Neurobiology of Aging 133 (2024) 67–77

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging.org

Sex modifies effects of imaging and CSF biomarkers on cognitive and functional outcomes: a study of Alzheimer's disease

Brian N. Lee [a], Junwen Wang [b], Molly A. Hall [c], Dokyoon Kim [a], Shana D. Stites [d], Li Shen [a,*], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, The University of Pennsylvania, Philadelphia, PA, USA
[b] Department of Health Sciences Research, Mayo Clinic Alix School of Medicine, Phoenix, AZ, USA
[c] Department of Veterinary and Biomedical Sciences, College of Agricultural Sciences, The Pennsylvania State University, University Park, PA, United States
[d] Department of Psychiatry, Perelman School of Medicine, The University of Pennsylvania, Philadelphia, PA

ARTICLE INFO

Keywords:
Alzheimer's disease
Sex-based disparity
Brain imaging
Cerebral spinal fluid (CSF) biomarker
Cognitive and functional outcome

ABSTRACT

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by memory and functional impairments. Two of 3 patients with AD are biologically female; therefore, the biological underpinnings of this diagnosis disparity may inform interventions slowing the AD progression. To bridge this gap, we conducted analyses of 1078 male and female participants from the Alzheimer's Disease Neuroimaging Initiative to examine associations between levels of cerebral spinal fluid (CSF)/neuroimaging biomarkers and cognitive/functional outcomes. The Chow test was used to quantify sex differences by determining if biological sex affects relationships between the studied biomarkers and outcomes. Multiple magnetic resonance imaging (whole brain, entorhinal cortex, middle temporal gyrus, fusiform gyrus, hippocampus), position emission tomography (AV45), and CSF (P-TAU, TAU) biomarkers were differentially associated with cognitive and functional outcomes. Post-hoc bootstrapped and association analyses confirmed these differential effects and emphasized the necessity of using separate, sex-stratified models. The studied imaging/CSF biomarkers may account for some of the sex-based variation in AD pathophysiology. The identified sex-varying relationships between CSF/imaging biomarkers and cognitive/functional outcomes warrant future biological investigation in independent cohorts.

1. Introduction

Alzheimer's disease (AD) is a complex neurodegenerative disease commonly characterized by memory impairments and other cognitive and functional problems, and the presence of both tau tangles and amyloid beta plaques (Dubois et al., 2016). As the leading cause of dementia, AD is influenced by a variety of environmental and genetic factors (Lane et al., 2018). Researchers are facing major challenges in developing effective preventative care and therefore have examined the role biological sex may play in AD (National Institutes of Health, 2018).

Previous studies have found significant differences in brain structure and function between biological male and female participants (Ruigrok et al., 2014; Ingalhalikar et al., 2014) that may result in differential effects of sex on the evolution and progression of neurodegenerative disorders such as AD (Ferretti et al., 2018). Sex differences in AD risk may be additionally attributable to differences in life expectancy or

Abbreviations: ABETA, Amyloid-beta; AD, Alzheimer's disease; ADAS13, Alzheimer's Disease Assessment Scale; ADNI, Alzheimer's Disease Neuroimaging Initiative; ANOVA, Analysis of variance; APOE, Apolipoprotein E; AV45, Florbetapir (18F-AV-45) PET imaging; CDRSB, Clinical Dementia Rating Scale, Sum of Boxes; CSF, Cerebral spinal fluid; DNA, Deoxyribonucleic acid; EC, Entorhinal cortex; FAQ, Functional Activities Questionnaire; FDG-PET, Fluorodeoxyglucose positron emission tomography; HC, Healthy control; ICV, Intracranial volume; MCI, Mild cognitive impairment; MMSE, Mini-Mental State Exam; MOCA, Montreal Cognitive Assessment; MRI, Magnetic resonance imaging; PET, Positron emission tomography; PTAU, Phosphorylated TAU; RAVLT.learning, Rey's auditory verbal learning test, learning subscore; SUVR, Standardized uptake value ratio; QT-PAD, Quantitative Templates for the Progression of Alzheimer's disease; US, United States.

* Correspondence to: B306 Richards Building, 3700 Hamilton Walk, Philadelphia, PA 19104, USA.
E-mail address: li.shen@pennmedicine.upenn.edu (L. Shen).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

https://doi.org/10.1016/j.neurobiolaging.2023.10.002
Received 6 September 2022; Received in revised form 9 October 2023; Accepted 10 October 2023
Available online 20 October 2023

expression of the apolipoprotein E (APOE) $\epsilon 4$ genotype (Li and Singh, 2014; Nebel et al., 2018; Laws et al., 2018; González Zarzar et al., 2022). However, to date, no data model or biological paradigm fully explains these sex-based differences in AD risk. Therefore, novel approaches for studying AD-related fluid biomarkers, imaging measurements, and cognitive data are necessary. Accurate diagnosis and treatment of complex diseases, such as AD, can be greatly supported by intimate knowledge of how sex affects disease biology and relevant cognitive/-functional outcomes. Currently, there are multiple approaches to quantifying the impact of sex on AD biology and relevant outcomes. The easiest and most prevalent technique is to use biological sex as a covariate. Although the main effect of sex on brain structure, function, and AD risk can be captured by including sex as a covariate, more accurate or complex models are needed to detect additional effects of sex that may interact with genetic, neurobiological, environmental and/or other variables (National Institutes of Health, 2019).

To the best of our knowledge, there are no comparable statistical tests that have been established as the unequivocal gold standard to quantify the significance of sex-based differences. However, sex-stratified models and sex interaction models are popular choices for examining sex differences. Sex-stratified models may be useful to identify a trend present within a specific sex group (i.e., a male population difference) but comparing the magnitude of trends between 2 sex-defined groups may be difficult. For example, a previous study has employed a test to examine the difference between effects estimated by stratified analyses in order to detect sex differences in metabolomics during AD progression (González Zarzar et al., 2022). Sex interaction models may allow for the identification of trends only present in an individual sex group but might not be able to detect subtle (but significant!) sex-based differences in trends that exist in both male and female populations.

In this work, we propose the Chow test as a novel alternative method to perform scalable, biologically-informed sex-stratified analyses. The Chow test can determine (1) if a specific independent variable (i.e., a fluid or imaging quantitative trait) has different impacts on a specific dependent variable (i.e., a cognitive and functional outcome) across different subgroups of a population (i.e., the stratifying variable, biological sex) and (2) report the magnitude of the sex-based differences. Specifically, the Chow test compares the parameters of 2 sex-stratified linear regression models (1 for male, 1 for female) in order to quantitatively detect the presence of a sex-based difference and report its magnitude. To our current knowledge, this is a novel application of the Chow test to the AD sphere generally and studying sex-based differences in AD specifically. The Chow test has been utilized to find genetic subtypes in cancers but is under-explored in the study of AD (Fu et al., 2019; Liang et al., 2016).

Unlike sex-stratified or sex-interaction analyses, the Chow test allows for the synthesis of multiple regression parameters within the scope of a single model via the use of an F statistic. Doing so allows researchers to quickly ascertain if there are statistically-significant differences between sex groups (i.e., male vs. female participants), this is the primary innovation that differentiates the Chow test from sex-stratified and sex-interaction models. Because it reports a holistic p value quantifying the significance of a sex difference, the Chow test can be used in conjunction with sex-stratified or sex-interaction models.

In this study, the Chow test is used to perform a systematic examination of how male and female sex modifies associations between AD quantitative traits and cognitive and functional outcomes (henceforth referred to as trait-outcome associations). We hypothesize that there will be a statistically significant difference between the regression coefficients of imaging and cerebrospinal fluid (CSF) biomarkers between male and female participants, which would greatly support efforts to adopt models for AD that more fairly and accurately account for sex-based differences.

2. Materials and methods

2.1. . Quantitative trait and diagnosis data

Since we are specifically interested in examining the differential effects of biological sex in AD pathophysiology and thus diagnosis, we started our analysis by extracting patient demographic data (including self-reported biological sex, age, and years of education) as well as key AD biomarker measurements from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (ADNI Team, 2021).

ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD, to test whether serial magnetic resonance imaging (MRI), position emission tomography (PET), and biological markers can be combined with clinical and neuropsychological assessments to accurately measure the progression of mild cognitive impairment (MCI) and early AD. For up-to-date information, see www.adni.loni.usc.edu.

Participants included individuals who were members of ADNI GO and 2 cohorts, as described by the ADNI "Quantitative Template for the Progression of Alzheimer's Disease" (QT-PAD) project. Please refer to QT-PAD Team (2021) for details about the QT-PAD data and how participants were chosen. There were 16 AD traits included in the QT-PAD, which included 5 cognitive and functional outcomes (ADAS13, CDRSB, RAVLT.learning, MMSE, and FAQ), 2 PET quantitative traits (FDG PET and Amyloid PET/AV45), 3 CSF quantitative traits (ABETA, tau, and p-tau) (Kang et al., 2015), and 6 MRI FreeSurfer quantitative traits (WholeBrain, Entorhinal, Ventricles, MidTemp, Fusiform, and Hippocampus). Of note, FDG-PET measure is the average measure of angular, temporal, and posterior cingulate regions. AV45 measure is the average standardized uptake value ratio measure of frontal, anterior cingulate, precuneus, and parietal cortex relative to the cerebellum.

