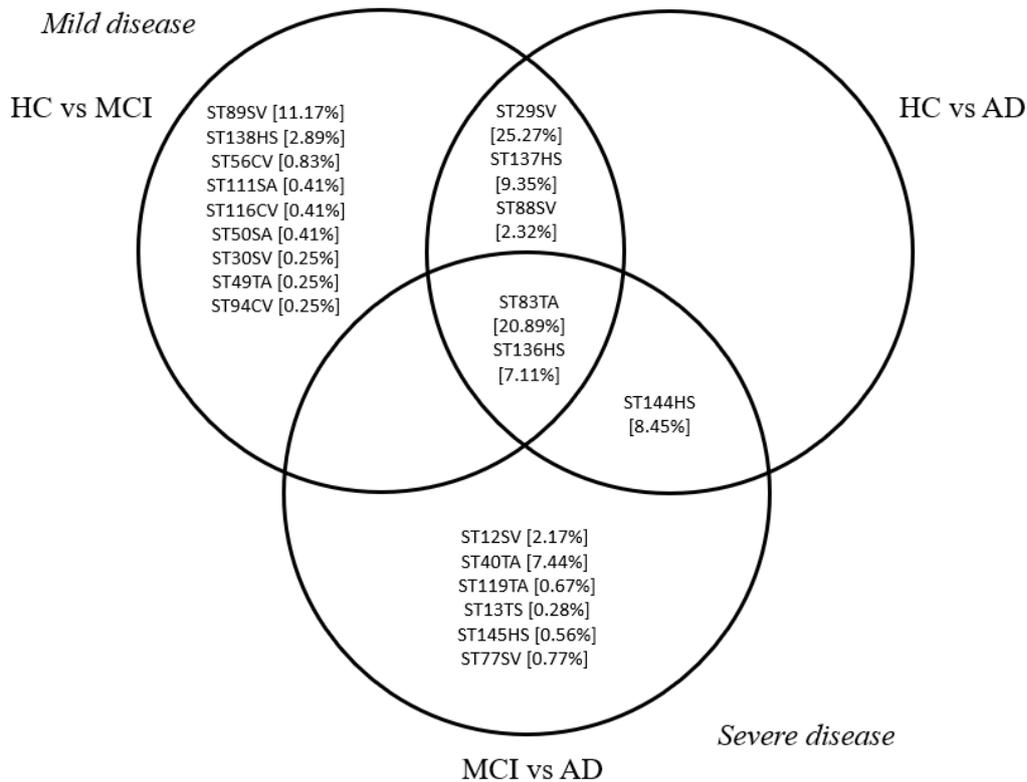
- ST120SV - Subcortical Volume (aseg.stats) of RightThalamus. 6, -16, -6 [Androulakis, 2017]
- ST12SV - Subcortical Volume (aseg.stats) of LeftAmygdala. -18, -4, -12 [Peters, 2013]
- ST136HS - Hippocampal Subfields Volume of LeftPresubiculum. -16.00, -19.00, 18.50 [Palomero-Gallagher, 2020 *]
- ST137HS - Hippocampal Subfields Volume of LeftSubiculum. -20.00, -19.00, -17 [Palomero-Gallagher, 2020 *]
- ST144HS - Hippocampal Subfields Volume of RightPresubiculum. 17.00, -18.00, -20.50 [Palomero-Gallagher, 2020 *]
- ST26TA - Thickness Average (aparc.stats) of LeftFusiform. -29, -57, -14 [Zhang, 2016]
- ST29SV - Subcortical Volume (aseg.stats) of LeftHippocampus. -25.03, -20.74, -10.13 [Borchardt, 2015]
- ST35CV - Cortical Volume (aparc.stats) of LeftLateralOccipital. -30, -90, 6 [Heckendorf, 2016]
- ST40TA - Thickness Average (aparc.stats) of LeftMiddleTemporal. -44.00, -67.00, -19.00 [Crone, 2011]
- ST83TA - Thickness Average (aparc.stats) of RightEntorhinal. 25.00, -2.00, -33.00 [Konishi, 2018]
- ST89SV - Subcortical Volume (aseg.stats) of RightInferiorLateralVentricle. 20, -25, 35 [Rorden, 2007 *]
- ST90TA - Thickness Average (aparc.stats) of RightInferiorParietal. 52, -42, 50 [Rojas, 2018]

