## 1 Comprehensive evaluation of AT(N) imaging biomarkers for

## 2 predicting cognition

- 3 Tom Earnest<sup>1</sup>, Braden Yang<sup>1</sup>, Deydeep Kothapalli<sup>1</sup>, Aristeidis Sotiras<sup>1,2</sup>, for the
- 4 Alzheimer's Disease Neuroimaging Initiative\*
- <sup>1</sup> Mallinckrodt Institute of Radiology, Washington University School of Medicine in St Louis; 4525 Scott
- 6 Ave, Saint Louis, MO 63110
- 7 <sup>2</sup> Institute for Informatics, Data Science & Biostatistics, Washington University School of Medicine in St
- 8 Louis; 660 S. Euclid Ave, Campus Box 8132, Saint Louis, MO 63110
- 9 \* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging
- 10 Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to
- 11 the design and implementation of ADNI and/or provided data but did not participate in analysis or writing
- 12 of this report. A complete listing of ADNI investigators can be found at: <u>http://adni.loni.usc.edu/wp-</u>
- 13 content/uploads/how to apply/ADNI Acknowledgement List.pdf

It is made available under a CC-BY-NC 4.0 International license .

# 14 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.

21 **Methods:** We included individuals from ADNI who underwent amyloid-PET ([<sup>18</sup>F]-

22 Florbetapir), tau-PET ([<sup>18</sup>F]-Flortaucipir), and volumetric MRI imaging. We compiled a

23 large collection of imaging biomarker definitions (42 in total) spanning different

24 pathologies (amyloid, tau, neurodegeneration) and variable types (continuous, binary,

25 non-binary categorical). Using cross-validation, we trained regression models to predict

26 neuropsychological performance, both globally and across different subdomains

27 (Phenotype Harmonization Consortium composites), using different combinations of

28 biomarkers. We also compared these biomarker models to support vector machines

29 (SVMs) trained to predict cognition directly from imaging regions of interest. In a

30 subsample of individuals with CSF biomarker readouts, we repeated experiments

31 comparing the accuracy of models using imaging and fluid biomarkers. Additional

32 analyses tested the predictive strength of imaging biomarkers when limited to specific

33 clinical stages of disease (cognitive unimpaired vs. impaired) and when modeling

34 longitudinal cognitive change.

35 **Results:** Our sample included 490 people (247 female) with a mix of no impairment 36 (n=288), mild impairment (n=163), and dementia (n=39). While almost all biomarkers 37 tested were predictive of cognitive performance, we observed substantial variability in 38 accuracy, even for measures of the same pathology. Tau biomarkers were the single 39 most accurate single predictors, though combination of biomarkers spanning multiple 40 pathologies were more accurate overall. SVM models were generally more accurate 41 than models using traditional biomarkers. Incorporating continuous or non-binary 42 categorical biomarkers was beneficial only for tau and neurodegeneration, but not 43 amyloid. Patterns of results were largely consistent when considering different clinical

- 44 stages of disease, neuropsychological domains, and longitudinal cognition. In the CSF
- 45 subsample (n=246), imaging biomarkers strongly outperformed CSF versions for
- 46 cognitive prediction.
- 47 **Discussion:** We demonstrated that different imaging biomarker definitions can lead to
- 48 variability in downstream predictive tasks. Researchers should consider how their
- 49 biomarker operationalizations may help or hinder the assessment of disease severity.

It is made available under a CC-BY-NC 4.0 International license .

# 50 Introduction

51 The modern biological definition of Alzheimer's Disease (AD) relies on biomarkers.<sup>1–3</sup> Biomarkers can accurately quantify pathobiological disease processes 52 53 which are specific to AD, particularly the aggregation of amyloid-beta (A $\beta$ ) plaques and 54 the neocortical spread of tau neurofibrillary tangles. Importantly, biomarkers can detect 55 and measure these pathologies prior to symptomatic onset. Because of their 56 capabilities, biomarkers have been used in a variety of research settings including 57 disease classification<sup>4</sup>, cognitive forecasting<sup>5</sup>, subtype identification<sup>6</sup>, clinical trial 58 stratification<sup>7</sup>, disease staging<sup>8,9</sup>, and more. Moreover, biomarkers are becoming 59 increasingly important for clinical management of AD<sup>2</sup>. For instance, recently approved 60 anti-A $\beta$  treatments for AD require the presence of A $\beta$ -pathology as assessed by 61 biomarkers.

