Comprehensive evaluation of AT(N) imaging biomarkers for predicting cognition

Tom Earnest[1], Braden Yang[1], Deydeep Kothapalli[1], Aristeidis Sotiras[1,2], for the Alzheimer's Disease Neuroimaging Initiative*

[1] Mallinckrodt Institute of Radiology, Washington University School of Medicine in St Louis; 4525 Scott Ave, Saint Louis, MO 63110

[2] Institute for Informatics, Data Science & Biostatistics, Washington University School of Medicine in St Louis; 660 S. Euclid Ave, Campus Box 8132, Saint Louis, MO 63110

* 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

Abstract

Background and Objectives: Imaging biomarkers enable *in vivo* quantification of amyloid, tau, and neurogenerative pathologies that develop in Alzheimer's Disease (AD). Interest in imaging biomarkers has led to a wide variety of biomarker definitions, some of which potentially offer less predictive value than others. We aimed to assess how different operationalizations of AD imaging biomarkers affect prediction of cognition.

Methods: We included individuals from ADNI who underwent amyloid-PET ([^{18}F]-Florbetapir), tau-PET ([^{18}F]-Flortaucipir), and volumetric MRI imaging. We compiled a large collection of imaging biomarker definitions (42 in total) spanning different pathologies (amyloid, tau, neurodegeneration) and variable types (continuous, binary, non-binary categorical). Using cross-validation, we trained regression models to predict neuropsychological performance, both globally and across different subdomains (Phenotype Harmonization Consortium composites), using different combinations of biomarkers. We also compared these biomarker models to support vector machines (SVMs) trained to predict cognition directly from imaging regions of interest. In a subsample of individuals with CSF biomarker readouts, we repeated experiments comparing the accuracy of models using imaging and fluid biomarkers. Additional analyses tested the predictive strength of imaging biomarkers when limited to specific clinical stages of disease (cognitive unimpaired vs. impaired) and when modeling longitudinal cognitive change.

Results: Our sample included 490 people (247 female) with a mix of no impairment (n=288), mild impairment (n=163), and dementia (n=39). While almost all biomarkers tested were predictive of cognitive performance, we observed substantial variability in accuracy, even for measures of the same pathology. Tau biomarkers were the single most accurate single predictors, though combination of biomarkers spanning multiple pathologies were more accurate overall. SVM models were generally more accurate than models using traditional biomarkers. Incorporating continuous or non-binary categorical biomarkers was beneficial only for tau and neurodegeneration, but not amyloid. Patterns of results were largely consistent when considering different clinical

stages of disease, neuropsychological domains, and longitudinal cognition. In the CSF subsample (n=246), imaging biomarkers strongly outperformed CSF versions for cognitive prediction.

Discussion: We demonstrated that different imaging biomarker definitions can lead to variability in downstream predictive tasks. Researchers should consider how their biomarker operationalizations may help or hinder the assessment of disease severity.

Introduction

The modern biological definition of Alzheimer's Disease (AD) relies on biomarkers.[1–3] Biomarkers can accurately quantify pathobiological disease processes which are specific to AD, particularly the aggregation of amyloid-beta (Aβ) plaques and the neocortical spread of tau neurofibrillary tangles. Importantly, biomarkers can detect and measure these pathologies prior to symptomatic onset. Because of their capabilities, biomarkers have been used in a variety of research settings including disease classification[4], cognitive forecasting[5], subtype identification[6], clinical trial stratification[7], disease staging[8,9], and more. Moreover, biomarkers are becoming increasingly important for clinical management of AD[2]. For instance, recently approved anti-Aβ treatments for AD require the presence of Aβ-pathology as assessed by biomarkers.

Interest in biological AD assessment has led to the creation of many AD-sensitive biomarkers which vary in terms of modality, underlying pathology, and statistical formulation. Idiosyncrasies of biomarker definitions may result in unwanted variability when applied for clinical and research uses. For example, estimated cut points for PET and CSF biomarker dichotomization are fairly application specific [10–12], and different approaches to pathological thresholding result in considerable variability for group assignment[13–16]. Less is known, however, about how variability in biomarker definitions affects prediction of cognition in AD. Identifying which specific biomarkers are most predictive of cognitive trajectories, particularly at different stages of disease, can provide insight into biological mechanisms of AD. Moreover, precise cognitive decline predictions are valuable for identifying candidates for early therapeutic interventions and for establishing meaningful cognitive endpoints in clinical trials. Despite these implications, investigations into the ramifications of different biomarker operationalizations remain limited. One previous study found that different biomarker definitions varied in their ability to predict longitudinal Mini-Mental State Examination (MMSE) scores, and that dichotomization hindered predictive power of some biomarkers relative to continuous values.[13] Similar analyses in separate cohorts with additional cognitive measures are needed to confirm and extend these findings,

particularly to establish optimal biomarker combinations for both prognostic accuracy and mechanistic insight.

Here, we developed a comprehensive set of neuroimaging measures (42 in total) covering the AD core biomarkers Aβ (A) and tau (T), as well as non-specific biomarkers of neurodegeneration ((N)). We systematically evaluated how different categories of biomarkers and individual variants differ in their ability to predict different cognitive outcomes. While we focused on cross-sectional cognition, we also extended analyses to measures of prospective longitudinal cognition and neuropsychological domains. We additionally incorporated machine learning to test how traditional biomarker approaches compare to methods which can detect more complex, multivariate patterns in imaging data. Finally, we tested multiple CSF biomarkers (20 definitions spanning 4 analytes) and compared their performance with imaging alternatives.

Methods

Participants

We selected a baseline, cross-sectional sample of Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with tau-PET, Aβ-PET, and structural MRI imaging data. Exclusion criteria were gaps between scans of greater than 1 year or missing values for any of the following variables: age, sex, *APOE* genotype, Clinical Dementia Rating® (CDR) status[17], Phenotype Harmonization Consortium (PHC) cognitive composite scores[18].

Standard protocol approvals, registrations, and patient consents

All participants provided informed written consent for participating in ADNI. Study protocols were approved by site-specific institutional review and ethical boards.

Image acquisition and processing

Detailed descriptions of imaging protocols are provided on the ADNI website[19]. Briefly, T1-weighted MRI acquisitions were collected on 3T scanners using an

accelerated MPRAGE sequence. Aβ-PET scans were acquired 50-70 minutes (4 frames × 5 minutes) after a 370 MBq (± 10%) injection of [^{18}F]-Florbetapir. Tau-PET scans were acquired 75-105 minutes (6 frames × 5 minutes) after a 370 MBq (± 10%) injection of [^{18}F]-Flortaucipir.

We accessed processed MRI and PET derivatives generated by the ADNI PET Core. A Freesurfer (v7.1.1) processing pipeline was applied to MRI scans to generate gray matter volumes within regions of interest (ROIs) of standard subcortical[20] and cortical atlases[21]. PET standardized uptake value ratios (SUVRs) were generated for these same ROIs after coregistration of each PET image to a contemporaneous MRI scan. Our analyses incorporated unilateral values from 68 cortical and 14 subcortical gray matter regions. Volumes were standardized relative to the intracranial volume. Aβ-PET uptakes were standardized to a whole cerebellum ROI, while tau-PET uptakes were standardized relative to an ROI containing inferior cerebellar gray matter[22].

Partial volume corrected (PVC) PET uptakes were available for tau (Geometric Transfer Matrix approach[23,24]) but not for Aβ. We used uncorrected SUVR values for most experiments, but we repeated some experiments with PVC-corrected tau SUVRs to evaluate the effect of PVC on cognitive prediction accuracy.

Cognitive and clinical assessments

Cognition was assessed using composite scores developed by the PHC[18]. We averaged the memory (PHC_{Memory}), executive functioning (PHC_{EF}), visuospatial (PHC_{Visual}), and language ($PHC_{Language}$) composites to create one global cognitive composite (PHC_{Global}). Composites are unitless factor loadings, with lower scores corresponding to more impairment.