Study inclusion criteria for the ADNI studies attempted to find members of the US population suitable for large-scale AD studies. General inclusion categories required individuals be between 55 and 90 years of age, have a study partner or caregiver willing to accompany the participant to all visits, be fluent in English or Spanish, be stable (for at least 4 weeks) on permitted medications, have adequate visual and auditory acuity to engage in neuropsychological testing, have good general health (and especially no other diseases expected to interfere with the study), are willing to complete all necessary assessments (including baseline assessments, neuroimaging, providing DNA and plasma samples), have sufficient evidence to exclude intellectual disability, a modified Hachinski score less than 5, and a Geriatric Depression scale less than 6. Additionally, female participants must be 2 years post-menopausal. Study participants presenting with MCI or Alzheimer's dementia (AD) must also have a memory complaint (by the patient or study partner) and abnormal memory scores on a number of cognitive measures (the Wechsler Memory Scale, Mini-Mental State Exam, and Clinical Dementia Rating).

Selection criteria for the present study include restricting analysis to participants with at least one AV45 data point (to control for the relatively fewer samples of PET, AV45, and tau measurements at the early stages of data collection). Visit codes for each participant were also manually adjusted to control for the relatively low number of CSF biomarker measurements; the first ADNI GO or 2 visit was set as the patient's "baseline" visit, with the "month 24" visit being defined as the visit closest to being 24 months after the established baseline. Redefining visit codes help keep sample sizes of biomarkers relatively consistent while maximizing the number of imaging and CSF biomarker measurements included in the study. After quality control, the analysis included 1078 individuals (578 male, 500 female). Characteristics of participants analyzed in analyses can be found in Table 1; effect sizes and p values (using a t-test for continuous variables and a χ^2 test for categorical variables) are included to facilitate comparison between males and females.

Table 1
Participant characteristics at the baseline visit

Characteristic	All (n = 1078)	Male (n = 578)	Female (n = 500)	Effect size	*p*
Diagnosis, n (HC, MCI, AD)	378, 556, 144	179, 315, 84	199, 241, 60	0.093	0.010
Age, mean (SD)	66.08 (0.69)	66.92 (0.96)	65.14 (1.01)	0.103	0.165
Education, mean (SD)	14.57 (0.20)	15.03 (0.28)	14.05 (0.29)	0.195	0.017
APOE4 dosage, mean (SD)	0.53 (0.03)	0.53 (0.04)	0.52 (0.04)	0.009	0.910
MMSE, mean (SD)	27.09 (0.12)	27.07 (0.08)	27.20 (0.19)	0.054	0.407
FAQ, mean (SD)	3.79 (0.18)	4.35 (0.27)	3.14 (0.26)	0.204	0.001
CDRSB, mean (SD)	1.63 (0.06)	1.77 (0.09)	1.46 (0.09)	0.154	0.013
ADAS13, mean (SD)	15.71 (0.32)	16.67 (0.40)	14.61 (0.48)	0.206	0.001
RAVLT.learning, mean (SD)	4.89 (0.11)	4.54 (0.15)	5.30 (0.16)	0.208	0.001

In columns 2–4, mean (standard deviation) is shown for each continuous variable. Columns 5 and 6 show effect size and *p* value, which are the results from statistical group comparison between males and females (using t-test for continuous variables and χ^2 test for categorical variables. Effect sizes for continuous variables are Cohen's d statistics; effect size for diagnosis is a Cramer's V statistic.
Key: AD, Alzheimer's disease; HC, healthy control; MCI, mild cognitive impairment; SD, standard deviation.

2.2. Chow test method

The Chow test was initially used in econometrics to determine if the relationship between an independent variable and a dependent variable changed after a major historical event (i.e., a war) or because of a categorical stratifying factor (i.e., race) (Chow, 1960). More generally, this test can be used to determine whether the true coefficients of 2 analogous linear regression models built from stratified data sets are equal. If the coefficients differ a statistically significant amount (as determined by a F test), one can conclude the change is due to the stratifying factor or significant temporal event. In this work, we implemented and conducted all the Chow tests in R. More details about the Chow test are available in Chow (1960).

Within the context of this study, the Chow test will determine if there are statistically significant sex-based differences in associations between each of 11 imaging and CSF QT-PAD quantitative traits and each of 5 cognitive and functional outcomes. Sex-based differences in these trait-outcome associations may indicate certain biological pathways, brain regions, or cognitive processes that account for the vast sex-based disparities in AD diagnosis reported in the literature. The 11 imaging and CSF quantitative traits include CSF quantitative traits, PET imaging quantitative traits (i.e., FDG measuring glucose metabolism, AV45 measuring amyloid burden), and MRI imaging quantitative traits (i.e., those extracted by the FreeSurfer or FS software). Prior to analysis via the Chow test, FreeSurfer volumetric measurements were adjusted for the intracranial volume using the regression coefficients derived from the ADNI healthy control participants (Shen et al., 2010). The five cognitive and functional outcomes include quantitative scores from the ADAS13, CDRSB, RAVLT.learning, MMSE, and FAQ cognitive tests.

The models evaluated by the Chow test attempted to measure the effect of a specific CSF or imaging quantitative trait (denoted x_{QT}) on each cognitive and functional outcome y_{cog} while also controlling for age (c_{AGE}) and education ($c_{EDUCATION}$). In R's statistical formula notation:

$$y_{cog} \sim x_{QT} + c_{AGE} + c_{EDUCATION}$$

An additional set of models factor for an individual's allelic dosage of known AD risk allele APOE4 (c_{APOE4}). Please note that our study modeled APOE4 continuously. In R's statistical formula notation, these models can be represented as:

$$y_{cog} \sim x_{QT} + c_{AGE} + c_{EDUCATION} + c_{APOE4}$$

The APOE and non-APOE linear regression models are fitted for each of 5 available cognitive and functional outcomes and all 11 available imaging and fluid biomarkers at 2 time points (i.e., using biomarker data collected at the baseline and month 24 visits). Additionally, these models are fitted for each of 2 subpopulations per biological sex (male or female). Significance between the individual stratified models and the summative model is determined using the F statistic and was expressed as a *p* value. Significant relationships were chosen using a Bonferroni threshold to correct for multiple comparisons across all Chow tests $\left(p < \frac{0.05}{(5\times11\times4)} = 2.27\times10^{-4}\right)$.

2.3. Post-hoc bootstrapping analyses

To visually confirm the differential effects on specific quantitative trait-cognitive and functional outcome associations, we bootstrapped the male and female linear regression coefficients calculated for several trait-outcome associations. Bootstrapped regression models utilized age and years of education as covariates but not APOE status. Only a handful of associations with a significant Chow test threshold were examined (as determined by the Bonferroni-corrected *p* value threshold of $p < 2.27\times10^{-4}$). To maximize the number of relationships that could be examined via bootstrapping, APOE was not factored as an additional covariate in these analyses.

Since there were too many individual relationships for each statistically-significant Chow test relationship to be visualized, only a handful of strong relationships were strategically chosen (as determined by Chow test *p* value). Visualization and analysis in post-hoc steps are primarily to confirm the significance of sex differences identified by the Chow test; all relationships with a significant Chow *p* comprise the primary findings of this manuscript. To allow for a diverse set of relationships to be determined, 2 of the significant imaging features and 1 significant CSF feature were visualized; the number of relationships studied was chosen manually. Predictor-outcome relationships involving the fusiform and middle temporal gyrus FreeSurfer biomarkers had relatively low Chow test *p* values among the individual FreeSurfer imaging biomarkers and were therefore chosen for visualization and post-hoc analysis. Tau had more significant Chow test *p* values than p-tau or A-beta and therefore was chosen for visualization and post-hoc analysis.

2.4. Post-hoc linear models

To visually confirm the differential effects on specific quantitative trait-cognitive and functional outcome associations, we also visualized the different trends between male participants, female participants, and all participants. Both individual data points representing measurements from individual participants and group-stratified linear regression lines are shown (1 for male participants, 1 for female participants, and 1 for all participants). Only a handful of associations with a significant Chow test threshold were examined (as determined by the Bonferroni-corrected *p* value threshold of $p < 2.27\times10^{-4}$). The relationship between CSF/imaging predictor and an adjusted cognitive/functional outcome value was plotted. The cognitive/functional outcome value was adjusted for the common covariates of age and years of education.

Since there were too many individual relationships for each statistically-significant Chow test relationship to be visualized, only a handful of the most significant relationships were strategically chosen (as determined by Chow test *p* value). The same logic used to select relationships for Fig. 2 was used to choose relationships to visualize in

Fig. 3. Additionally, to maximize the number of relationships that could be visualized, APOE was not factored as an additional covariate in these analyses.