62 Interest in biological AD assessment has led to the creation of many AD-sensitive 63 biomarkers which vary in terms of modality, underlying pathology, and statistical 64 formulation. Idiosyncrasies of biomarker definitions may result in unwanted variability 65 when applied for clinical and research uses. For example, estimated cut points for PET 66 and CSF biomarker dichotomization are fairly application specific <sup>10–12</sup>, and different 67 approaches to pathological thresholding result in considerable variability for group 68 assignment<sup>13–16</sup>. Less is known, however, about how variability in biomarker definitions affects prediction of cognition in AD. Identifying which specific biomarkers are most 69 70 predictive of cognitive trajectories, particularly at different stages of disease, can provide 71 insight into biological mechanisms of AD. Moreover, precise cognitive decline 72 predictions are valuable for identifying candidates for early therapeutic interventions and 73 for establishing meaningful cognitive endpoints in clinical trials. Despite these 74 implications, investigations into the ramifications of different biomarker 75 operationalizations remain limited. One previous study found that different biomarker 76 definitions varied in their ability to predict longitudinal Mini-Mental State Examination 77 (MMSE) scores, and that dichotomization hindered predictive power of some 78 biomarkers relative to continuous values.<sup>13</sup> Similar analyses in separate cohorts with 79 additional cognitive measures are needed to confirm and extend these findings,

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

82 Here, we developed a comprehensive set of neuroimaging measures (42 in total) 83 covering the AD core biomarkers A $\beta$  (A) and tau (T), as well as non-specific biomarkers of neurodegeneration ((N)). We systematically evaluated how different categories of 84 85 biomarkers and individual variants differ in their ability to predict different cognitive 86 outcomes. While we focused on cross-sectional cognition, we also extended analyses 87 to measures of prospective longitudinal cognition and neuropsychological domains. We 88 additionally incorporated machine learning to test how traditional biomarker approaches 89 compare to methods which can detect more complex, multivariate patterns in imaging 90 data. Finally, we tested multiple CSF biomarkers (20 definitions spanning 4 analytes) 91 and compared their performance with imaging alternatives.

# 92 Methods

## 93 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<sup>17</sup>, Phenotype Harmonization Consortium (PHC)
cognitive composite scores<sup>18</sup>.

## 100 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.

## 103 Image acquisition and processing

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

It is made available under a CC-BY-NC 4.0 International license .

106 accelerated MPRAGE sequence. A $\beta$ -PET scans were acquired 50-70 minutes (4 107 frames × 5 minutes) after a 370 MBq (± 10%) injection of [<sup>18</sup>F]-Florbetapir. Tau-PET 108 scans were acquired 75-105 minutes (6 frames × 5 minutes) after a 370 MBq (± 10%) 109 injection of [<sup>18</sup>F]-Flortaucipir.

110 We accessed processed MRI and PET derivatives generated by the ADNI PET 111 Core. A Freesurfer (v7.1.1) processing pipeline was applied to MRI scans to generate aray matter volumes within regions of interest (ROIs) of standard subcortical<sup>20</sup> and 112 113 cortical atlases<sup>21</sup>. PET standardized uptake value ratios (SUVRs) were generated for 114 these same ROIs after coregistration of each PET image to a contemporaneous MRI 115 scan. Our analyses incorporated unilateral values from 68 cortical and 14 subcortical 116 gray matter regions. Volumes were standardized relative to the intracranial volume. 117 Aβ-PET uptakes were standardized to a whole cerebellum ROI, while tau-PET uptakes 118 were standardized relative to an ROI containing inferior cerebellar gray matter<sup>22</sup>.

119 Partial volume corrected (PVC) PET uptakes were available for tau (Geometric 120 Transfer Matrix approach<sup>23,24</sup>) but not for A $\beta$ . We used uncorrected SUVR values for 121 most experiments, but we repeated some experiments with PVC-corrected tau SUVRs 122 to evaluate the effect of PVC on cognitive prediction accuracy.