CDR was used as a measure of dementia severity[17]. Subjects were assigned to the following groups based on CDR status: cognitively unimpaired (CU, CDR=0) or cognitively impaired (CI, CDR>=0.5).

Image-based biomarker definitions

We implemented a variety of biomarker definitions to use for predicting cognition. A full list of the biomarker definitions tested is provided in eTable 1. Biomarkers were categorized based on pathology (AT(N)) and variable type (binary [BIN], non-binary categorical [CAT], continuous [CON]). Lists of atlas regions used to form composites are provided in eTable2.

Continuous variables consisted of scalar MRI (volume) or PET (SUVR) measures in standard composite ROIs. For Aβ, continuous measures of Aβ included the average SUVR in a cortical summary region[25,26] (Aβ composite) and Centiloid[27]. Centiloids were provided by ADNI and derived from the Aβ composite using previously validated equations[28]. Continuous tau measures included the average uptakes in a meta-temporal (MT) composite region[11] and uptakes in ROIs corresponding to progressive Braak stages[29,30] (Braak I, Braak III/IV, Braak V/VI). Braak II was omitted due to off-target binding issues with flortaucipir[31,32]. Hippocampal volume and volume of the MT region were included as continuous assessments of neurodegeneration.

Binary predictors consisted of dichotomized versions of the continuous predictors listed above. There were three main methods tested for binarizing continuous variables: previously published cutoffs, Z-scoring, and Gaussian mixture modeling (GMM). Previously published cutoffs were included for the Aβ composite at the following SUVRs: 1.11[33], 1.24[34], 1.30[11], 1.42[11]. We also tested Centiloid cutoffs (15, 20, 25, 30) based on ranges reported in previous literature[28,35,36]. Z-scoring and GMMs were included as data-driven approaches for deriving cutoffs. These methods were applied to the Aβ composite SUVR, MT tau SUVR, MT volume, and hippocampal volume. Z-scores for each variable were computed relative to CU, Aβ-negative individuals (using an SUVR cutoff of 1.11 applied to the Aβ composite to determine Aβ-negativity, as recommended by the ADNI PET Core). Z-scores were dichotomized using cutoffs of 2 and 2.5 standard deviations away from the CU, Aβ-negative mean value. GMM binarization was implemented by fitting two-component Gaussian mixtures to the distribution of continuous variables. A cutoff point was estimated as the curve

intersection between the fitted Gaussians. GMMs were omitted for hippocampal volume, due to a lack of bimodal distribution.

Non-binary categorical biomarkers consisted of quartiles, binarization with an indeterminate zone, and staging systems. Quartiles were computed by binning continuous values at the 25th, 50th, and 75th percentiles. Binarization with an intermediate zone (BIZ) was used to model the uncertainty of assigning individuals who display biomarker values near the cutoff threshold[2]. BIZ was implemented with a GMM, where individuals were marked as uncertain if they showed less than 60% probability of being assigned to either Gaussian component.

Staging systems were included as non-binary categorical measures which assign disease severity grades based on the spatial extent of Aβ or tau pathology. For Aβ, we applied two previously published staging models[9,37]. For tau, we implemented two versions of Braak staging with different granularities: Braak staging (6) (I, III, IV, V, VI) and Braak staging (3) (I, III/IV, V/VI). Detailed description of the staging procedures for each of these systems is provided in the eMethods.

CSF-based biomarkers

We identified a subsample of individuals who had CSF immunoassays within 1 year of imaging. CSF samples were analyzed with Roche Elecsys kits sensitive to $A\beta_{42}$, $A\beta_{40}$, total-tau (tTau), and tau phosphorylated at threonine 181 (pTau181). CSF processing was administered by the ADNI Biomarker Core at the University of Pennsylvania.

CSF concentrations of analytes were grouped into biomarker categories as previously recommended[14,151] (Aβ: $A\beta_{42}$, $A\beta_{40}$, $A\beta_{42}/A\beta_{40}$, tau: pTau181, neurodegeneration: tTau). Raw concentrations were included as continuous measures. GMMs were used to define binary versions of CSF biomarkers. BIZ and quartiles were used to generate non-binary categorical versions.

Statistical analyses

We ran a series of cross-validated modeling experiments to assess how different biomarker definitions compared in their ability to model cognition. Complete details of these experiments are provided in the eMethods. Briefly, we used linear regression models to predict PHC_{Global} using single or multiple biomarker definitions. All regression models also included covariates of age, sex, and *APOE* E4 carriership. Models which only included these covariates were included as controls. Models with single biomarkers were used to test (1) if all included biomarker definitions improved cognitive prediction accuracy and (2) which individual definitions were most predictive of cognitive impairment. Next, we developed linear models combining multiple biomarker definitions as predictors of PHC_{Global}. To limit comparisons for these models, biomarkers were grouped based on the underlying pathology (AT(N)) and the variable type (binary, non-binary categorical, continuous), with nested cross-validation used to select the best predicting definition within each group. These models were used to test (3) if combination of biomarkers improved prediction accuracy, and (4) if models incorporating continuous or non-categorical binary biomarkers outperformed models with binary biomarkers. Next, we trained support vector machines (SVM) to predict PHC_{Global} from regional biomarker values to test if (5) multivariate modeling of AD pathology could improve prediction accuracy beyond that of pre-defined biomarker definitions. SVMs were trained with both linear and non-linear kernels with a grid search to select optimal hyperparameters (see eMethods).

All cross-validation experiments had 10 outer folds and were repeated 10 times to generate 100 out of sample error estimates for each tested model. Model error was assessed using root mean squared error (RMSE). Boxplots included in the Results show distributions of the 100 out-of-sample RMSE measurements from trained models. To statistically evaluate differences in accuracy for comparisons of interest, we used Nadeau-Bengio t-tests. The Nadeau-Bengio t-test includes a bias correction for the interdependency of out-of-sample error estimates when using repeated, cross-validated designs[38,39]. All tests were corrected for multiple comparisons with a false discovery rate method[40]. We investigated feature importance by plotting the distribution of

selected biomarkers across folds, the magnitude of the coefficients on biomarkers in linear regression models, and the covariance corrected weights for SVM models[41] (see eMethods). Additionally, we visualized the distribution of model-selected SUVR cutoffs for binary Aβ- and tau-PET biomarkers.

Finally, we ran a series of additional cross-validated experiments using alternative features, target variables, or clinical disease states. Specific experiments were as follows: (a) predicting the prospective slope of PHC_{Global} instead of the cross-sectional value, (b) predicting neuropsychological domains instead of PHC_{Global}, (c) using PVC tau data instead of non-PVC, (d) using CSF-based biomarkers instead of imaging-based versions, and (e) using only CU or CI individuals for model selection and out-of-sample evaluation. For (a), we only included individuals who had longitudinal cognitive measurements following baseline. To estimate longitudinal change in PHC_{Global}, linear mixed effect models were fit to model longitudinal scores following the baseline assessment. Models were fit with random slopes and intercepts for participants.

Data availability

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org. All data used in this study are accessible from ADNI following formal data usage agreements. Data were downloaded on May 10th, 2024. All R (v4.4.0) and Python (v3.10) code for this project will be shared at the following repository: https://github.com/sotiraslab/earnest_ad_biomarker_modeling.

Results

Sample characteristics

We selected 490 individuals with baseline biomarker imaging (Table 1). The cohort consisted of a mix of individuals with no cognitive impairment (CDR=0, n=288), very mild dementia (CDR=0.5, n=163), and mild to severe dementia (CDR>0.5, n=39). We observed significant differences in the distribution of age (p=0.001), sex (p=0.005), Aβ-burden (Centilioid, p<0.001), and PHC_{Global} (p<0.001) across dementia status. Mean age and Aβ-burden increased with dementia status, while PHC_{Global} decreased. Relatively more females were observed in the CU group (56.6%) than those with very mild (41.1%) or mild to severe dementia (43.6%). *APOE* E4 status was not significantly different across groups (p=0.200).