Figure S2. Second-ranking ROIs distribution in Axial, Sagittal and Coronal views



Notes. Clockwise from top to bottom left: Axial, Sagittal and Coronal views. The brain networks were generated using BrainNet Viewer (<http://www.nitrc.org/projects/bnv/>) (Xia et al., 2013)

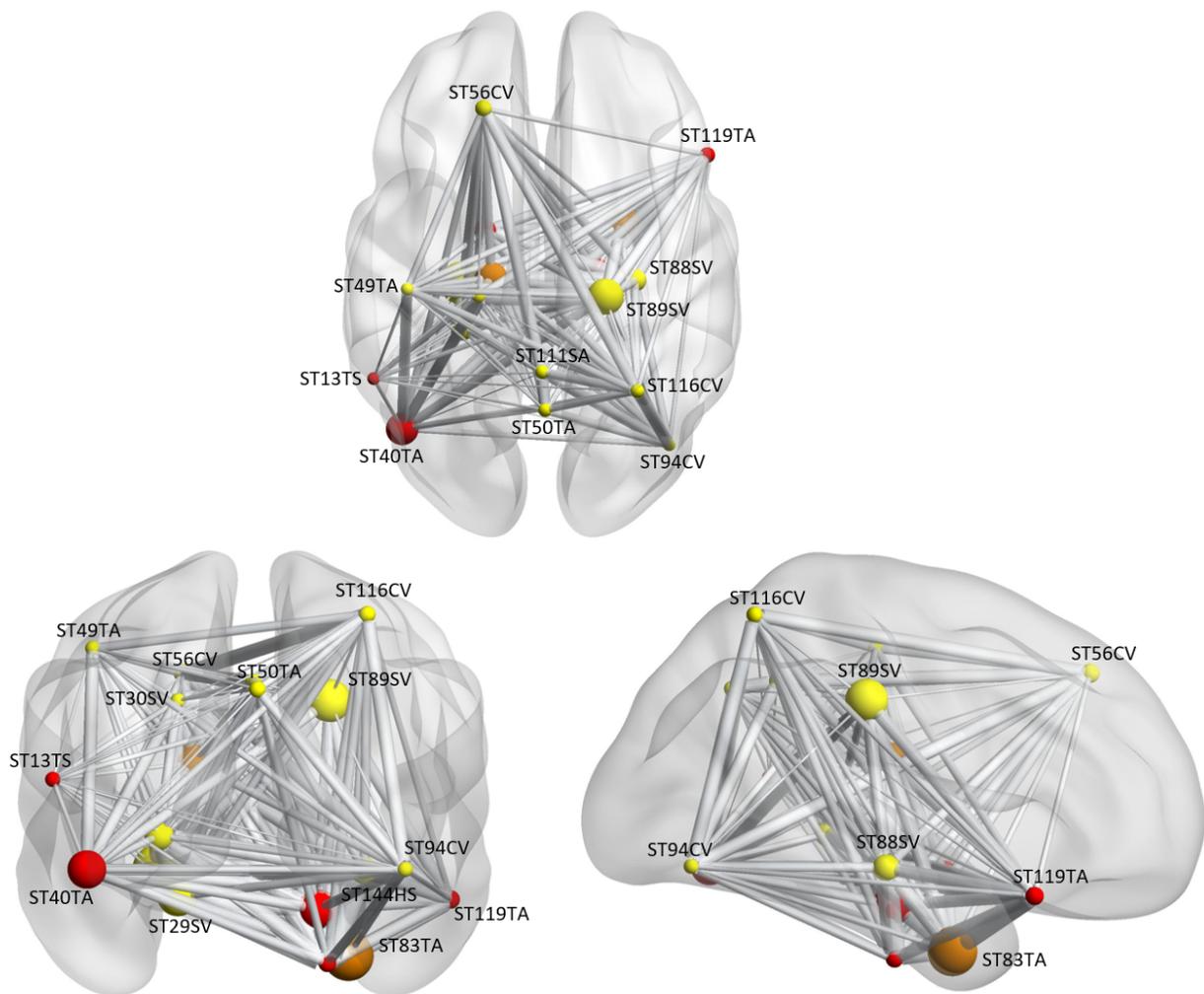
Figure S3. Joint probability of ranking third most important feature over 150,000 iterations



Notes:

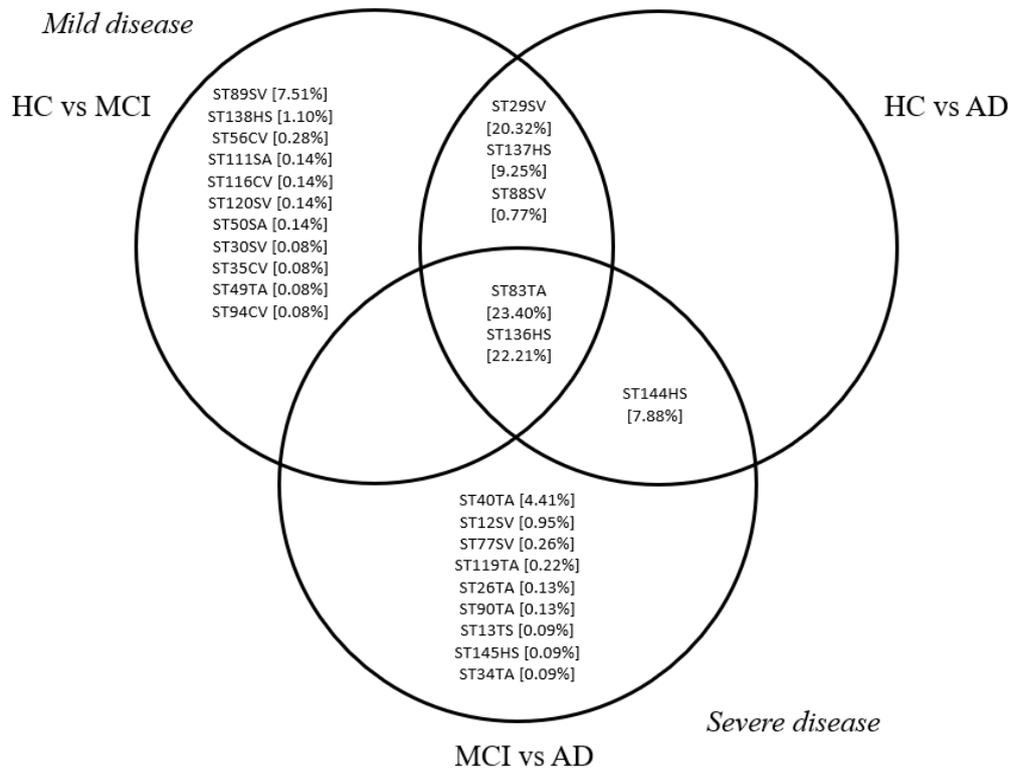
- ST111SA - Surface Area (aparc.stats) of RightPrecuneus. 0, -49, 40 [Androulakis, 2017]
 - ST116CV - Cortical Volume (aparc.stats) of RightSuperiorParietal. 30, -55, 58 [Li, 2012]
 - ST119TA - Thickness Average (aparc.stats) of RightTemporalPole. 52, 20, -18 [Androulakis, 2017]
 - ST12SV - Subcortical Volume (aseg.stats) of LeftAmygdala. -18, -4, -12 [Peters, 2013]
 - ST136HS - Hippocampal Subfields Volume of LeftPresubiculum. -16.00, -19.00, 18.50 [Palomero-Gallagher, 2020 *]
 - ST137HS - Hippocampal Subfields Volume of LeftSubiculum. -20.00, -19.00, -17 [Palomero-Gallagher, 2020 *]
 - ST138HS - Hippocampal Subfields Volume of LeftTail. -25.00, -35.00, -1.00 [Rorden, 2007 *]
 - ST13TS - Thickness Standard Deviation (aparc.stats) of LeftBankssts. -53, -51, 14 [Pitcher, 2020]
 - ST144HS - Hippocampal Subfields Volume of RightPresubiculum. 17.00, -18.00, -20.50 [Palomero-Gallagher, 2020 *]
 - ST145HS - Hippocampal Subfields Volume of RightSubiculum. 19.5, -17.5, -35 [Palomero-Gallagher, 2020 *]
 - ST29SV - Subcortical Volume (aseg.stats) of LeftHippocampus. -25.03, -20.74, -10.13 [Borchardt, 2015]
 - ST30SV - Subcortical Volume (aseg.stats) of LeftInferiorLateralVentricle. -20, -25, 35 [Rorden, 2007 *]
 - ST40TA - Thickness Average (aparc.stats) of LeftMiddleTemporal. -44.00, -67.00, -19.00 [Crone, 2011]
 - ST49TA - Thickness Average (aparc.stats) of LeftPostcentral. -42.46, -22.63, 48.92 [Borchardt, 2015]
 - ST50SA - Surface Area (aparc.stats) of LeftPosteriorCingulate. 1, -61, 38 [Hur, 2021]
 - ST56CV - Cortical Volume (aparc.stats) of LeftSuperiorFrontal. -18.45 34.81 42.2 [Borchardt, 2015]
 - ST77SV - Subcortical Volume (aseg.stats) of RightCerebellumWM. 12 -66 -48 [Stoodley, 2012]
 - ST83TA - Thickness Average (aparc.stats) of RightEntorhinal. 25.00, -2.00, -33.00 [Konishi, 2018]
 - ST88SV - Subcortical Volume (aseg.stats) of RightHippocampus. 29.23 -19.78 -10.33 [Borchardt, 2015]
 - ST89SV - Subcortical Volume (aseg.stats) of RightInferiorLateralVentricle. 20, -25, 35 [Rorden, 2007 *]
 - ST94CV - Cortical Volume (aparc.stats) of RightLateralOccipital. [Heckendorf, 2016]
- The asterisk "*" indicates further authors' elaboration on referenced publication.