## 123 Cognitive and clinical assessments

124 Cognition was assessed using composite scores developed by the PHC<sup>18</sup>. We 125 averaged the memory (PHC<sub>Memory</sub>), executive functioning (PHC<sub>EF</sub>), visuospatial 126 (PHC<sub>Visual</sub>), and language (PHC<sub>Language</sub>) composites to create one global cognitive 127 composite (PHC<sub>Global</sub>). Composites are unitless factor loadings, with lower scores 128 corresponding to more impairment.

129 CDR was used as a measure of dementia severity<sup>17</sup>. Subjects were assigned to 130 the following groups based on CDR status: cognitively unimpaired (CU, CDR=0) or 131 cognitively impaired (CI, CDR>=0.5).

It is made available under a CC-BY-NC 4.0 International license .

## 132 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.

138 Continuous variables consisted of scalar MRI (volume) or PET (SUVR) measures 139 in standard composite ROIs. For A $\beta$ , continuous measures of A $\beta$  included the average SUVR in a cortical summary region<sup>25,26</sup> (A<sub>β</sub> composite) and Centiloid<sup>27</sup>. Centiloids were 140 141 provided by ADNI and derived from the AB composite using previously validated 142 equations<sup>28</sup>. Continuous tau measures included the average uptakes in a meta-143 temporal (MT) composite region<sup>11</sup> and uptakes in ROIs corresponding to progressive 144 Braak stages<sup>29,30</sup> (Braak I, Braak III/IV, Braak V/VI). Braak II was omitted due to offtarget binding issues with flortaucipir<sup>31,32</sup>. Hippocampal volume and volume of the MT 145 146 region were included as continuous assessments of neurodegeneration.

147 Binary predictors consisted of dichotomized versions of the continuous predictors 148 listed above. There were three main methods tested for binarizing continuous 149 variables: previously published cutoffs, Z-scoring, and Gaussian mixture modeling 150 (GMM). Previously published cutoffs were included for the A $\beta$  composite at the following SUVRs: 1.11<sup>33</sup>, 1.24<sup>34</sup>, 1.30<sup>11</sup>, 1.42<sup>11</sup>. We also tested Centiloid cutoffs (15, 20, 151 25, 30) based on ranges reported in previous literature<sup>28,35,36</sup>. Z-scoring and GMMs 152 153 were included as data-driven approaches for deriving cutoffs. These methods were 154 applied to the A $\beta$  composite SUVR, MT tau SUVR, MT volume, and hippocampal 155 volume. Z-scores for each variable were computed relative to CU,  $A\beta$ -negative 156 individuals (using an SUVR cutoff of 1.11 applied to the A<sup>β</sup> composite to determine A<sup>β</sup>-157 negativity, as recommended by the ADNI PET Core). Z-scores were dichotomized 158 using cutoffs of 2 and 2.5 standard deviations away from the CU, A<sub>β</sub>-negative mean 159 value. GMM binarization was implemented by fitting two-component Gaussian mixtures 160 to the distribution of continuous variables. A cutoff point was estimated as the curve

161 intersection between the fitted Gaussians. GMMs were omitted for hippocampal162 volume, due to a lack of bimodal distribution.

163 Non-binary categorical biomarkers consisted of quartiles, binarization with an 164 indeterminate zone, and staging systems. Quartiles were computed by binning 165 continuous values at the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles. Binarization with an 166 intermediate zone (BIZ) was used to model the uncertainty of assigning individuals who 167 display biomarker values near the cutoff threshold<sup>2</sup>. BIZ was implemented with a GMM, 168 where individuals were marked as uncertain if they showed less than 60% probability of 169 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 A $\beta$  or tau pathology. For A $\beta$ , we applied two previously published staging models<sup>9,37</sup>. 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.

#### 176 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 Aβ<sub>42</sub>,
Aβ<sub>40</sub>, total-tau (tTau), and tau phosphorylated at threonine 181 (pTau181). CSF
processing was administered by the ADNI Biomarker Core at the University of
Pennsylvania.