We also selected subsamples of individuals who had longitudinal PHC_{Global} assessments following baseline (n=383) and those who had CSF biomarker measurements as well as imaging (n=246). Characteristics of these samples are shown in eTables 4 and 5, respectively.

Assessment of modeling performance for biomarkers

Relative to a control model which included covariates (RMSE=0.531), almost all tested biomarkers led to a significant improvement in prediction accuracy for modeling cognitive scores (Figure 1). The only exception was hippocampal volume binarized at -2.5 Z-scores (RMSE=0.525 [0.06], p=0.09). While these results indicated that most biomarker definitions provided some predictive value, gains in performance were not equal across pathologies and variable types (range in RMSE reduction: 3.4-21.1%). Tau biomarkers led to the largest improvements in accuracy, with 9/10 of the best performing biomarkers being tau-based. Furthermore, SVM models which were trained on regional pathology were more accurate than linear models using single biomarker definitions. The tau SVM was the best performing model overall (RMSE=0.419 [0.05], p<0.001), while the Aβ SVM (RMSE=0.472 [0.06], p<0.001) and volume SVM (RMSE=0.452 [0.05], p<0.001) were the best performing Aβ and neurodegeneration

models, respectively. Outside of SVMs, the best performing models for each AT(N) category were the Aβ SUVR binarized at 1.24 (RMSE=0.492 [0.06], p<0.001), continuous MT tau SUVR (RMSE=0.440 [0.05], p<0.001), and continuous MT volume (RMSE=0.471 [0.05], p<0.001).

Combination of biomarkers

Our next experiments applied a model selection to identify the best performing biomarker predictors based on AT(N) category and variable type. We observed that all biomarker varieties caused a reduction in error over the covariate-only model (Figure 2A; mean RMSEs: Covariates=0.531, A_{BIN}=0.498, A_{CAT}=0.495, A_{CON}=0.495, T_{BIN}=0.470, T_{CAT}=0.443, T_{CON}=0.440, N_{BIN}=0.495, N_{CAT}=0.492, N_{CON}=0.471; all p<0.01). Like our experiments (Figure 1), benefits were largest for tau predictors relative to Aβ and neurodegeneration.

Combination models which included assessments for all AT(N) categories generally outperformed models with only one category included. All combination models were more accurate than the covariate-only model in predicting global cognition (mean RMSEs: $A_{BIN}/T_{BIN}/N_{BIN}$=0.446, $A_{CAT}/T_{CAT}/N_{CAT}$=0.428, $A_{CON}/T_{CON}/N_{CON}$=0.415, $A_{SVM}/T_{SVM}/N_{SVM}$=0.405, all p<0.001). Additionally, models which combined biomarkers resulted in significantly higher accuracy than models which only included one pathology assessment (Figure 2A). The benefit of combination was evident for binary ($A_{BIN}/T_{BIN}/N_{BIN}$ vs. A_{BIN}: t=3.96, p<0.001; vs. T_{BIN}: t=2.61, p<0.05; vs. N_{BIN}: t=3.92, p<0.001), non-binary categorical ($A_{CAT}/T_{CAT}/N_{CAT}$ vs. A_{CAT}: t=4.46, p<0.001; vs. T_{CAT}: t=2.22, p<0.05; vs. N_{CAT}: t=4.31 p<0.001), and continuous ($A_{CON}/T_{CON}/N_{CON}$ vs. A_{CON}: t=5.22, p<0.001; vs. T_{CON}: t=3.34, p<0.01; vs. N_{CON}: t=4.48 p<0.001) biomarkers. The combination SVM outperformed the Aβ (t=5.29, p<0.001) and gray matter (t=3.62, p<0.001) SVMs, but not the tau SVM (t=1.56, p=0.12), indicating that the improved accuracy of the multimodal SVM may be primarily driven by tau.

Direct comparison of biomarkers based on variable types indicated that biomarker binarization reduced the accuracy for tau and neurodegeneration predictors, but not Aβ (Figure 2B). Relative to the model with all binary predictors (mean RMSE for

300 $A_{BIN}/T_{BIN}/N_{BIN}$=0.446), reductions in error were seen when incorporating non-binary
301 categorical tau ($A_{BIN}/T_{CAT}/N_{BIN}$: RMSE=0.426, t=2.14, p<0.05), continuous tau
302 ($A_{BIN}/T_{CON}/N_{BIN}$: RMSE=0.424, t=2.42, p<0.05), or continuous neurodegeneration
303 ($A_{BIN}/T_{BIN}/N_{CON}$: RMSE=0.432, t=2.10, p<0.05) biomarkers. An improvement was also
304 observed when including all continuous biomarkers ($A_{CON}/T_{CON}/N_{CON}$: RMSE=0.415,
305 t=2.99, p<0.05), but not when including all non-binary biomarkers ($A_{CAT}/T_{CAT}/N_{CAT}$:
306 RMSE=0.428, t=1.56, p=0.10). The tau SVM (RMSE=0.419, t=2.02, p<0.05) and the
307 AT(N) SVM model (RMSE=0.405, t=3.42, p<0.01) also outperformed the all-binary
308 model. Models which replaced the binary Aβ definition with a non-binary categorical
309 ($A_{CAT}/T_{BIN}/N_{BIN}$: RMSE=0.446, t=0.26, p=0.60) or continuous ($A_{CON}/T_{BIN}/N_{BIN}$:
310 RMSE=0.447, t=0.24, p=0.71) version did not improve accuracy. Improvements were
311 also not seen for Aβ and neurodegeneration SVMs (p>0.05).

312 We found no differences in accuracy between models which used PVC for tau
313 SUVRs and ones with no correction (eFigure 1, all p>0.05).

314 Feature importance and model interpretation

315 For models which applied nested cross-validation to group biomarkers based on
316 AT(N) category and variable type, the best performing predictors were highly consistent
317 across folds, suggesting that some biomarker definitions generally outperformed others
318 measuring the same pathology (Figure 3A). This was particularly true for tau and
319 neurodegeneration models, where the same biomarker definitions were selected in
320 more than 95% of folds. There was slightly more variation for Aβ, but the best
321 performing biomarker was still chosen at least 67% of the time. For PET binarization,
322 previously published cutoffs accounted for 100% of selected Aβ biomarkers, but only
323 1% of tau biomarkers (the other 99% being a GMM applied to the MTT SUVR). For
324 non-binary categorical measures of PET, staging systems appeared to generally
325 outperform other approaches, appearing in 76% of selected Aβ models and 100% of tau
326 models.

327 Inspection of model coefficients highlighted the relative importance of tau for
328 predicting cognition. For linear models which included multiple biomarkers as

predictors, coefficients were highest for tau, followed by neurodegeneration and Aβ (Figure 3B). When considering continuous biomarkers in particular, weights for Aβ were much lower than those of tau or neurodegeneration. Similarly, the weights for tau features were on average higher than those of Aβ or atrophy in the multimodal SVM model (Figure 3C). Cortical weights for tau and neurodegeneration highlighted medial and lateral temporal structures, while Aβ weights were more homogenous. Subcortical regions were weighted lower than cortical ones, except for tau uptake and gray matter volume in the amygdala and hippocampus. Similar spatial patterns were observed when considering the SVM weights from separate Aβ, tau, and neurodegeneration models (eFigure 2).

We used our cross-validated modeling to identify the optimal cutoffs for Aβ and tau binarization in our cognitive modeling experiments. Our results indicated SUVR cutoffs of 1.26 (range: 1.24-1.30) for Aβ and 1.44 (range: 1.33-1.45) for tau (Figure 4).

Cognitive modeling in CU and CI populations

We also trained separate models to optimize the prediction of cognitive scores for individuals who were CU (CDR=0) and those who were CI (CDR>0). Like our results in the whole sample, we found that almost all biomarker models tested resulted in a significant improvement in accuracy relative to a baseline model with just covariates (eFigure 3). These benefits were observed for both CU (range in RMSE reduction: 3.9-15.1%) and CI (range in RMSE reduction: 7.2-30.2%) settings. The only exception was for categorical neurodegeneration models in those CU, where the difference was non-significant ($t=1.65$, $p=0.05$). The best performing models in the CU and CI populations were the all-continuous ($A_{CON}/T_{CON}/N_{CON}$: RMSE=0.349, $t=6.1$, $p<0.001$) and AT(N) SVM models ($A_{SVM}/T_{SVM}/N_{SVM}$: RMSE=0.452, $t=7.5$, $p<0.001$), respectively.