Figure S4. Third-ranking ROIs distribution in Axial, Sagittal and Coronal views



Notes. Clockwise from top to bottom left: Axial, Sagittal and Coronal views. The brain networks were generated using BrainNet Viewer (<http://www.nitrc.org/projects/bnv/>) (Xia et al., 2013)

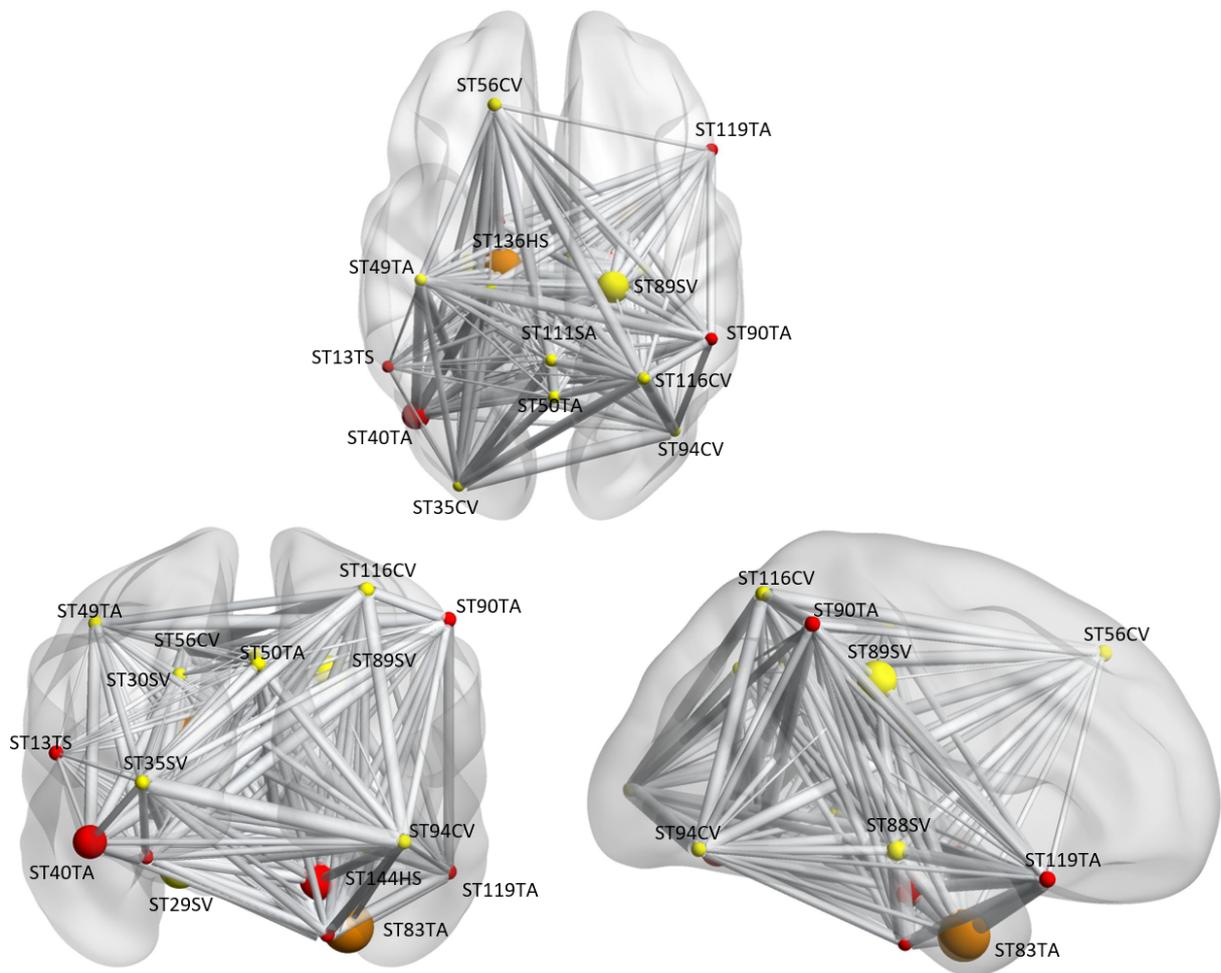
Figure S5. Joint probability of ranking in the first three most important features over 150,000 iterations