182 CSF concentrations of analytes were grouped into biomarker categories as 183 previously recommended<sup>14,151</sup> (A $\beta$ : A $\beta_{42}$ , A $\beta_{40}$ , A $\beta_{42}$ /A $\beta_{40}$ , tau: pTau181,

184 neurodegeneration: tTau). Raw concentrations were included as continuous measures.

185 GMMs were used to define binary versions of CSF biomarkers. BIZ and quartiles were

186 used to generate non-binary categorical versions.

It is made available under a CC-BY-NC 4.0 International license .

## 187 Statistical analyses

188 We ran a series of cross-validated modeling experiments to assess how different 189 biomarker definitions compared in their ability to model cognition. Complete details of 190 these experiments are provided in the eMethods. Briefly, we used linear regression 191 models to predict PHC<sub>Global</sub> using single or multiple biomarker definitions. All regression 192 models also included covariates of age, sex, and APOE E4 carriership. Models which 193 only included these covariates were included as controls. Models with single 194 biomarkers were used to test (1) if all included biomarker definitions improved cognitive 195 prediction accuracy and (2) which individual definitions were most predictive of cognitive 196 impairment. Next, we developed linear models combining multiple biomarker definitions 197 as predictors of PHC<sub>Global</sub>. To limit comparisons for these models, biomarkers were 198 grouped based on the underlying pathology (AT(N)) and the variable type (binary, non-199 binary categorical, continuous), with nested cross-validation used to select the best 200 predicting definition within each group. These models were used to test (3) if 201 combination of biomarkers improved prediction accuracy, and (4) if models 202 incorporating continuous or non-categorical binary biomarkers outperformed models 203 with binary biomarkers. Next, we trained support vector machines (SVM) to predict 204 PHC<sub>Global</sub> from regional biomarker values to test if (5) multivariate modeling of AD 205 pathology could improve prediction accuracy beyond that of pre-defined biomarker 206 definitions. SVMs were trained with both linear and non-linear kernels with a grid 207 search to select optimal hyperparameters (see eMethods).

208 All cross-validation experiments had 10 outer folds and were repeated 10 times 209 to generate 100 out of sample error estimates for each tested model. Model error was 210 assessed using root mean squared error (RMSE). Boxplots included in the Results 211 show distributions of the 100 out-of-sample RMSE measurements from trained models. 212 To statistically evaluate differences in accuracy for comparisons of interest, we used 213 Nadeau-Bengio t-tests. The Nadeau-Bengio t-test includes a bias correction for the 214 interdependency of out-of-sample error estimates when using repeated, cross-validated 215 designs<sup>38,39</sup>. All tests were corrected for multiple comparisons with a false discovery 216 rate method<sup>40</sup>. We investigated feature importance by plotting the distribution of

It is made available under a CC-BY-NC 4.0 International license .

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

221 Finally, we ran a series of additional cross-validated experiments using 222 alternative features, target variables, or clinical disease states. Specific experiments 223 were as follows: (a) predicting the prospective slope of PHC<sub>Global</sub> instead of the cross-224 sectional value, (b) predicting neuropsychological domains instead of  $PHC_{Global}$ , (c) 225 using PVC tau data instead of non-PVC, (d) using CSF-based biomarkers instead of 226 imaging-based versions, and (e) using only CU or CI individuals for model selection and 227 out-of-sample evaluation. For (a), we only included individuals who had longitudinal 228 cognitive measurements following baseline. To estimate longitudinal change in 229 PHC<sub>Global</sub>, linear mixed effect models were fit to model longitudinal scores following the 230 baseline assessment. Models were fit with random slopes and intercepts for 231 participants.

#### 232 Data availability

233 Data used in the preparation of this article were obtained from the Alzheimer's 234 Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was 235 launched in 2003 as a public-private partnership, led by Principal Investigator Michael 236 W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, 237 other biological markers, and clinical and neuropsychological assessment can be 238 combined to measure the progression of MCI and early AD. For up-to-date information, 239 see www.adni-info.org. All data used in this study are accessible from ADNI following 240 formal data usage agreements. Data were downloaded on May 10<sup>th</sup>, 2024. All R 241 (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. 242

It is made available under a CC-BY-NC 4.0 International license .