2.5. Post-hoc ANOVA

Analysis of variance (ANOVA) was also performed to measure the differences between levels of specific quantitative traits between individuals of male and female sex. The magnitude of the differences between the mean levels of specific biomarkers for male and female participants is reported for each biomarker at each available visit code in the form of a *Cohen's d* statistic.

For the same reasons as explained in Sections 2.3 and 2.4, only a select number of CSF/imaging predictor variables were analyzed in post-hoc ANOVA analyses: baseline and month 24 fusiform gyrus, baseline and month 24 middle temporal gyrus, and baseline tau.

2.6. Post-hoc comparison association analyses

To confirm the utility of stratifying AD participants by biological sex before modeling, sets of linear regression association tests were performed and the resulting p values compared. The first set of association analyses will be framed similar to a traditional association study with regard to biological sex: the association between a cognitive/functional outcome (such as ADAS13) and an imaging/CSF quantitative trait (such as WholeBrain Free Surfer volume) will be measured using biological sex, years of education, age, and APOE4 allelic dosage as covariates. The second set of association analyses will stratify the ADNI participants on their biological sex before calculating the exact same associations. Note that these stratified associations are intended to emulate future studies of biological sex effects and make a broader argument in favor of the Chow test's stratified approach to association studies; therefore, all covariates—notably, including APOE4—are included in these analyses. This is distinct from the bootstrapping analyses and scatterplots; those post-hoc analyses are primarily intended to visualize as many significant relationships as possible to verify the veracity of the Chow test findings.

Notably, the individual hypotheses assessed in these stratified associations are distinct from the Chow test associations. Please note that the Chow test p values merely represent if a significant difference between the regression coefficients of the 3 stratified associations exists; such a difference would be indicative of a biological sex effect. A Chow test p value is not indicative of the significance of any 1 of the 3 individual stratified models. However, in the individual stratified association analyses performed in this subsection, a significant p value would represent that there exists a significant association between an imaging/CSF biomarker and the cognitive/functional outcome (after factoring for relevant covariates) within the specified sex-based subpopulation. Therefore, a significant Chow test p value (see Fig. 1) does not imply significance in the stratified associations assessed here, and vice versa.

Significance across all association tests will be determined using a Bonferroni-corrected p value threshold to account for all comparisons being made $\left(p < \frac{0.05}{5\times11\times6} = 1.52 \times 10^{-4}\right)$.

After the calculation of all models, it will be possible to compare the p

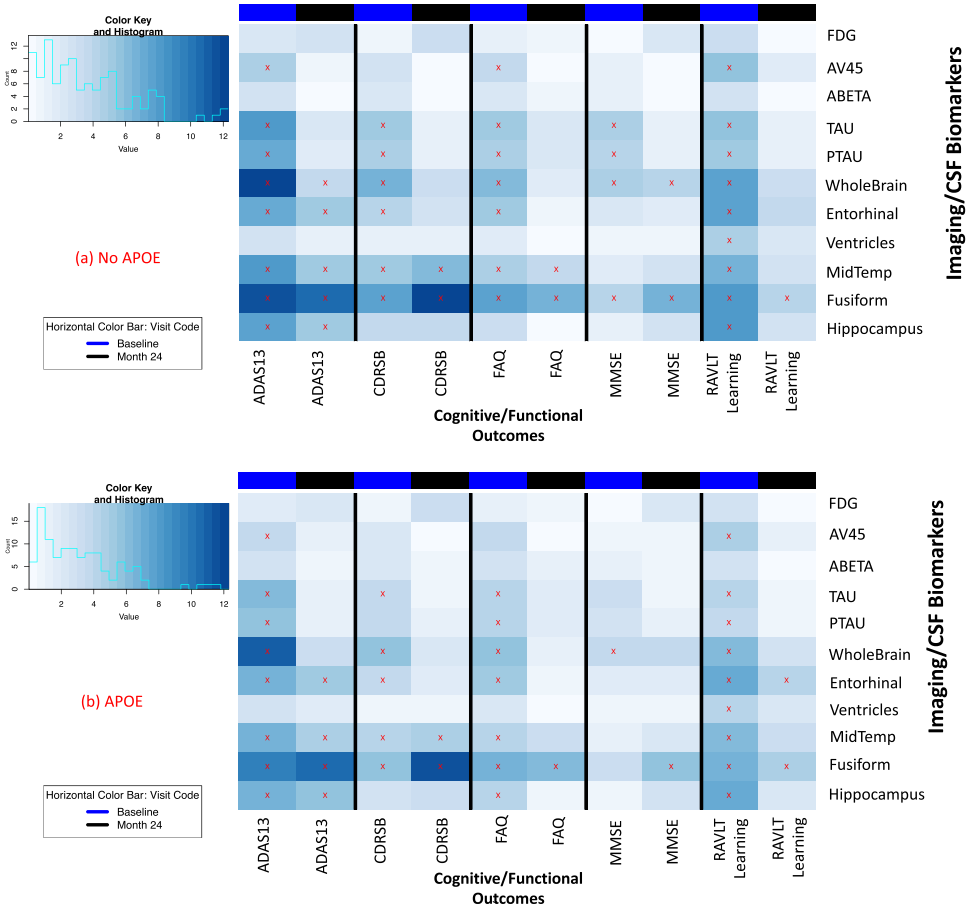

Fig. 1. Heat map showing results of Chow test. Regression: associate cognitive and functional outcomes (horizontal axis) using a variety of imaging and CSF biomarkers (vertical axis) when factoring for age and years of education as covariates. Horizontal color bar represents specific QT-PAD visit code data selection is from and shade of cells denotes relative-log(Chow Test p). Significance was determined by a Bonferroni threshold ($p < 2.27 \times 10^{-4}$) with significant relationships denoted as X. Subfigure (a) uses age and education as covariates; subfigure (b) uses APOE4 status, age, and education as covariates. Horizontal color bar signifies visit code; blue-labeled columns correspond to baseline relationships and black-labeled columns correspond to month 24 relationships.

values of the traditional model to the 2 p values corresponding to the relevant male/female models. In this way, it can be determined if the sex-stratified association tests reveal sex-specific patterns in AD pathogenesis that would otherwise be missed in the traditional association test paradigm.

3. Results

3.1. Baseline ADNI participant demographics

Although ADNI participants had a similar number of self-identified male and female participants at the baseline visit, there were significant differences in the population regarding multiple measured traits. For instance, male and female participants had significantly different average years of education, FAQ mean scores, CDRSB mean scores,