Notes:

- ST111SA - Surface Area (aparc.stats) of RightPrecuneus. 0, -49, 40 [Androulakis, 2017]
- ST116CV - Cortical Volume (aparc.stats) of RightSuperiorParietal. 30, -55, 58 [Li, 2012]
- ST119TA - Thickness Average (aparc.stats) of RightTemporalPole. 52, 20, -18 [Androulakis, 2017]
- ST120SV - Subcortical Volume (aseg.stats) of RightThalamus. 6, -16, -6 [Androulakis, 2017]
- ST12SV - Subcortical Volume (aseg.stats) of LeftAmygdala. -18, -4, -12 [Peters, 2013]
- ST136HS - Hippocampal Subfields Volume of LeftPresubiculum. -16.00, -19.00, 18.50 [Palomero-Gallagher, 2020 *]
- ST137HS - Hippocampal Subfields Volume of LeftSubiculum. -20.00, -19.00, -17 [Palomero-Gallagher, 2020 *]
- ST138HS - Hippocampal Subfields Volume of LeftTail. -25.00, -35.00, -1.00 [Rorden, 2007 *]
- ST13TS - Thickness Standard Deviation (aparc.stats) of LeftBankssts. -53, -51, 14 [Pitcher, 2020]
- ST144HS - Hippocampal Subfields Volume of RightPresubiculum. 17.00, -18.00, -20.50 [Palomero-Gallagher, 2020 *]
- ST145HS - Hippocampal Subfields Volume of RightSubiculum. 19.5, -17.5, -35 [Palomero-Gallagher, 2020 *]
- ST26TA - Thickness Average (aparc.stats) of LeftFusiform. -29, -57, -14 [Zhang, 2016]
- ST29SV - Subcortical Volume (aseg.stats) of LeftHippocampus. -25.03 -20.74 -10.13 [Borchardt, 2015]
- ST30SV - Subcortical Volume (aseg.stats) of LeftInferiorLateralVentricle. -20, -25, 35 [Rorden, 2007 *]
- ST35CV - Cortical Volume (aparc.stats) of LeftLateralOccipital. -30, -90, 6 [Heckendorf, 2016]
- ST40TA - Thickness Average (aparc.stats) of LeftMiddleTemporal. -44.00, -67.00, -19.00 [Crone, 2011]
- ST49TA - Thickness Average (aparc.stats) of LeftPostcentral. -42.46, -22.63, 48.92 [Borchardt, 2015]
- ST50SA - Surface Area (aparc.stats) of LeftPosteriorCingulate. 1, -61, 38 [Hur, 2021]
- ST56CV - Cortical Volume (aparc.stats) of LeftSuperiorFrontal. -18.45 34.81 42.2 [Borchardt, 2015]
- ST77SV - Subcortical Volume (aseg.stats) of RightCerebellumWM. 12 -66 -48 [Stoodley, 2012]
- ST83TA - Thickness Average (aparc.stats) of RightEntorhinal. 25.00, -2.00, -33.00 [Konishi, 2018]
- ST88SV - Subcortical Volume (aseg.stats) of RightHippocampus. 29.23 -19.78 -10.33 [Borchardt, 2015]
- ST89SV - Subcortical Volume (aseg.stats) of RightInferiorLateralVentricle. 20, -25, 35 [Rorden, 2007 *]
- ST90TA - Thickness Average (aparc.stats) of RightInferiorParietal. 52, -42, 50 [Rojas, 2018]
- ST94CV - Cortical Volume (aparc.stats) of RightLateralOccipital. 40, -72, -10 [Heckendorf, 2016]