Similarly to the whole-sample results, non-binary measures of tau and neurodegeneration, but not Aβ, provided additional accuracy for modeling PHC_{Global} in CU and CI individuals (eFigure 4). Larger benefits were seen for models including non-binary tau and neurodegeneration in CI individuals relative to CU individuals. For CU individuals (mean RMSE for $A_{BIN}/T_{BIN}/N_{BIN}$: 0.381), we observed improvements when

including non-binary categorical tau ($A_{BIN}/T_{CAT}/N_{BIN}$: RMSE=0.366, t=2.8, p<0.05), continuous tau ($A_{BIN}/T_{CON}/N_{BIN}$: RMSE=0.366, t=2.6, p<0.05), or continuous biomarkers ($A_{CON}/T_{CON}/N_{CON}$: RMSE=0.359, t=2.5, p<0.05). Considering CI people (mean RMSE for $A_{BIN}/T_{BIN}/N_{BIN}$: 0.523), we observed increases in accuracy for the categorical tau ($A_{BIN}/T_{CAT}/N_{BIN}$: RMSE=0.490, t=2.6, p<0.05), continuous tau ($A_{BIN}/T_{CON}/N_{BIN}$: RMSE=0.485, t=3.3, p<0.01), continuous neurodegeneration ($A_{BIN}/T_{BIN}/N_{CON}$: RMSE=0.495, t=3.4, p<0.01), continuous AT(N) ($A_{CON}/T_{CON}/N_{CON}$: RMSE=0.479, t=3.4, p<0.01), tau SVM (T_{SVM}: RMSE=0.484, t=2.4, p<0.05), and AT(N) SVM ($A_{SVM}/T_{SVM}/N_{SVM}$: RMSE=0.442, t=4.8, p<0.001) models.

Modeling longitudinal cognition

The pattern of results we observed were largely consistent when modeling prospective change in cognition. All model varieties tested were significantly more accurate for predicting longitudinal change in PHC_{Global} (range in RMSE reduction: 4.1-30.3%, all p<0.05) relative to a covariate-only model (eFigure 5a), with the largest benefits seen for the all-continuous ($A_{CON}/T_{CON}/N_{CON}$: RMSE=0.386, t=5.83, p<0.001), multimodal SVM ($A_{SVM}/T_{SVM}/N_{SVM}$: RMSE=0.368, t=5.44, p<0.001), and tau SVM (T_{SVM}: RMSE=0.373, t=5.44, p<0.001). No biomarker linear models with non-binary measures improved prediction accuracy relative to an all-binary baseline (eFigure 5b, all p>0.05). However, the tau SVM (T_{SVM}: RMSE=0.373, t=3.05, p<0.01) and multimodal SVM ($A_{SVM}/T_{SVM}/N_{SVM}$: RMSE=0.368, t=3.54, p<0.01) were still significantly more accurate at predicting change in PHC_{Global} than a model consisting of all-binary predictors.

Modeling of individual neuropsychological domains

We also observed similar patterns of accuracy differences for non-binary biomarker definitions when modeling neuropsychological domains instead of PHC_{Global}. Significant benefits were only observed for models which included non-binary or SVM-based assessments of tau and neurodegenerative pathology (eFigure 6). Models which included non-binary definitions of Aβ alone did not surpass the all-binary model for any neuropsychological domain. Continuous tau and neurodegeneration measures improved accuracy for prediction of PHC_{EF} (p<0.05) and PHC_{Visual} (p<0.05), while non-

binary categorical measures only improved prediction of PHC_{Visual} ($p<0.05$). The AT(N) SVM had significantly higher accuracy for modeling PHC_{EF} ($p<0.01$), PHC_{Visual} ($p<0.05$), and PHC_{Memory} ($p<0.05$). No significant differences were observed for $PHC_{Language}$ prediction accuracy (all $p>0.05$).

Comparison of image-based and CSF-based models

We observed that CSF-based models performed relatively poorly for modeling cognition. Models which incorporated CSF-based biomarkers, as opposed to imaging-based ones, did not perform better than a baseline model consisting of only covariates (Figure 5 & eFigure 7, all $p>0.05$). Moreover, imaging-based models were significantly more accurate than CSF-based models. This was true for binary ($t=2.81$, $p<0.01$), non-binary categorical ($t=3.68$, $p<0.01$), and continuous ($t=3.96$, $p<0.01$) biomarker definitions.

Discussion

AD biomarkers differ from each other along various axes such as the underlying pathology they measure (e.g. Aβ, tau), the modality (e.g., imaging, CSF, blood), and measurement characteristics (e.g., variable type). Our analyses indicate that differences along these dimensions result in considerable variability when imaging biomarkers are utilized in downstream tasks. We show that even biomarkers which assess the same pathology exhibit a range in accuracy when applied for modeling cognition. Additionally, we demonstrate that multivariate machine learning approaches can surpass traditional biomarker definitions for cognitive prediction in AD. Careful consideration should be applied when selecting biomarker definitions for predictive tasks, as certain operationalizations may be relatively less informative than other variants. While we specifically focus on cognitive prediction, our results may be relevant for other settings where researchers wish to quantify AD pathology.

While nearly all tested biomarkers provided predictive gains when modeling cognition, some categories of biomarkers yielded consistently larger improvements than others. Multiple analyses demonstrated that tau biomarkers exhibited stronger

associations with cognition than assessments of Aβ or neurodegeneration, a well-documented finding.[42–44] Feature importance analyses indicated that tau predictors were weighted higher than Aβ or neurodegeneration measures in models which incorporated all three AT(N) categories. However, combined AT(N) models generally outperformed unimodal ones, even when tau was the single biomarker included. Thus, while measures of tau are important indicators of cognitive decline, incorporation of measures spanning other pathologies is warranted for enhancing predictive accuracy.

We observed that SVM models outperformed more traditional linear models of cognition. Tau and multimodal SVMs were the best predictors overall in many experiments. These models were also the only models which outperformed binary biomarker models for prediction of longitudinal cognitive decline. Aβ and neurodegeneration SVMs were also relatively stronger than other individual biomarker definitions of these pathologies. The superior performance of SVM models suggest that there may be key predictive signal occurring in brain regions external to the manually defined meta-ROIs which are utilized in most of the biomarker definitions we tested. However, the SVMs also allowed for non-linear transformations of input features, making them relatively more powerful models.

Inclusion of non-binary tau and neurodegeneration predictors led to small but consistent improvements in accuracy relative to binary alternatives. However, binary Aβ measures performed equally to non-binary ones. As such, our findings indicate that binarization along dimensions of tau and neurodegeneration (e.g., labeling individuals as T+/- or N+/-) may obfuscate information relevant to the prediction of cognitive decline. On the other hand, dichotomization of Aβ status may be sufficient. These notion agrees with revised criteria for diagnosis and staging of AD[2]: their proposed PET staging system includes binary assessment of Aβ (i.e., has AD or not) and multi-level staging of tau based on the extent of progression outside the medial temporal lobe. Interestingly, the specific benefits for tau and neurodegeneration (and not Aβ) were consistent when considering only CU or CI individuals and when modeling some neuropsychological domains (executive functioning and visuospatial performance). These results agree with a previous study which found similar non-dichotomized tau

and neurodegeneration for modeling longitudinal cognitive decline[13]. However, we did not replicate their findings showing accuracy improvements when modeling prospective cognition in CU individuals and including non-binary measures of Aβ.