# 243 **Results**

## 244 Sample characteristics

245 We selected 490 individuals with baseline biomarker imaging (Table 1). The 246 cohort consisted of a mix of individuals with no cognitive impairment (CDR=0, n=288). verv mild dementia (CDR=0.5, n=163), and mild to severe dementia (CDR>0.5, n=39). 247 248 We observed significant differences in the distribution of age (p=0.001), sex (p=0.005), 249 A<sub>β</sub>-burden (Centilioid, p<0.001), and PHC<sub>Global</sub> (p<0.001) across dementia status. Mean 250 age and A $\beta$ -burden increased with dementia status, while PHC<sub>Global</sub> decreased. 251 Relatively more females were observed in the CU group (56.6%) than those with very 252 mild (41.1%) or mild to severe dementia (43.6%). APOE E4 status was not significantly 253 different across groups (p=0.200).

We also selected subsamples of individuals who had longitudinal PHC<sub>Global</sub>
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.

## 258 Assessment of modeling performance for biomarkers

259 Relative to a control model which included covariates (RMSE=0.531), almost all 260 tested biomarkers led to a significant improvement in prediction accuracy for modeling 261 cognitive scores (Figure 1). The only exception was hippocampal volume binarized at -262 2.5 Z-scores (RMSE=0.525 [0.06], p=0.09). While these results indicated that most 263 biomarker definitions provided some predictive value, gains in performance were not 264 equal across pathologies and variable types (range in RMSE reduction: 3.4-21.1%). 265 Tau biomarkers led to the largest improvements in accuracy, with 9/10 of the best 266 performing biomarkers being tau-based. Furthermore, SVM models which were trained 267 on regional pathology were more accurate than linear models using single biomarker 268 definitions. The tau SVM was the best performing model overall (RMSE=0.419 [0.05], 269 p<0.001), while the A $\beta$  SVM (RMSE=0.472 [0.06], p<0.001) and volume SVM 270 (RMSE=0.452 [0.05], p<0.001) were the best performing A<sub>β</sub> and neurodegeneration

It is made available under a CC-BY-NC 4.0 International license .

- 271 models, respectively. Outside of SVMs, the best performing models for each AT(N)
- 272 category were the A $\beta$  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
- 274 (RMSE=0.471 [0.05], p<0.001).

## 275 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<sub>CAT</sub>=0.443, T<sub>CON</sub>=0.440, N<sub>BIN</sub>=0.495, N<sub>CAT</sub>=0.492, N<sub>CON</sub>=0.471; all p<0.01). Like our experiments (Figure 1), benefits were largest for tau predictors relative to A $\beta$  and neurodegeneration.

283 Combination models which included assessments for all AT(N) categories 284 generally outperformed models with only one category included. All combination 285 models were more accurate than the covariate-only model in predicting global cognition 286 (mean RMSEs: ABIN/TBIN/NBIN=0.446, ACAT/TCAT/NCAT=0.428, ACON/TCON/NCON=0.415, 287 A<sub>SVM</sub>/T<sub>SVM</sub>/N<sub>SVM</sub>=0.405, all p<0.001). Additionally, models which combined biomarkers 288 resulted in significantly higher accuracy than models which only included one pathology 289 assessment (Figure 2A). The benefit of combination was evident for binary (ABIN/TBIN/NBIN vs. ABIN: t=3.96, p<0.001; vs. TBIN: t=2.61, p<0.05; vs. NBIN: t=3.92, 290 291 p<0.001), non-binary categorical (A<sub>CAT</sub>/T<sub>CAT</sub>/N<sub>CAT</sub> vs. A<sub>CAT</sub>: t=4.46, p<0.001; vs. T<sub>CAT</sub>: 292 t=2.22, p<0.05; vs. N<sub>CAT</sub>: t=4.31 p<0.001), and continuous (A<sub>CON</sub>/T<sub>CON</sub>/N<sub>CON</sub> vs. A<sub>CON</sub>: 293 t=5.22, p<0.001; vs. T<sub>CON</sub>: t=3.34, p<0.01; vs. N<sub>CON</sub>: t=4.48 p<0.001) biomarkers. The 294 combination SVM outperformed the A<sub> $\beta$ </sub> (t=5.29, p<0.001) and gray matter (t=3.62, 295 p<0.001) SVMs, but not the tau SVM (t=1.56, p=0.12), indicating that the improved 296 accuracy of the multimodal SVM may be primarily driven by tau.