Figure S6. Third-ranking ROIs distribution in Axial, Sagittal and Coronal views



Notes. Clockwise from top to bottom left: Axial, Sagittal and Coronal views. The brain networks were generated using BrainNet Viewer (<http://www.nitrc.org/projects/bnv/>) (Xia et al., 2013)

References

- Ahmad MA, Eckert C, Teredesai A. Interpretable machine learning in healthcare. In Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics 2018 Aug 15 (pp. 559-560).
- Androulakis XM, Krebs K, Peterlin BL, Zhang T, Maleki N, Sen S, Rorden C, Herath P. Modulation of intrinsic resting-state fMRI networks in women with chronic migraine. *Neurology*. 2017 Jul 11;89(2):163-9.
- Bonner-Jackson A, Mahmoud S, Miller J, Banks SJ. Verbal and non-verbal memory and hippocampal volumes in a memory clinic population. *Alzheimer's research & therapy*. 2015 Dec;7(1):1-0.
- Borchardt V, Krause AL, Li M, van Tol MJ, Demenescu LR, Buchheim A, Metzger CD, Sweeney-Reed CM, Nolte T, Lord AR, Walter M. Dynamic disconnection of the supplementary motor area after processing of dismissive biographic narratives. *Brain and Behavior*. 2015 Oct;5(10):e00377.
- Bottino CM, Castro CC, Gomes RL, Buchpiguel CA, Marchetti RL, Neto MR. Volumetric MRI measurements can differentiate Alzheimer's disease, mild cognitive impairment, and normal aging. *International Psychogeriatrics*. 2002 Mar;14(1):59-72.
- Bozzali M, Filippi M, Magnani G, Cercignani M, Franceschi M, Schiatti E, Castiglioni S, Mossini R, Falautano M, Scotti G, Comi G. The contribution of voxel-based morphometry in staging patients with mild cognitive impairment. *Neurology*. 2006 Aug 8;67(3):453-60.
- Chételat G, Landeau B, Eustache F, Mézenge F, Viader F, de La Sayette V, Desgranges B, Baron JC. Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: a longitudinal MRI study. *Neuroimage*. 2005 Oct 1;27(4):934-46.
- Du Y, Rafferty AR, McAuliffe FM, Wei L, Mooney C. An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus. *Sci Rep*. 2022 Jan 21;12(1):1170. doi: 10.1038/s41598-022-05112-2. PMID: 35064173; PMCID: PMC8782851.
- Galton CJ, Erzinclioglu S, Sahakian BJ, Antoun N, Hodges JR. A comparison of the Addenbrooke's Cognitive Examination (ACE), conventional neuropsychological assessment, and simple MRI-based medial temporal lobe evaluation in the early diagnosis of Alzheimer's disease. *Cogn Behav Neurol*. 2005;18(3):144-150.
- Guo S, Xiao B, Wu C, Alzheimer's Disease Neuroimaging Initiative. Identifying subtypes of mild cognitive impairment from healthy aging based on multiple cortical features combined with volumetric measurements of the hippocampal subfields. *Quantitative Imaging in Medicine and Surgery*. 2020 Jul;10(7):1477.

Hämäläinen A, Tervo S, Grau-Olivares M, Niskanen E, Pennanen C, Huuskonen J, Kivipelto M, Hänninen T, Tapiola M, Vanhanen M, Hallikainen M. Voxel-based morphometry to detect brain atrophy in progressive mild cognitive impairment. *Neuroimage*. 2007 Oct 1;37(4):1122-31.

Hauskrecht M, Pelikan R, Valko M, Lyons-Weiler J. Feature selection and dimensionality reduction in genomics and proteomics. *Fundamentals of data mining in genomics and proteomics*. 2007:149-72.