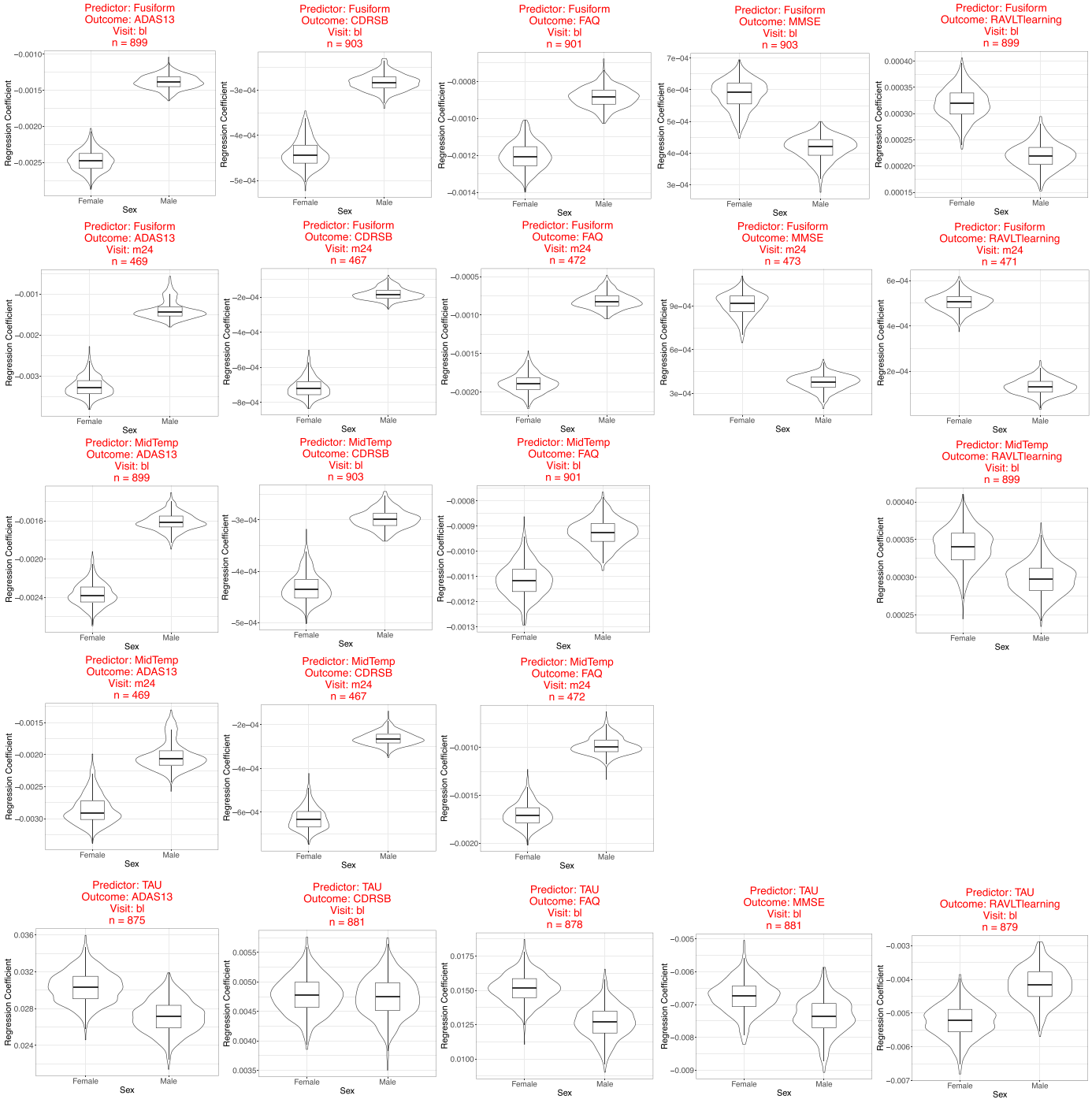

Fig. 2. Violin plots showing the results of bootstrapping analysis ($n = 600$; randomly sampling 80% of the available data per sex per iteration) to evaluate the results of the Chow tests. A select number of top significant correlations from the Chow tests (see Fig. 1) were chosen. Each plot depicts a unique predictor-outcome relationship assessed. The horizontal axis of each plot indicates the stratified sex (male or female) and the vertical axis of each plot indicates the magnitude of the (bootstrapped) regression coefficient. For ease of viewing, plots have been sorted into rows and columns based on the specific predictor and outcome utilized. The box plot shows the median and interquartile range of calculated regression coefficients and specific points highlight outliers.

ADAS13 mean scores, RAVLT learning mean scores, and proportions of individuals with varying diagnosis codes (all $p < 0.05$, see Table 1). No sex differences were noticed in age or APOE4 dosage between male and female participants at the baseline visit ($p \geq 0.05$).

3.2. Chow test results

Our Chow test results are summarized and shown in Fig. 1. Several FreeSurfer imaging biomarkers predicted AD cognitive scores with differing regression coefficients between male and female participants over the two time points. All statistically significant pairs of outcomes (as determined by a Chow Test p value smaller than the Bonferroni threshold of 2.27×10^{-4}) are marked with a red "X." Results from models not using APOE as an additional covariate are visualized in subfigure (a) and results from models using APOE as an additional covariate are visualized in subfigure (b).

The strongest sex-based differences involved the FreeSurfer quantitative traits of the fusiform gyrus, midtemporal cortex, and whole brain predicting cognitive and functional outcomes (with non-APOE p values like 2.53×10^{-12} for the relationship where fusiform gyrus volume predicts ADAS13 using baseline data). Many of these strong sex-based differences were also noted at 24 months (the fusiform gyrus-ADAS13 relationship has a Chow test $p = 1.57 \times 10^{-11}$. Noticeably, many of the strongest signals are present in both the APOE and non-APOE models, with slightly less significant P values in the APOE models. There were also statistically significant relationships noted involving imaging and CSF biomarkers such as AV45 and tau. Baseline AV45 measurements were associated with sex-based differences in the ADAS13, FAQ, and RAVLT.learning cognitive/functional scores as determined by the Chow test (non-APOE p values include $p = 3.14 \times 10^{-5}$, 1.21×10^{-4}, 1.22×10^{-6}, respectively). Baseline tau measurements were associated with sex-based differences in the ADAS13, CDRSB, FAQ, MMSE, and RAVLT learning cognitive/functional scores as determined by the Chow test (non-APOE P values include $p = 1.26 \times 10^{-8}$, $p = 6.69 \times 10^{-6}$, $p = 4.44 \times 10^{-6}$, $p = 2.59 \times 10^{-5}$, and $p = 2.01 \times 10^{-6}$, respectively). Baseline p-tau measurements were associated with sex-based differences in the ADAS13, CDRSB, FAQ, MMSE, and RAVLT learning cognitive/functional scores as determined by the Chow test (non-APOE p values include $p = 5.08 \times 10^{-8}$, $p = 1.77 \times 10^{-5}$, $p = 1.44 \times 10^{-5}$, $p = 6.07 \times 10^{-5}$, and $p = 4.50 \times 10^{-6}$, respectively).

3.3. Post-hoc bootstrapping analysis

To confirm the significance of sex-related differences in quantitative-trait-related predictions, the proposed relationships for each noted quantitative trait and cognitive and functional outcome association were bootstrapped ($n = 600$) (Wilcox, 2010). The resulting regression coefficients from the bootstrapped analysis are shown in Fig. 2.

Large differences can be seen in the male versus female regression coefficient distributions associated many of the quantitative trait-cognitive and functional outcome pairs with the lowest Chow test p values (including the fusiform gyrus and midtemporal cortex quantitative traits predicting each of the 5 cognitive and functional outcomes using both baseline and month 24 data). Baseline and month 24 plots for each respective quantitative trait-cognitive and functional outcome pair also have a similar directionality: for instance, the fusiform gyrus-ADAS13 baseline and m24 plots similarly note a smaller mean regression coefficient in female participants. Additionally, there are some difference between the male and female regression coefficients of tau quantitative trait-cognitive and functional outcome pairs; however, the difference is less stark than the imaging pairs (i.e., the fusiform gyrus and midtemporal cortex-related trait-outcome pairs).

3.4. Post-hoc linear models

To visually confirm the fit of the regression coefficients calculated in the bootstrapped analyses and the sex-based differences found via the Chow test, a series of scatterplots were made (Fig. 3). These plots conveniently depict a sex-difference-associated change in the regression coefficients as differences in the slope of the male and female linear regression models' plotted lines. Individual points (male in orange, female in blue) represent data from the QT-PAD cohort; for each patient, their quantitative trait is plotted against an adjusted cognitive and functional outcome. The adjusted cognitive score is calculated by subtracting the intercept from the quantitative trait-cognitive and functional outcome regression model (using all participants, as opposed to using 2 different intercepts from the male and female participants) from the participant's original cognitive and functional outcome score. (Plotting the adjusted cognitive or functional score allows for a more informative visualization—slope differences between the male and female models are significantly more apparent.) Then, regression lines for the quantitative trait-cognitive and functional outcome were plotted using the data from all participants (in green), male participants (in orange), and female participants (in blue).

In the scatterplots shown in Fig. 3, sex-related differences can be seen via the different slopes of the male, female, and cumulative regression lines. The slope differences are especially noticeable in the fusiform gyrus plots (the first 2 rows) and the midtemporal cortex plots (the next 2 rows) but less so in the tau plots. Additionally, the individual-sex regression lines better represent trends in their respective strata than the cumulative regression line does for members of either sex.

3.5. Post-hoc ANOVA

To confirm the sex-based differences in the data, ANOVA was performed to assess the significance of differences in the means. Fusiform gyrus, midtemporal cortex, and tau quantitative traits data from male and female participants at both time points were used (Fig. 4). ANOVA found statistically significant differences between male and female patients in the baseline and month 24 fusiform gyrus trait as well as baseline tau.

3.6. Post-hoc comparison association analyses

To establish the utility of using 2 biological-sex-based models instead of factoring biological sex as a covariate, 3 sets of association analyses (using all male participants, using all female participants, using all participants) predicting a cognitive score using a quantitative trait were performed. Fig. 5 displays the results (i.e., p-values) of these analyses side-by-side for direct comparison. Data from both relevant time points (baseline visit and month 24 visit) were used. The large majority of trait-outcome associations were deemed statistically significant ($p < 1.52 \times 10^{-4}$) when using all patient data. However, doing so glosses over differences between the male and female sex. For example, with the Whole Brain Volume - RAVLT learning association using baseline data, the association is statistically significant in female participants only.

4. Discussion

The Chow test has been successfully used to identify quantitative trait-cognitive and functional outcome relationships with statistically significant differences between the male and female sexes. The most striking disparities occur in relationships involving imaging traits pertaining to regions of interest such as the fusiform gyrus and midtemporal cortex. There were also notable sex differences in CSF total tau and phosphorylated tau relationships.