While nearly all the imaging biomarkers we tested improved prediction of cognitive impairment, the same was not true for CSF counterparts. Models which incorporated CSF biomarkers as predictors did not perform better than baseline models which only included standard covariates, regardless of the analyte or its operationalization. Previous findings have similarly demonstrated stronger associations for imaging biomarkers and cognitive scores in AD, relative to fluid biomarkers[45]. Importantly, the CSF analytes tested largely reflect earlier pathological cascades which likely develop and saturate prior to the onset of neurodegeneration and cognitive decline[2,46]. As such, they may be less suited for providing direct associations with cognitive decline, and more suited for diagnosis or prediction of future decline.

Our study has limitations which should be considered. First, while we performed a comprehensive and rigorous evaluation of multiple biomarkers, biomarkers such as fluorodeoxyglucose-PET, cortical thickness, and functional imaging were not included in this study. Future studies are warranted to examine them. Second, this study relied only on ADNI because inclusion of other sources posed issues of harmonization and biomarker availability. While ADNI is one of few databases which can enable the analyses we conducted, it is also relatively limited in its inclusion of demographic diversity[47]. As such, our results warrant replication in other datasets.

The growth of large data initiatives has led to an explosion of approaches for biomarker assessment of AD. While picking from the myriads of methods for quantification of AD pathology, it is important for researchers to mind the biological, statistical, and practical characteristics of each approach. Our results demonstrate that different operationalizations of the same pathology can result in variable performance for downstream predictive tasks. More complex indices of pathology may be superior to dichotomous alternatives, particularly for measures of neurodegeneration and tau. Finally, data-driven, machine learning approaches may be preferable for identifying biomarker contributions to cognitive decline.

Acknowledgement

The authors thank the staff for the Washington University Center for High Performance Computing who helped enable this work. Computations were performed using the facilities of the Washington University Research Computing and Informatics Facility (RCIF). The RCIF has received funding from NIH S10 program grants: 1S10OD025200-01A1 and 1S10OD030477-01.

Study Funding

This work was supported by the National Institutes of Health (NIH) (R01-AG067103) and the BrightFocus Foundation (ADR A2021042S).

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.

Disclosures

Author AS has equity in TheraPanacea and have received personal compensation for serving as grant reviewer for BrightFocus Foundation. The remaining authors have no conflicting interests to report.

References

1. Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*. 2018;14(4):535-562. doi:10.1016/j.jalz.2018.02.018

2. Jack CR, Andrews JS, Beach TG, et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & Dementia*. Published online June 27, 2024:alz.13859. doi:10.1002/alz.13859

3. Hansson O. Biomarkers for neurodegenerative diseases. *Nat Med*. 2021;27(6):954-963. doi:10.1038/s41591-021-01382-x

4. Rathore S, Habes M, Iftikhar MA, Shacklett A, Davatzikos C. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *NeuroImage*. 2017;155:530-548. doi:10.1016/j.neuroimage.2017.03.057

5. McConathy J, Sheline YI. Imaging Biomarkers Associated With Cognitive Decline: A Review. *Biological Psychiatry*. 2015;77(8):685-692. doi:10.1016/j.biopsych.2014.08.024

6. Chen P, Zhang S, Zhao K, Kang X, Rittman T, Liu Y. Robustly uncovering the heterogeneity of neurodegenerative disease by using data-driven subtyping in neuroimaging: A review. *Brain Research*. 2024;1823:148675. doi:10.1016/j.brainres.2023.148675

7. Abdelnour C, Agosta F, Bozzali M, et al. Perspectives and challenges in patient stratification in Alzheimer's disease. *Alzheimers Res Ther*. 2022;14:112. doi:10.1186/s13195-022-01055-y

8. Earnest T, Bani A, Ha SM, et al. Data-driven decomposition and staging of flortaucipir uptake in Alzheimer's disease. *Alzheimer's & Dementia*. 2024;20(6). doi:10.1002/alz.13769

9. Collij LE, Heeman F, Salvadó G, et al. Multitracer model for staging cortical amyloid deposition using PET imaging. *Neurology*. 2020;95(11):e1538-e1553. doi:10.1212/WNL.0000000000010256

10. Klunk W, Cohen A, Bi W, et al. Why we need two cutoffs for amyloid imaging: Early versus Alzheimer's-like amyloid-positivity. *Alzheimer's & Dementia*. 2012;8(4, Supplement):P453-P454. doi:10.1016/j.jalz.2012.05.1208

11. Jack CR, Wiste HJ, Weigand SD, et al. Defining imaging biomarker cut-points for brain aging and Alzheimer's disease. *Alzheimers Dement*. 2017;13(3):205-216. doi:10.1016/j.jalz.2016.08.005

12. Weigand AJ, Maass A, Eglit GL, Bondi MW. What's the cut-point?: a systematic investigation of tau PET thresholding methods. *Alzheimers Res Ther*. 2022;14(1):49. doi:10.1186/s13195-022-00986-w

13. Mattsson-Carlgren N, Leuzy A, Janelidze S, et al. The implications of different approaches to define AT(N) in Alzheimer disease. *Neurology*. 2020;94(21):e2233-e2244. doi:10.1212/WNL.0000000000009485

14. Salimi Y, Domingo-Fernández D, Hofmann-Apitius M, et al. Data-Driven Thresholding Statistically Biases ATN Profiling across Cohort Datasets. *J Prev Alzheimers Dis*. 2024;11(1):185-195. doi:10.14283/jpad.2023.100

15. Provost K, Iaccarino L, Soleimani-Meigooni DN, et al. Comparing ATN-T designation by tau PET visual reads, tau PET quantification, and CSF PTau181 across three cohorts. *Eur J Nucl Med Mol Imaging*. 2021;48(7):2259-2271. doi:10.1007/s00259-020-05152-8

16. Bucci M, Chiotis K, Nordberg A. Alzheimer's disease profiled by fluid and imaging markers: tau PET best predicts cognitive decline. *Mol Psychiatry*. 2021;26(10):5888-5898. doi:10.1038/s41380-021-01263-2

17. Morris JC. Clinical Dementia Rating: A Reliable and Valid Diagnostic and Staging Measure for Dementia of the Alzheimer Type. *International Psychogeriatrics*. 1997;9(S1):173-176. doi:10.1017/S1041610297004870

18. Mukherjee S, Choi SE, Lee ML, et al. Cognitive Domain Harmonization and Cocalibration in Studies of Older Adults. *Neuropsychology*. 2023;37(4):409-423. doi:10.1037/neu0000835

19. ADNI Study Documents. Alzheimer's Disease Neuroimaging Initiative. 2024. Accessed May 15, 2024. https://adni.loni.usc.edu/methods/documents/

20. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002;33(3):341-355. doi:10.1016/s0896-6273(02)00569-x

21. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*. 2006;31(3):968-980. doi:10.1016/j.neuroimage.2006.01.021

22. Diedrichsen J. A spatially unbiased atlas template of the human cerebellum. *Neuroimage*. 2006;33(1):127-138. doi:10.1016/j.neuroimage.2006.05.056

23. Baker SL, Maass A, Jagust WJ. Considerations and code for partial volume correcting [18F]-AV-1451 tau PET data. *Data in Brief*. 2017;15:648-657. doi:10.1016/j.dib.2017.10.024

24. Rousset OG, Ma Y, Evans AC. Correction for partial volume effects in PET: principle and validation. *J Nucl Med*. 1998;39(5):904-911.