Heckendorf E, Huffmeijer R, Bakermans-Kranenburg MJ, Van IJzendoorn MH. Neural processing of familiar and unfamiliar children's faces: Effects of experienced love withdrawal, but no effects of neutral and threatening priming. *Frontiers in human neuroscience*. 2016 May 26;10:231.

Huang Y, Huang L, Wang Y, Liu Y, Lo CY, Guo Q. Differential associations of visual memory with hippocampal subfields in subjective cognitive decline and amnesic mild cognitive impairment. *BMC geriatrics*. 2022 Dec;22(1):1-0.

Hur JW, Kim T, Cho KI, Kwon JS. Attenuated resting-state functional anticorrelation between attention and executive control networks in schizotypal personality disorder. *Journal of clinical medicine*. 2021 Jan 15;10(2):312.

Jain D, Singh V. Feature selection and classification systems for chronic disease prediction: A review. *Egyptian Informatics Journal*. 2018 Nov 1;19(3):179-89.

Kong Y, Deng Y, Dai Q. Discriminative clustering and feature selection for brain MRI segmentation. *IEEE Signal Processing Letters*. 2014 Oct 23;22(5):573-7.

Li Q, Liu C, Oster J, Clifford GD. Signal processing and feature selection preprocessing for classification in noisy healthcare data. *Machine Learning for Healthcare Technologies*. 2016 Oct 28;2(33):2016.

Molnar C. *Interpretable machine learning*. Lulu. com; 2020.

Peters J, Miedl SF, Büchel C. Elevated functional connectivity in a striatal-amygdala circuit in pathological gamblers. *PLoS one*. 2013 Sep 2;8(9):e74353.

Petrella JR, Krishnan S, Slavin MJ, Tran TT, Murty L, Doraiswamy PM. Mild cognitive impairment: evaluation with 4-T functional MR imaging. *Radiology*. 2006 Jul;240(1):177-86.

Pitcher D, Pilkington A, Rauth L, Baker C, Kravitz DJ, Ungerleider LG. The human posterior superior temporal sulcus samples visual space differently from other face-selective regions. *Cerebral Cortex*. 2020 Feb;30(2):778-85.

Polat H, Danaei Mehr H, Cetin A. Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. *Journal of medical systems*. 2017 Apr;41:1-1.

Rojas GM, Alvarez C, Montoya CE, De la Iglesia-Vaya M, Cisternas JE, Gálvez M. Study of resting-state functional connectivity networks using EEG electrodes position as seed. *Frontiers in neuroscience*. 2018 Apr 24;12:235.

Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*. 2015 Mar 4;10(3):e0118432.

Simon N, Friedman J, Hastie T, Tibshirani R (2013) A Sparse-Group Lasso, *Journal of Computational and Graphical Statistics*, 22:2, 231-245, DOI: 10.1080/10618600.2012.681250

Stoodley CJ, Valera EM, Schmahmann JD. Functional topography of the cerebellum for motor and cognitive tasks: an fMRI study. *Neuroimage*. 2012 Jan 16;59(2):1560-70.

Sutter JM, Kalivas JH. Comparison of forward selection, backward elimination, and generalized simulated annealing for variable selection. *Microchemical journal*. 1993 Feb 1;47(1-2):60-6.

Therriault J, Zimmer ER, Benedet AL, Pascoal TA, Gauthier S, Rosa-Neto P. Staging of Alzheimer's disease: past, present, and future perspectives. *Trends in Molecular Medicine*. 2022 Sep 1;28(9):726-41.

Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 1996 Jan;58(1):267-88.

Xia M, Wang J, He Y (2013) BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *PLoS ONE* 8: e68910.

Xing EP, Jordan MI, Karp RM. Feature selection for high-dimensional genomic microarray data. *Inlcml* 2001 Jun 28 (Vol. 1, No. 3, pp. 601-608).

Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2006 Feb;68(1):49-67.

Zhang N, Ruan S, Lebonvallet S, Liao Q, Zhu Y. Kernel feature selection to fuse multi-spectral MRI images for brain tumor segmentation. *Computer Vision and Image Understanding*. 2011 Feb 1;115(2):256-69.

Zhang W, Wang J, Fan L, Zhang Y, Fox PT, Eickhoff SB, Yu C, Jiang T. Functional organization of the fusiform gyrus revealed with connectivity profiles. *Human brain mapping*. 2016 Aug;37(8):3003-16.

Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2005 Apr;67(2):301-20.