The magnitude of the sex-based differences has been confirmed and visualized through a variety of visualizations (violin plots, scatterplots)

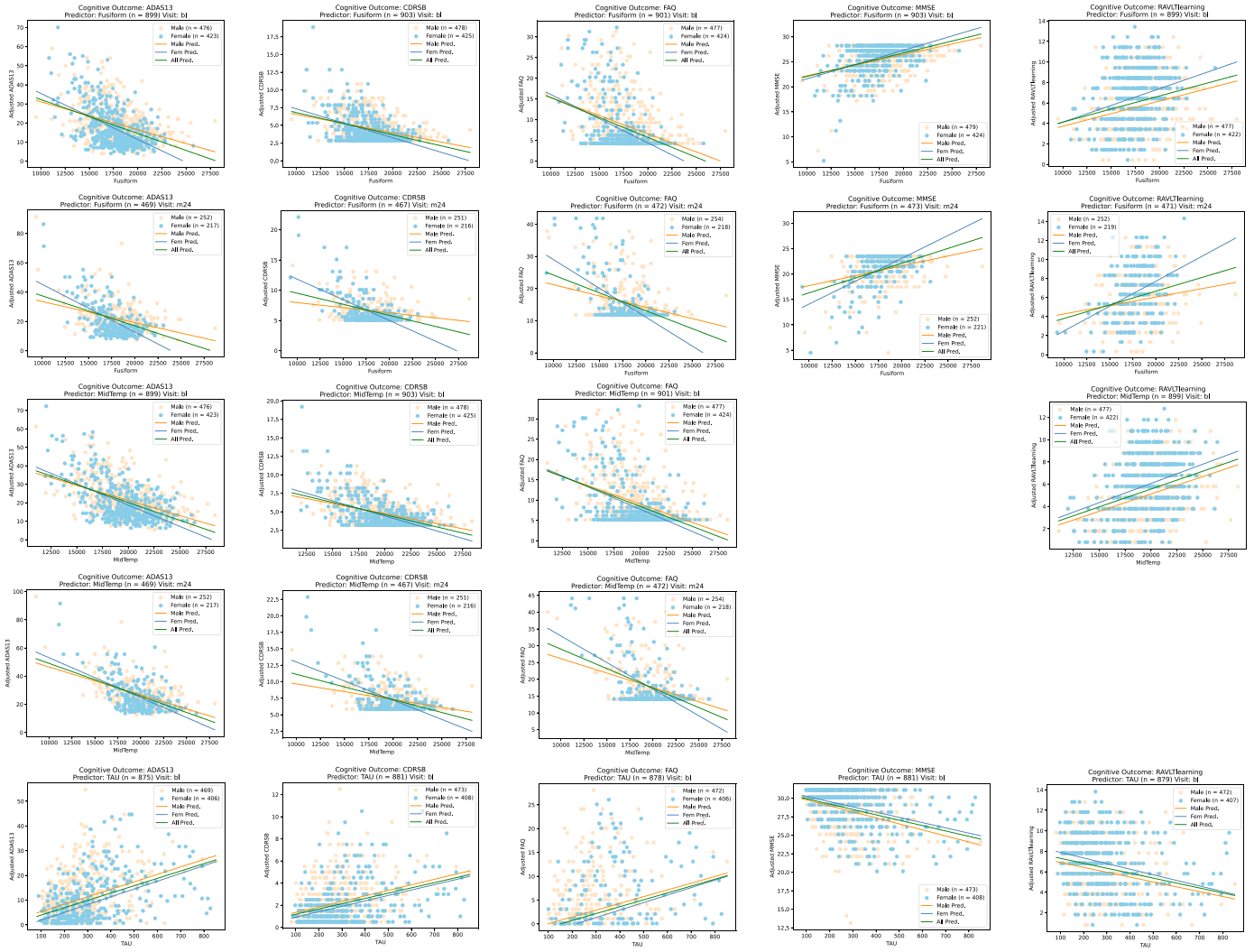

Fig. 3. Example linear models learned from male subjects, female subjects, and all subjects, respectively, using baseline and month 24 data. A select number of top significant correlations from the Chow tests (see Fig. 1) were chosen; these are the same relationships as those chosen in Fig. 5. Each plot depicts a unique predictor-outcome relationship assessed. The horizontal axis of each plot indicates the CSF/imaging predictor (imaging features are in units of cubic centimeters). The vertical axis of each plot indicates an age-and-education-adjusted cognitive/functional outcome. Male participants' data are represented with orange dots and the general regression-fitted trend visualized with an orange trend line. Female participants' data are represented with blue dots and the general regression-fitted trend visualized with a blue trend line. The overall trend line for all participants is visualized with a green trend line. For ease of viewing, plots have been sorted into rows and columns based on the specific predictor and outcome utilized. Importantly, for a given predictor-outcome pair, these visualizations depict differences in the distribution of male/female participants and show the differences in the trend lines (note the varying slopes and intercepts between lines of different colors); the Chow test is valuable for identifying these differences in predictor-outcome relationships.

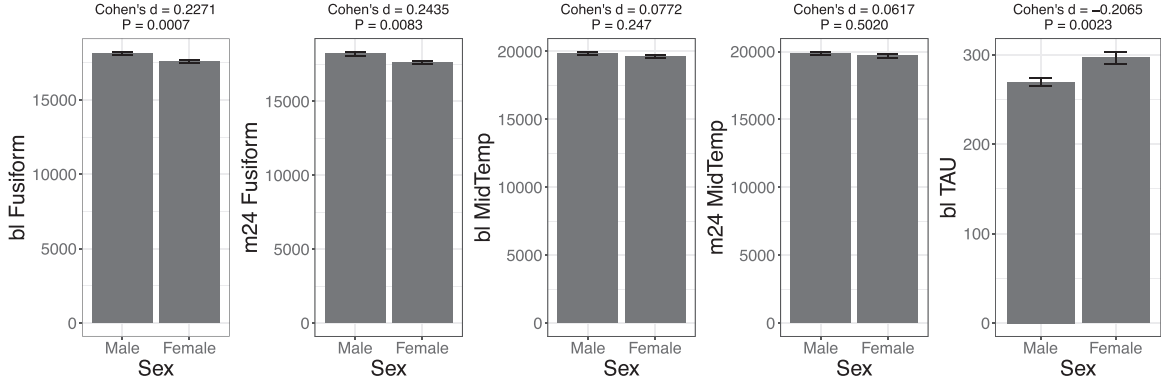

Fig. 4. ANOVA to confirm the significance of the difference in AD biomarkers between sex-stratified subpopulations of ADNI QT-PAD samples. Horizontal axis depicts subpopulation (male or female); vertical axis describes the predictor biomarker; vertical error bars depict the standard error of data; title reports Cohen's d and 2-tailed t-test *p* value measuring significance and magnitude of differences between reported means.

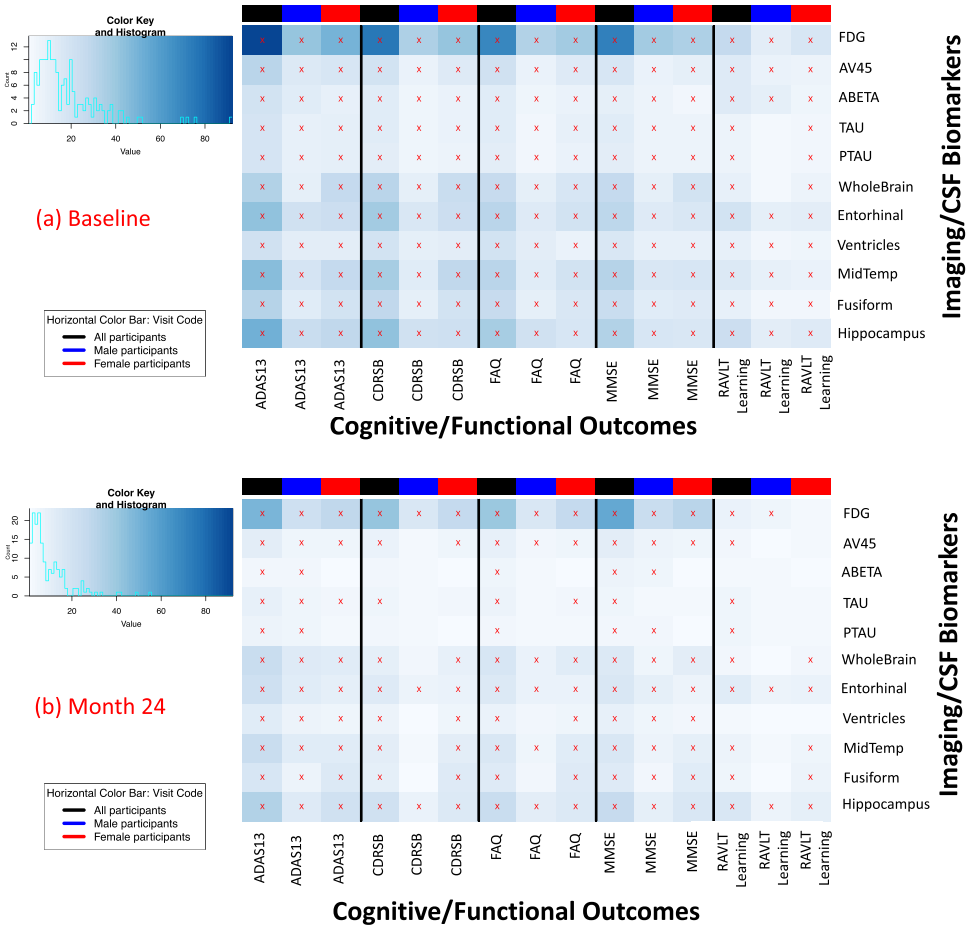

Fig. 5. Heat map showing results of confirmatory association tests. Regression: associate functional and cognitive and functional outcomes (horizontal axis) using a variety of imaging and CSF biomarkers (vertical axis) when factoring for age, years of education, and APOE4 status as covariates. Horizontal color bar represents population used (male, female, or all subjects) and shade of cells denotes relative-log(Association Test p). Significance was determined by a Bonferroni threshold ($p < 1.52 \times 10^{-4}$) with significant relationships denoted as X.