25. Mormino EC, Kluth JT, Madison CM, et al. Episodic memory loss is related to hippocampal-mediated beta-amyloid deposition in elderly subjects. *Brain*. 2009;132(Pt 5):1310-1323. doi:10.1093/brain/awn320

26. Jagust WJ, Landau SM, Shaw LM, et al. Relationships between biomarkers in aging and dementia. *Neurology*. 2009;73(15):1193-1199. doi:10.1212/WNL.0b013e3181bc010c

27. Klunk WE, Koeppe RA, Price JC, et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimers Dement*. 2015;11(1):1-15.e1-4. doi:10.1016/j.jalz.2014.07.003

28. Royse SK, Minhas DS, Lopresti BJ, et al. Validation of amyloid PET positivity thresholds in centiloids: a multisite PET study approach. *Alzheimers Res Ther*. 2021;13(1):99. doi:10.1186/s13195-021-00836-1

29. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol*. 1991;82(4):239-259. doi:10.1007/BF00308809

30. Braak H, Alafuzoff I, Arzberger T, Kretzschmar H, Del Tredici K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol*. 2006;112(4):389-404. doi:10.1007/s00401-006-0127-z

31. Lemoine L, Leuzy A, Chiotis K, Rodriguez-Vieitez E, Nordberg A. Tau positron emission tomography imaging in tauopathies: The added hurdle of off-target binding. *Alzheimers Dement (Amst)*. 2018;10:232-236. doi:10.1016/j.dadm.2018.01.007

32. Biel D, Brendel M, Rubinski A, et al. Tau-PET and in vivo Braak-staging as prognostic markers of future cognitive decline in cognitively normal to demented individuals. *Alzheimer's Research & Therapy*. 2021;13(1):137. doi:10.1186/s13195-021-00880-x

606 33. Landau SM, Mintun MA, Joshi AD, et al. Amyloid deposition, hypometabolism, and
607 longitudinal cognitive decline. *Annals of Neurology*. 2012;72(4):578-586.
608 doi:10.1002/ana.23650

609 34. Su Y, Flores S, Wang G, et al. Comparison of Pittsburgh compound B and
610 florbetapir in cross-sectional and longitudinal studies. *Alzheimers Dement (Amst)*.
611 2019;11:180-190. doi:10.1016/j.dadm.2018.12.008

612 35. Salvadó G, Molinuevo JL, Brugulat-Serrat A, et al. Centiloid cut-off values for
613 optimal agreement between PET and CSF core AD biomarkers. *Alzheimer's*
614 *Research & Therapy*. 2019;11(1):27. doi:10.1186/s13195-019-0478-z

615 36. Farrell ME, Jiang S, Schultz AP, et al. Defining the Lowest Threshold for Amyloid-
616 PET to Predict Future Cognitive Decline and Amyloid Accumulation. *Neurology*.
617 2021;96(4):e619-e631. doi:10.1212/WNL.0000000000011214

618 37. Mattsson N, Palmqvist S, Stomrud E, Vogel J, Hansson O. Staging β-Amyloid
619 Pathology With Amyloid Positron Emission Tomography. *JAMA Neurology*.
620 2019;76(11):1319-1329. doi:10.1001/jamaneurol.2019.2214

621 38. Nadeau C, Bengio Y. Inference for the Generalization Error. *Machine Learning*.
622 2003;52(3):239-281. doi:10.1023/A:1024068626366

623 39. Bouckaert RR, Frank E. Evaluating the Replicability of Significance Tests for
624 Comparing Learning Algorithms. In: Dai H, Srikant R, Zhang C, eds. *Advances in*
625 *Knowledge Discovery and Data Mining*. Lecture Notes in Computer Science.
626 Springer; 2004:3-12. doi:10.1007/978-3-540-24775-3_3

627 40. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and
628 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society:*
629 *Series B (Methodological)*. 1995;57(1):289-300. doi:10.1111/j.2517-
630 6161.1995.tb02031.x

631 41. Haufe S, Meinecke F, Görgen K, et al. On the interpretation of weight vectors of
632 linear models in multivariate neuroimaging. *NeuroImage*. 2014;87:96-110.
633 doi:10.1016/j.neuroimage.2013.10.067

634 42. Gordon BA, McCullough A, Mishra S, et al. Cross-sectional and longitudinal atrophy
635 is preferentially associated with tau rather than amyloid β positron emission
636 tomography pathology. *Alzheimers Dement (Amst)*. 2018;10:245-252.
637 doi:10.1016/j.dadm.2018.02.003

638 43. La Joie R, Visani AV, Baker SL, et al. Prospective longitudinal atrophy in
639 Alzheimer's disease correlates with the intensity and topography of baseline tau-
640 PET. *Sci Transl Med*. 2020;12(524):eaau5732. doi:10.1126/scitranslmed.aau5732

641 44. Ossenkoppele R, Smith R, Mattsson-Carlgren N, et al. Accuracy of Tau Positron
642 Emission Tomography as a Prognostic Marker in Preclinical and Prodromal

643 Alzheimer Disease: A Head-to-Head Comparison Against Amyloid Positron
644 Emission Tomography and Magnetic Resonance Imaging. *JAMA Neurology*.
645 2021;78(8):961-971. doi:10.1001/jamaneurol.2021.1858

646 45. Lu J, Ma X, Zhang H, et al. Head-to-head comparison of plasma and PET imaging
647 ATN markers in subjects with cognitive complaints. *Transl Neurodegener*.
648 2023;12(1):34. doi:10.1186/s40035-023-00365-x

649 46. Tissot C, Therriault J, Kunach P, et al. Comparing tau status determined via plasma
650 pTau181, pTau231 and [18F]MK6240 tau-PET. *eBioMedicine*. 2022;76:103837.
651 doi:10.1016/j.ebiom.2022.103837

652 47. Weiner MW, Veitch DP, Miller MJ, et al. Increasing participant diversity in AD
653 research: Plans for digital screening, blood testing, and a community-engaged
654 approach in the Alzheimer's Disease Neuroimaging Initiative 4. *Alzheimer's &*
655 *Dementia*. 2023;19(1):307-317. doi:10.1002/alz.12797

656

Tables

	CDR=0.0	CDR=0.5	CDR=1.0+	p-value
n	288	163	39	
Age	73.84 (7.44)	75.78 (8.46)	78.29 (8.58)	**0.001**
Sex (M/F)	125/163	96/67	22/17	**0.005**
***APOE* E4+**	99 (34.4%)	62 (38.0%)	19 (48.7%)	0.200
Centiloid	19.90 (36.10)	45.57 (54.95)	78.41 (49.61)	**<0.001**
PHC_{Global}	0.81 (0.36)	0.32 (0.46)	-0.42 (0.54)	**<0.001**

Table 1: Sample characteristics. The last column shows p-values for significance tests comparing distributions of variables across dementia status groups (CDR). Chi-squared tests of association were used for categorical variables (sex, *APOE* status) and one-way ANOVAs were used for continuous variables (age, Centiloid, PHC_{Global}).

Figures

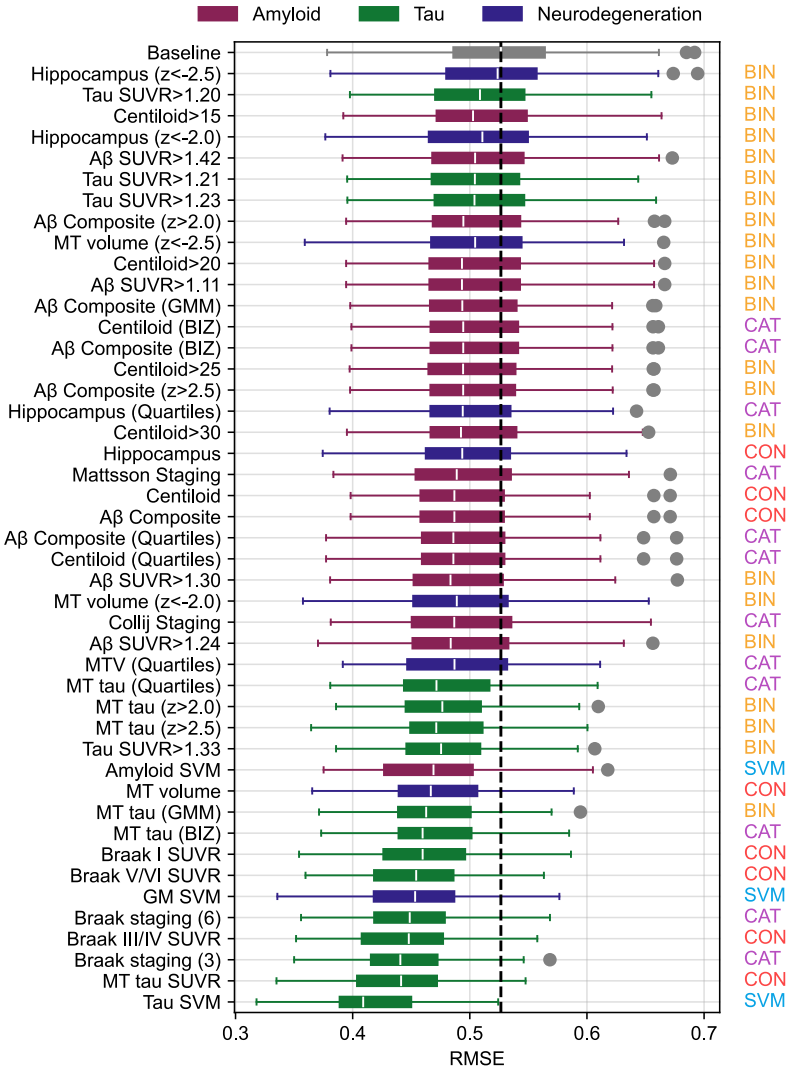

Figure 1: Boxplots showing performance of individual biomarkers for predicting PHC_{Global}. Values plotted are RMSEs taken from out of sample predictions for 100 cross-validation instances (lower value is more accurate). The baseline model only included covariates as independent variables (mean performance for this model indicated by the dotted line). All other models included the same covariates and a single amyloid (maroon), tau (green), or neurodegeneration (blue) biomarker. Labels on the right indicate the variable type of each biomarker (BIN=binary, CAT=non-binary categorical, CON=continuous, SVM=support vector machine). All models exhibited significantly lower RMSE than the baseline model, except for hippocampal volume binarized at -2.5 Z-scores [Hippocampus (z<-2.5)].

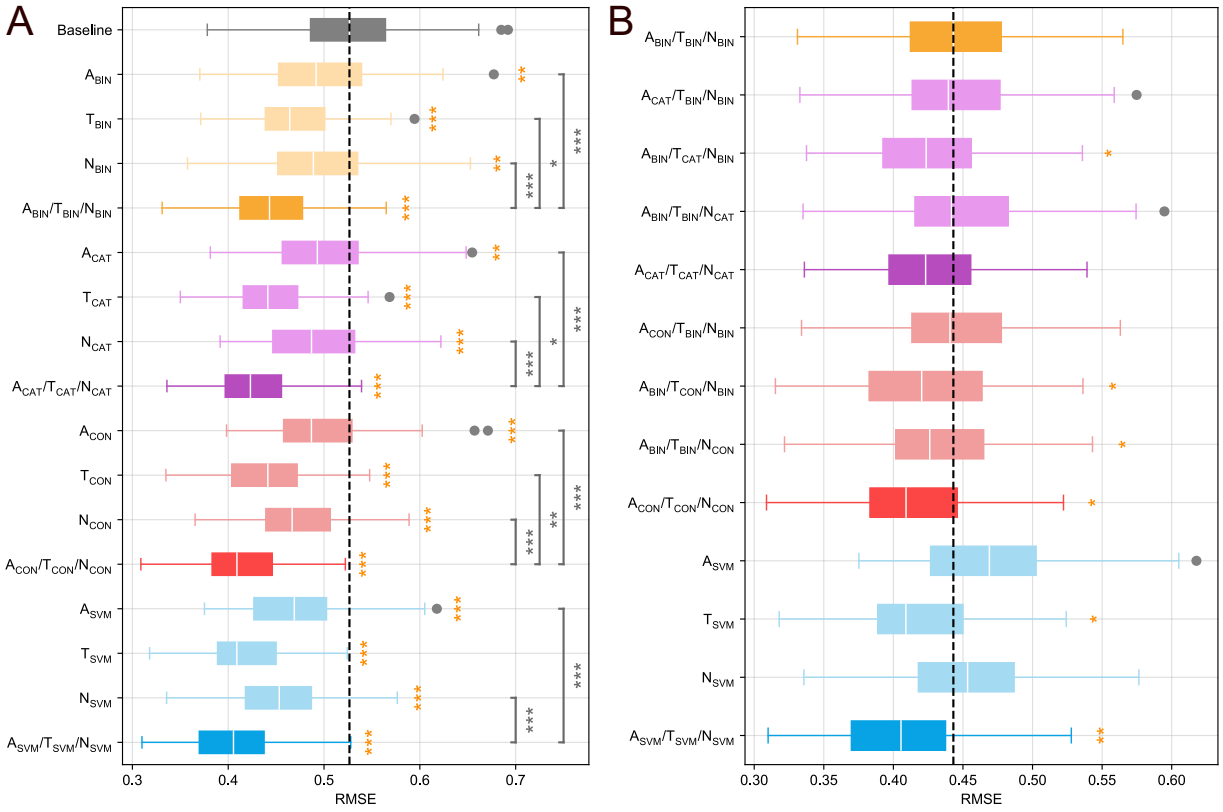

Figure 2: Boxplots showing RMSE performance of combination biomarkers for predicting global neuropsychological performance (PHC_{Global}). **A.** Individual and combination biomarker models are compared against a baseline model using only covariates (mean performance indicated by dotted line) to predict PHC_{Global}. **B.** Combination biomarker models with non-binary variable types are compared against a baseline model with binary biomarker definitions (mean performance indicated by dotted line). In both panels, colors are used to indicate the variable type of included biomarkers (yellow: binary, purple: non-binary categorical, red: continuous, blue: SVM). Lighter coloring indicates models which only have a single pathology assessment, while darker coloring indicates models which have Aβ, tau, and neurodegeneration biomarkers. Gold stars indicate a significant improvement in accuracy relative to the topmost model. Gray stars and bars highlight significant pairwise differences between individual models. Statistical results are derived from Nadeau-Bengio t-tests with correction for multiple comparisons (*p<0.05, **p<0.01, ***p<0.001).