and post-hoc comparative analyses (bootstrapping, ANOVA, sex-stratified association tests). Bootstrapping analysis visually confirms that differences in regression coefficients for trait-outcome relationships exist between female and male-participant populations. The small difference in absolute effect sizes (i.e., regression coefficients) between males and females is likely an artifact of the scaling units we used. However, statistically these differences are significant, which suggests a sex disparity. Scatterplot visualizations confirm this intuition by visually depicting how sex-specific best fit lines better fit populations of male and female participants; therefore, there likely exist significant biological effects of biological sex on the identified relationships. ANOVA results confirm that some identified CSF/imaging biomarkers may have inherent biological differences that proceed to influence the difference in the Chow test-identified trait-outcome relationships. Ad-hoc sex-stratified association tests work to confirm that factoring biological sex as a covariate would miss the insignificance of key associations in male versus female participants. Therefore, to obtain optimal predictions, it is very necessary to embrace the stratified approach utilized by the Chow test.

Additionally, significant sex-based differences in the QT-PAD quantitative imaging and CSF traits have been well-documented in the literature, corroborating the abilities of a stratified approach to more accurately model indicators of AD. As visualized by the relationship-representative scatterplots (Fig. 3) and represented by the violin plots of differing regression coefficients (Fig. 2), constant changes in biomarkers (i.e., a noted gray matter loss, a noted increase in measured amyloid burden) are associated with differing effects on adjusted cognition/functional scores in male versus female participants. For instance, with the same amount of middle temporal gyrus gray matter loss, performance loss in the FAQ functional score is faster in female than male participants. Therefore, female patients can be seen as more vulnerable for Alzheimer's disease.

As noted in Section 3, although many relationships are similarly significant in the APOE and non-APOE Chow tests, the APOE Chow test results are slightly less significant. This difference of significance between APOE and non-APOE relationships assessed by the Chow test implies that APOE explains some of the identified sex differences in AD-relevant relationships. Previous work corroborates this intuition: studies have identified a significant interaction between the biological effects of APOE4 and self-reported biological sex (Altmann et al., 2014; Sampedro et al., 2015; Shinohara et al., 2016). Although the interaction between APOE4 and biological sex may account for some of the difference in significance between the APOE4 and non-APOE4 Chow test results, significant sex differences remain in many of the noted relationships.

A majority of the quantitative trait-cognitive and functional outcome associations with statistically significant Chow test p values specifically involve the ADAS13, CDRSB, and FAQ cognitive and functional outcomes. This is likely due to the increased granularity of these 3 clinical scales as well as the larger sample sizes of these scales, given their prevalence in the clinic. Given their ties to clinical measures of cognitive impairment, these scores (as well as other clinical/pathophysiological scales such as the MOCA or Braak stage score) may also serve as ideal proxies for AD diagnosis and allow us to investigate the changing ability of our 11 non-cognitive biomarkers to directly predict AD diagnosis.

Many of the sex-based differences in key quantitative traits found in our analyses were also reported in the literature. These findings in the literature corroborate the utility of the Chow test in finding sex-based differences in relationships between quantitative traits and clinical outcomes. All 20 quantitative trait-cognitive and functional outcome associations (including both time points, the APOE and non-APOE trials, and 5 cognitive and functional outcomes) involving the FreeSurfer fusiform gyrus quantitative trait had a statistically significant Chow test *p* value. For example, Lotze et al. (2019) found in a cohort of 2838 individuals of an age range from 21–90 years that male participants had a larger gray matter volume in the fusiform gyrus (Cohen's d of 0.40). These significant differences may lead to the differing regression coefficients involving this imaging phenotype predicting a variety of cognitive and functional outcomes. Given its role in multiple complex cognitive pathways including face recognition, word recognition, and within-category identification, it is understandable this brain region is very important in AD. Therefore, AD biomarkers involving the fusiform gyrus warrant investigation perhaps as therapeutic targets or early predictors of disease.

Similarly, all 20 quantitative trait-cognitive and functional outcome associations (including both time points, the APOE and non-APOE trials, and 5 cognitive and functional outcomes) involving the FreeSurfer entorhinal cortex quantitative trait had a statistically significant Chow test *p* value. The entorhinal cortex is another brain region that likely plays a functional role in AD pathogenesis. Normally playing a key role in cognitive processes including memory, navigation, and the perception of time, the entorhinal cortex suffers a significant loss of neurons during the first stages of AD and serves as a primary location for neurofibrillary tangles. Therefore, it is understandable this region may be associated with several AD biomarkers. Additionally, other studies have documented significant differences in the entorhinal cortex that occur on the basis of sex, corroborating the findings of this study (Arsenault et al., 2020).

Additionally, all 20 quantitative trait-cognitive and functional outcome associations (including both time points, the APOE and non-APOE trials, and 5 cognitive and functional outcomes) involving the FreeSurfer temporal lobe quantitative trait had a statistically significant Chow test *p* value. The temporal lobe also plays a large role in memory, processing sensory information for the retention of visual memory, language comprehension, and emotion association. A component of the temporal lobe denoted the medial temporal lobe is also very important for episodic and spatial memory consolidation and storage (Cutsuridis and Yoshida, 2017). Therefore, it is understandable that alterations in this brain region may be associated with lower cognitive and functional outcome scores associated with AD. A previous study has also found that the volumes of the temporal lobes varies between male and female participants in the context of AD diagnosis, confirming the sex differences noted in these analyses (Cutsuridis and Yoshida, 2017). An additional study identified cortical thickness declines in the bilateral temporal regions of male participants throughout their progression from normal control to Alzheimer's dementia diagnosis while female participants had more stable cortical thickness until mild cognitive impairment but sharper declines from mild cognitive impairment to Alzheimer's dementia (Cieri et al., 2022).

The total-tau biomarker used in this study is specifically associated with neurodegeneration or neuronal injury while phosphorylated tau (or p-tau) is associated with the accumulation of neurofibrillary tangles (Jack et al., 2018). Although these biomarkers are implicated with distinct processes, both tau and p-tau are likely highly significant to AD etiology and likely to have an effect—direct or indirect—on the cognitive and functional outcomes examined in this study. There exist multiple significant trait-outcome relationships involving both tau and p-tau, implying that the processes of both neurodegeneration/neuronal injury and the accumulation of neurofibrillary tangles may have differential effects on female versus male participants. Multiple studies have confirmed significant sex differences in total tau and phosphorylated tau biomarkers through autopsy-based (Oveisgharan et al., 2018), neuroimaging-based (Buckley et al., 2019), and CSF-based studies (Altmann et al., 2014; Hohman et al., 2018) in the study of AD. Our results confirm the significance of studying these fluid-based biomarkers in order to clarify the magnitude of these sex-based differences. Please recall that both tau and p-tau were evaluated in the Chow tests but only tau was evaluated in post-hoc analyses.

A small minority of the significant relationships involved AV45; no significant relationships involve FDG nor ABETA. This general lack of significance is surprising given the tendency for 18F-AV-45 to serve as a biomarker for beta-amyloid plaque buildup (Choi et al., 2009), FDG-PET to identify significant functional changes in the brain (Marcus et al., 2014), and beta-amyloid imaging to accurately estimate a participant's amyloid-beta burden (Chun, 2018). Additionally, the literature suggests the presence of some sex-based effects with the amyloid-beta biomarkers: a previous study found a significant interaction between sex and amyloid-beta-42 on declines in memory and executive function (Koran et al., 2017). The relatively few number of identified relationships is likely attributable to low sample sizes and large proportions of missing data in the ADNI cohort; although the handful of expected AV45 relationships identified are reassuring, perhaps a larger and more diverse cohort of individuals would allow for the identification of more significant relationships.

Interestingly, an almost-universal majority of the stronger biomarker-clinical outcome relationships tend to be in female participants. This is visualized by the larger-magnitude regression coefficients and steeper slopes of trait-outcome relationship best-fit lines corresponding to female-participant populations. These results may imply that female participants are therefore more at-risk for Alzheimer's disease and/or related cognitive decline, which is generally corroborated with the significant sex-based disparities in AD risk reported in the literature (Ferretti et al., 2018).

The different *p* values for certain associations between male and female subjects highlight the usefulness of stratified analyses; only a stratified analysis would accurately reflect differences in male and female AD pathophysiology that must be accounted for in models and patient care. If sufficient data allow, researchers should use independent models for male and female patients when creating models for key AD biomarkers associated with AD diagnosis or pathology. This is because combining all subjects in a single analysis and factoring patient sex as a covariate—the standard approach in large genome-wide association studies and machine learning approaches—may overemphasize associations in just male subjects (as is the case with predicting levels of p-tau using FreeSurfer whole brain volume with month 24 data) or just female subjects (as is the case with predicting CDRSB scores using FreeSurfer whole brain volume with month 24 data). Only in stratified analyses are these differential associations visible.