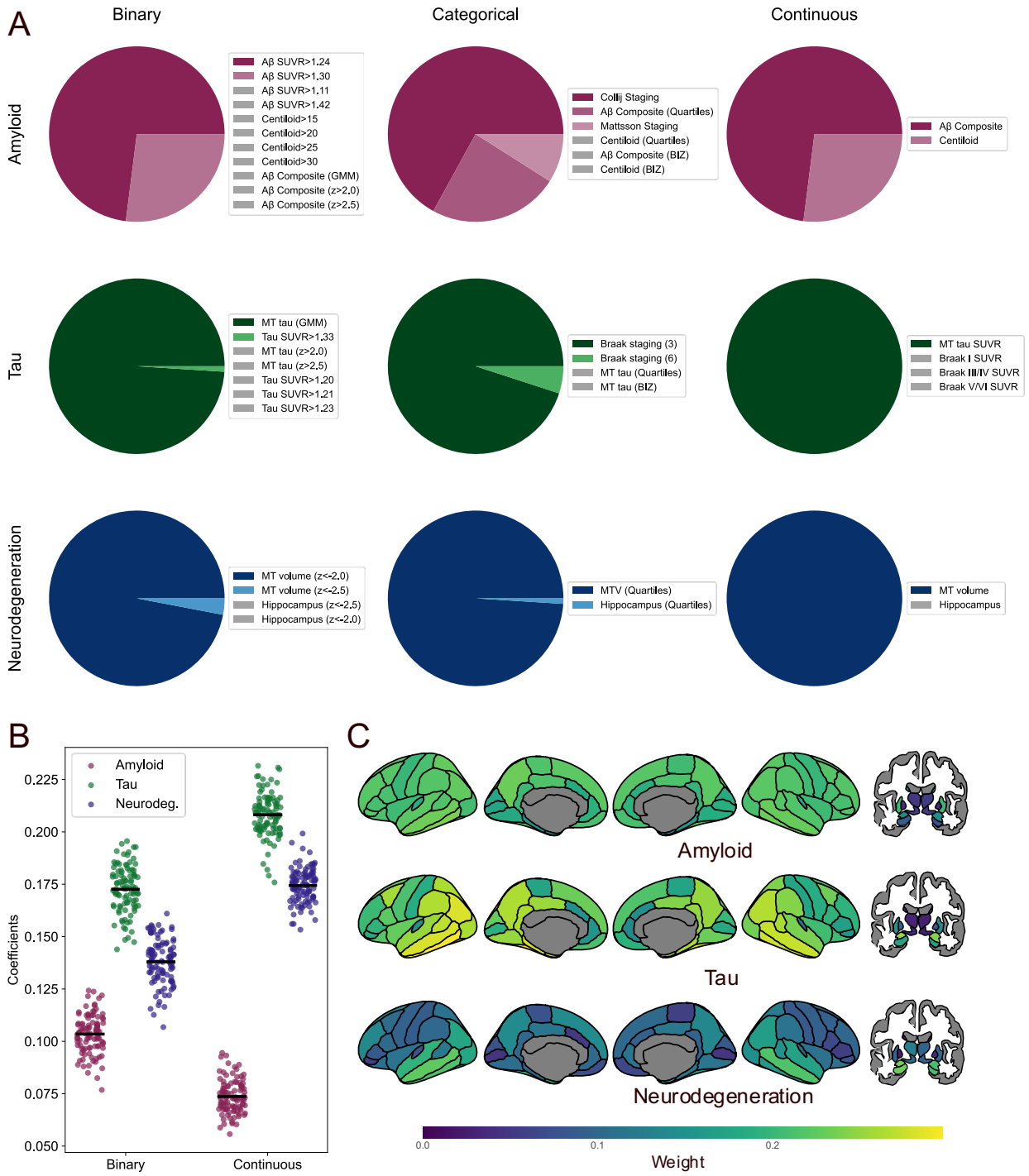

Figure 3: Feature importance analysis. **A.** Pie charts showing which biomarkers were selected as the best performing from cross-validation (100 training fold instances). Biomarkers shown with gray coloring were not selected in any iteration. **B.** Coefficients for the Aβ, tau, and neurodegeneration predictor in cross-validated linear models. Values are taken from 100 instances of the all binary ($A_{BIN}/T_{BIN}/N_{BIN}$) and all continuous ($A_{CON}/T_{CON}/N_{CON}$) models. **C.** Brain maps showing average regional feature importance derived from the $A_{SVM}/T_{SVM}/N_{SVM}$ model.

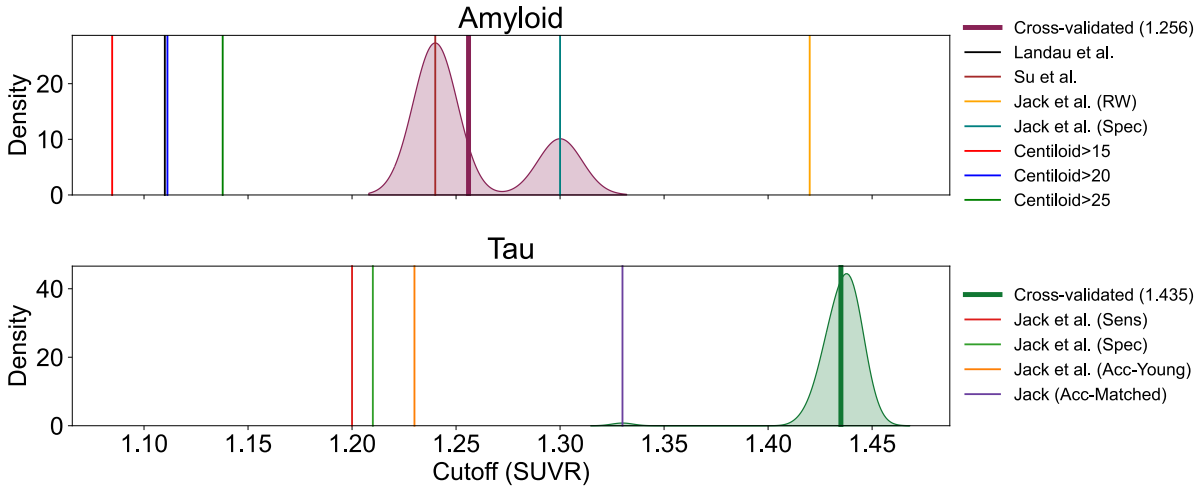

Figure 4. Estimated cutoffs for Aβ- and tau-PET binarization. **A.** The kernel density estimation of selected cutoffs for Aβ (100 cross-validation iterations) is shown in maroon, with the mean value (1.256) highlighted with the bold vertical line. **B.** Same as A., but for tau and shown in green (mean value: 1.435). In both panels, other vertical lines show other pre-defined cutoff values that were tested.

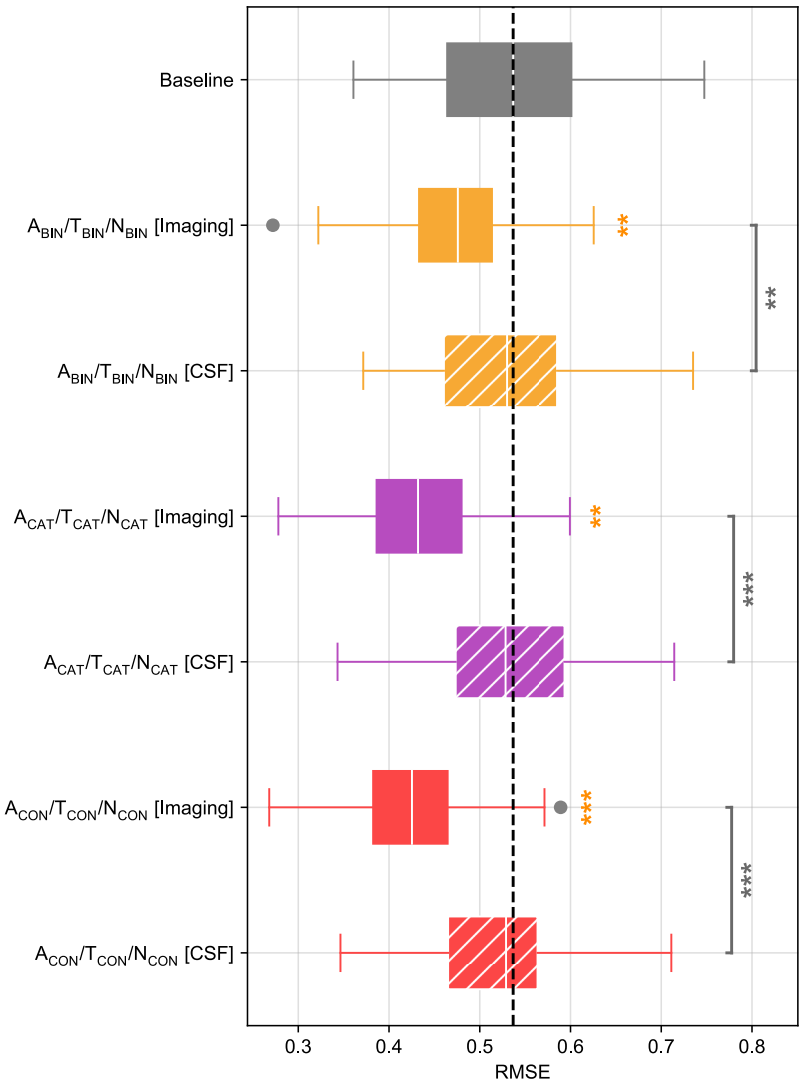

Figure 5. Boxplots comparing accuracy cognition predictions using image-based (solid) and CSF-based (hatched) biomarkers. Values are RMSEs taken from 100 cross-validation instances. Colors are used to indicated the variable type of included biomarkers (yellow: binary, purple: non-binary categorical, red: continuous). Gold stars indicate a significant improvement in accuracy relative to the baseline model (only covariates). Gray stars and bars highlight significant pairwise differences between individual models. Statistical results are derived from Nadeau-Bengio t-tests with correction for multiple comparisons (*p<0.05, **p<0.01, ***p<0.001).