The Chow test is useful for identifying situations in which the predictive power of certain AD biomarkers varies between male and female subjects. Many of the proposed relationships with significant sex-based differences seen in the association studies confirmed by the violin plots (Fig. 2), linear models (Fig. 3), and ANOVA (Fig. 4) were noted by the Chow test. As such, the Chow test may serve as a useful first step in determining which relationships may be worth studying in each sex group. This intuition can also help create more robust models for AD diagnosis, allowing researchers and clinicians to account for the differences between male and female participants in AD research and care.

One concern about the specific results of this study includes the overrepresentation of White, non-Hispanic participants. This is one weakness of the ADNI cohorts. Future investigations may involve using more diverse cohorts to determine if similar sex disparities exist in additional patient populations. Additionally, the Chow test can be used in conjunction with stratified approaches to locate additional AD-related disparities when stratifying on other demographic variables including race or socioeconomic status.

Additionally, there were relatively few measurements available for

analysis at the month 24 visit code. It may be possible that the lower number of samples has deflated the significance of any noted differences between male and female patients. However, given the significance of relationships analyzing sex-mediated differences in the AV45 biomarker using sufficiently-powerful and numerous baseline data, one can conclude that differences involving the AV45 phenotype and other imaging biomarkers (i.e., the FDG-PET biomarker) warrant further analysis in an independent cohort.

Biological sex in the current study was measured via self-report. Future studies that measure other aspects of sex, such as genetic sex, may help to further inform this area of research. Additionally, there is a distinct lack of gender-based analysis; given that sex and gender are distinct, future work may also examine gender-dependent effects (Heidari et al., 2016).

Lastly, there may be some concerns regarding the sample size used in our study as well as the male/female balance. Ideally, it would be possible to utilize a full data set with samples from multiple AD biomarkers with a suitably large cohort with equal numbers of male and female patients. Future studies would involve efforts to verify the results of these analyses, perhaps in a different cohort of patients entirely or via use of updated statistics and biomarker measurements from the ADNI data.

The primary strengths of this study are (1) the novel use of the Chow test and (2) identification of differences in useful trait-outcome relationships. To the best of our knowledge, finding and quantifying sex disparities as performed in this manuscript is a novel application of the Chow test. The quantification of sex differences in imaging traits and CSF traits can improve current models of AD, allowing for more precise predictions of disease progression.

The Chow test also has the unique methodological advantage of quantifying the magnitude of sex-based differences. This is a useful measure not ordinarily attainable via standard stratified analyses (see Fig. 5). Although it is possible to notice a significant sex-based difference via direct comparison of p values from different strata, as is done in Fig. 5, the Chow test's utilization of the F statistic can facilitate the comparison of different relationships (i.e., comparing individual quantitative trait-cognitive and functional outcome pairs). In addition, sex interaction models may allow for the identification of trends only present in an individual sex group but might not be able to detect subtle (but significant!) sex-based differences in trends that exist in both male and female populations. Chow test is specifically designed to use a single model to detect such difference, and thus is more powerful for this specific purpose. Therefore, the Chow test differs from—and therefore is well-poised to accompany—other sex-stratified and sex-interaction studies because it can evaluate the significance of the difference between the effect sizes of a pair of models.

The differences noted in our analyses here have the potential to inform the models used to predict AD pathogenesis: although using patient biological sex is a decent first step, given the magnitude of these differences (with $p < 10^{-6}$ in some cases), it may be necessary to specifically account for sex-related differences in AD biomarkers in future analyses. In doing so, our analysis provides guidance for researchers searching for neurobiological factors that may explain sex-based differences in cognition and daily functioning.

5. Conclusions

While prior studies mainly investigated sex effects on AD biomarkers, this work examined how sex modified the effects of imaging and CSF biomarkers on AD cognitive and functional outcomes at a variety of time points. Chow tests were performed to determine the magnitude and statistical significance of sex differences and found several significant differences involving imaging and CSF biomarkers such as the Fusiform and MidTemp FreeSurfer outcomes and measurements of Tau protein. These results are also consistent with prior results showing significant sex differences in imaging/CSF biomarkers, highlighting the viability of our approach in measuring the quantitative effect(s) sex has on both key AD-related biological measures. This novel application of the Chow test to quantify the magnitude of sex differences in addition to the analysis of AD-relevant cognitive scores in lieu of AD diagnosis have enhanced the robustness of statistical analyses. The imaging/CSF quantitative trait-cognitive score pairs highlighted by the Chow tests and post-hoc analyses should be studied in more detail from a biological perspective to confirm the presence of such effects. Future directions may include applying the intuition gained from these analyses to help build fairer sex-stratified predictive models. Such models may promote precision medicine and help elucidate how biological factors drive the sex-based pathological disparity in AD.

CRediT authorship contribution statement

Brian N. Lee: Conceptualization, Methodology, Formal analysis, Validation, Writing - Original Draft. **Junwen Wang**: Software, Writing - Review & Editing. **Molly A. Hall**: Conceptualization, Writing - Review & Editing. **Dokyoon Kim**: Conceptualization, Writing - Review & Editing. **Shana D. Stites**: Conceptualization, Writing - Review & Editing. **Li Shen**: Supervision, Conceptualization, Methodology, Writing - Review & Editing.

Disclosure statements

The authors have no actual or potential conflicts of interest.

Acknowledgements

This work is supported in part by the Alzheimer's Association (AARF-17–528934) and the National Institutes of Health (K23AG065442, U24AG058556, R01 AG071470, U01 AG066833, U01 AG068057, and R01 LM013463).

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12–2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.neurobiolaging.2023.10.002.

References

ADNI Team, 2021.Alzheimer's Disease Neuroimaging Initiative. Available at ⟨http://adni.loni.usc.edu/⟩ Accessed 2021-08-22.

Altmann, A., Tian, L., Henderson, V.W., Greicius, M.D., 2014. Sex modifies the APOE-related risk of developing Alzheimer disease. Ann. Neurol. 75, 563–573. https://doi.org/10.1002/ANA.24135.

Arsenault, D., Tremblay, C., Emond, V., Calon, F., 2020. Sex-dependent alterations in the physiology of entorhinal cortex neurons in old heterozygous 3xTg-AD mice. Biol. Sex Differ. 11, 1–19. https://doi.org/10.1186/s13293-020-00337-0.

Buckley, R.F., Mormino, E.C., Rabin, J.S., Hohman, T.J., Landau, S., Hanseeuw, B.J., Jacobs, H.I.L., Papp, K.V., Amariglio, R.E., Properzi, M.J., Schultz, A.P., Kirn, D., Scott, M.R., Hedden, T., Farrell, M., Price, J., Chhatwal, J., Rentz, D.M., Villemagne, V.L., Johnson, K.A., Sperling, R.A., 2019. Sex differences in the association of global amyloid and regional tau deposition measured by positron emission tomography in clinically normal older adults. JAMA Neurol 76, 542–551. https://doi.org/10.1001/JAMANEUROL.2018.4693.

Choi, S.R., Golding, G., Zhuang, Z., Zhang, W., Lim, N., Hefti, F., Benedum, T.E., Kilbourn, M.R., Skovronsky, D., Kung, H.F., 2009. Preclinical properties of 18F-AV-45: a PET agent for abeta plaques in the brain. J. Nucl. Med. 50, 1887–1894.

Chow, G.C., 1960. Tests of equality between sets of coefficients in two linear regressions. Econometrica 28, 591. https://doi.org/10.2307/1910133.

Chun, K.A., 2018. Beta-amyloid imaging in dementia. Yeungnam Univ. J. Med. 35, 1–6.

Cieri, F., Zhuang, X., Cordes, D., Kaplan, N., Cummings, J., Caldwell, J., 2022. Relationship of sex differences in cortical thickness and memory among cognitively healthy subjects and individuals with mild cognitive impairment and Alzheimer disease. Alzheimers Res. Ther. 14, 36.

Cutsuridis, V., Yoshida, M., 2017. Editorial: Memory processes in medial temporal lobe: experimental, theoretical and computational approaches. Front. Syst. Neurosci. 11, 19. https://doi.org/10.3389/FNSYS.2017.00019.

Dubois, B., Hampel, H., Feldman, H., Scheltens, P., Aisen, P., Andrieu, S., Bakardjian, H., Benali, H., Bertram, L., Blennow, K., Broich, K., Cavedo, E., Crutch, S., Dartigues, J. F., Duyckaerts, C., Epelbaum, S., Frisoni, G.B., Gauthier, S., Genthon, R., Gouw, A.A., Habert, M.O., Holtzman, D.M., Kivipelto, M., Lista, S., Molinuevo, J.L., O'Bryant, S. E., Rabinovici, G.D., Rowe, C., Salloway, S., Schneider, L.S., Sperling, R., Teichmann, M., Carrillo, M.C., Cummings, J., Jack, C.R., 2016. Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. Alzheimers Dement. 12, 292–323. https://doi.org/10.1016/J.JALZ.2016.02.002.

Ferretti, M.T., Iulita, M.F., Cavedo, E., Chiesa, P.A., Dimech, A.S., Chadha, A.S., Baracchi, F., Girouard, H., Misoch, S., Giacobini, E., Depypere, H., Hampel, H., 2018. Sex differences in Alzheimer disease—the gateway to precision medicine. Nat. Rev. Neurol. 14, 457–469. https://doi.org/10.1038/s41582-018-0032-9.

Fu, J., Ruan, H., Zheng, H., Cai, C., Zhou, S., Wang, Q., Chen, W., Fu, W., Du, J., 2019. Impact of old age on resectable colorectal cancer outcomes. PeerJ 7, e6350. https://doi.org/10.7717/PEERJ.6350.

González Zarzar, T., Lee, B., Coughlin, R., Kim, D., Shen, L., Hall, M.A., 2022. Sex differences in the metabolome of Alzheimer's disease progression. Front. Radiol. 2, 782864 https://doi.org/10.3389/fradi.2022.782864.

Heidari, S., Babor, T.F., De Castro, P., Tort, S., Curno, M., 2016. Sex and gender equity in research: rationale for the SAGER guidelines and recommended use. Res. Integr. Peer Rev. 1, 2.

Hohman, T.J., Dumitrescu, L., Barnes, L.L., Thambisetty, M., Beecham, G., Kunkle, B., Gifford, K.A., Bush, W.S., Chibnik, L.B., Mukherjee, S., De Jager, P.L., Kukull, W., Crane, P.K., Resnick, S.M., Keene, C.D., Montine, T.J., Schellenberg, G.D., Haines, J. L., Zetterberg, H., Blennow, K., Larson, E.B., Johnson, S.C., Albert, M., Bennett, D.A., Schneider, J.A., Jefferson, A.L., 2018. Sex-specific association of apolipoprotein E with cerebrospinal fluid levels of tau. JAMA Neurol. 75, 989–998. https://doi.org/10.1001/JAMANEUROL.2018.0821.

Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T.D., Elliott, M.A., Ruparel, K., Hakonarson, H., Gur, R.E., Gur, R.C., Verma, R., 2014. Sex differences in the structural connectome of the human brain. Proc. Natl. Acad. Sci. 111, 823–828. https://doi.org/10.1073/PNAS.1316909110.

Jack, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J.L., Montine, T., Phelps, C., Rankin, K.P., Rowe, C.C., Scheltens, P., Siemers, E., Snyder, H.M., Sperling, R., Elliott, C., Masliah, E., Ryan, L., Silverberg, N., 2018. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement. 14, 535–562.

Kang, J.H., Korecka, M., Figurski, M.J., Toledo, J.B., Blennow, K., Zetterberg, H., Waligorska, T., Brylska, M., Fields, L., Shah, N., Soares, H., Dean, R.A., Vanderstichele, H., Petersen, R.C., Aisen, P.S., Saykin, A.J., Weiner, M.W., Trojanowski, J.Q., Shaw, L.M., 2015. The alzheimer's disease neuroimaging initiative 2 biomarker core: a review of progress and plans. Alzheimers. Dement. 11, 772–791.

Koran, M.E.I., Wagener, M., Hohman, T.J., Alzheimer's Neuroimaging Initiative, 2017. Sex differences in the association between AD biomarkers and cognitive decline. Brain Imaging Behav. 11, 205–213.

Lane, C.A., Hardy, J., Schott, J.M., 2018. Alzheimer's disease. Eur. J. Neurol. 25, 59–70. https://doi.org/10.1111/ENE.13439.

Laws, K., Irvine, K., Gale, T., 2018. Sex differences in Alzheimer's disease. Curr. Opin Psychiatry 31, 133–139. https://doi.org/10.1097/YCO.0000000000000401.

Li, R., Singh, M., 2014. Sex differences in cognitive impairment and Alzheimer's disease. Front Neuroendocrinol. 35, 385–403. https://doi.org/10.1016/J.YFRNE.2014.01.002.

Liang, W., He, J., Shen, Y., Shen, J., He, Q., Zhang, J., Jiang, G., Wang, Q., Liu, L., Gao, S., Liu, D., Wang, Z., Zhu, Z., Ng, C.S.H., Liu, C.-C., Petersen, R.H., Rocco, G., D'Amico, T., Brunelli, A., Chen, H., Zhi, X., Liu, B., Yang, Y., Chen, W., Zhou, Q., He, J., 2016. Impact of examined lymph node count on precise staging and long-term survival of resected non-small-cell lung cancer: a population study of the US SEER Database and a Chinese Multi-Institutional Registry. J. Clin. Oncol. 35, 1162–1170. https://doi.org/10.1200/JCO.2016.67.5140.

Lotze, M., Domin, M., Gerlach, F.H., Gaser, C., Lueders, E., Schmidt, C.O., Neumann, N., 2019. Novel findings from 2,838 adult brains on sex differences in gray matter brain volume. Sci. Rep. 9, 1671. https://doi.org/10.1038/S41598-018-38239-2.

Marcus, C., Mena, E., Subramaniam, R.M., 2014. Brain PET in the diagnosis of Alzheimer's disease. Clin. Nucl. Med. 39, e413–e422 quiz e423–426.

National Institutes of Health, 2018. Research on current topics in Alzheimer's disease and its related dementias (R01 Clinical Trial Optional). Available at: ⟨https://grants.nih.gov/grants/guide/notice-files/NOT-AG-18-051.html⟩ Accessed 2018-11-29.

, 2019National Institutes of Health, 2019. Summary of the NIH office of research on women's health GWAS, sex, and chromosomes think tank on February 27, 2019. Available at: ⟨https://orwh.od.nih.gov/sites/orwh/files/docs/2019_SABV_in_GWAS_Think_Tank_Summary.pdf⟩ Accessed 2019-02-27.

Nebel, R.A., Aggarwal, N.T., Barnes, L.L., Gallagher, A., Goldstein, J.M., Kantarci, K., Mallampalli, M.P., Mormino, E.C., Scott, L., Yu, W.H., Maki, P.M., Mielke, M.M., 2018. Understanding the impact of sex and gender in Alzheimer's disease: a call to action. Alzheimers Dement. 14, 1171–1183. https://doi.org/10.1016/J.JALZ.2018.04.008.

Oveisgharan, S., Arvanitakis, Z., Yu, L., Farfel, J., Schneider, J.A., Bennett, D.A., 2018. Sex differences in Alzheimer's disease and common neuropathologies of aging. Acta Neuropathol. 136, 887–900. https://doi.org/10.1007/S00401-018-1920-1.

QT-PAD Team, 2021. Alzheimer's disease modelling challenge. Available at ⟨http://www.pi4cs.org/qt-pad-challenge⟩ Accessed 2021-08-22.

Ruigrok, A.N.V., Salimi-Khorshidi, G., Lai, M.C., Baron-Cohen, S., Lombardo, M.V., Tait, R.J., Suckling, J., 2014. A meta-analysis of sex differences in human brain structure. Neurosci. Biobehav. Rev. 39, 34–50. https://doi.org/10.1016/J.NEUBIOREV.2013.12.004.

Sampedro, F., Vilaplana, E., de Leon, M.J., Alcolea, D., Pegueroles, J., Montal, V., Carmona-Iragui, M., Sala, I., Sánchez-Saudinos, M.B., Antón-Aguirre, S., Morenas-Rodríguez, E., Camacho, V., Falcón, C., Pavía, J., Ros, D., Clarimón, J., Blesa, R., Lleó, A., Fortea, J., 2015. APOE-by-sex interactions on brain structure and metabolism in healthy elderly controls. Oncotarget 6, 26663–26674.

Shen, L., Kim, S., Risacher, S.L., Nho, K., Swaminathan, S., West, J.D., Foroud, T., Pankratz, N., Moore, J.H., Sloan, C.D., Huentelman, M.J., Craig, D.W., DeChairo, B. M., Potkin, S.G., Jack, C.R., Weiner, M.W., Saykin, A.J., 2010. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. Neuroimage 53, 1051–1063. https://doi.org/10.1016/j.neuroimage.2010.01.042.

Shinohara, M., Murray, M.E., Frank, R.D., Shinohara, M., DeTure, M., Yamazaki, Y., Tachibana, M., Atagi, Y., Davis, M.D., Liu, C.C., Zhao, N., Painter, M.M., Petersen, R. C., Fryer, J.D., Crook, J.E., Dickson, D.W., Bu, G., Kanekiyo, T., 2016. Impact of sex and APOE4 on cerebral amyloid angiopathy in Alzheimer's disease. Acta Neuropathol. 132, 225–234.

Wilcox, R., 2010. Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy. Springer,, New York. ⟨https://books.google.ie/books?id=PUkOBwAAQBAJ⟩.