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

It is made available under a CC-BY-NC 4.0 International license .

- $A_{BIN}/T_{BIN}/N_{BIN}=0.446$ ), reductions in error were seen when incorporating non-binary
- 301 categorical tau (A<sub>BIN</sub>/T<sub>CAT</sub>/N<sub>BIN</sub>: RMSE=0.426, t=2.14, p<0.05), continuous tau
- 302 (A<sub>BIN</sub>/T<sub>CON</sub>/N<sub>BIN</sub>: RMSE=0.424, t=2.42, p<0.05), or continuous neurodegeneration
- 303 (A<sub>BIN</sub>/T<sub>BIN</sub>/N<sub>CON</sub>: RMSE=0.432, t=2.10, p<0.05) biomarkers. An improvement was also
- 304 observed when including all continuous biomarkers (A<sub>CON</sub>/T<sub>CON</sub>/N<sub>CON</sub>: RMSE=0.415,
- 305 t=2.99, p<0.05), but not when including all non-binary biomarkers (A<sub>CAT</sub>/T<sub>CAT</sub>/N<sub>CAT</sub>:
- 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  $A\beta$  definition with a non-binary categorical
- 309 (A<sub>CAT</sub>/T<sub>BIN</sub>/N<sub>BIN</sub>: RMSE=0.446, t=0.26, p=0.60) or continuous (A<sub>CON</sub>/T<sub>BIN</sub>/N<sub>BIN</sub>:
- 310 RMSE=0.447, t=0.24, p=0.71) version did not improve accuracy. Improvements were
- also not seen for A $\beta$  and neurodegeneration SVMs (p>0.05).
- We found no differences in accuracy between models which used PVC for tau 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 A $\beta$ , 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 A<sup>β</sup> models and 100% of tau 326 models.

327 Inspection of model coefficients highlighted the relative importance of tau for328 predicting cognition. For linear models which included multiple biomarkers as

It is made available under a CC-BY-NC 4.0 International license .

329 predictors, coefficients were highest for tau, followed by neurodegeneration and A $\beta$ 330 (Figure 3B). When considering continuous biomarkers in particular, weights for A $\beta$  were 331 much lower than those of tau or neurodegeneration. Similarly, the weights for tau 332 features were on average higher than those of A $\beta$  or atrophy in the multimodal SVM 333 model (Figure 3C). Cortical weights for tau and neurodegeneration highlighted medial 334 and lateral temporal structures, while A $\beta$  weights were more homogenous. Subcortical 335 regions were weighted lower than cortical ones, except for tau uptake and gray matter 336 volume in the amygdala and hippocampus. Similar spatial patterns were observed 337 when considering the SVM weights from separate A $\beta$ , tau, and neurodegeneration 338 models (eFigure 2).

339 We used our cross-validated modeling to identify the optimal cutoffs for A $\beta$  and 340 tau binarization in our cognitive modeling experiments. Our results indicated SUVR 341 cutoffs of 1.26 (range: 1.24-1.30) for A $\beta$  and 1.44 (range: 1.33-1.45) for tau (Figure 4).

#### 342 Cognitive modeling in CU and CI populations

343 We also trained separate models to optimize the prediction of cognitive scores 344 for individuals who were CU (CDR=0) and those who were CI (CDR>0). Like our 345 results in the whole sample, we found that almost all biomarker models tested resulted 346 in a significant improvement in accuracy relative to a baseline model with just covariates 347 (eFigure 3). These benefits were observed for both CU (range in RMSE reduction: 3.9-348 15.1%) and CI (range in RMSE reduction: 7.2-30.2%) settings. The only exception was 349 for categorical neurodegeneration models in those CU, where the difference was non-350 significant (t=1.65, p=0.05). The best performing models in the CU and CI populations 351 were the all-continuous ( $A_{CON}/T_{CON}/N_{CON}$ : RMSE=0.349, t=6.1, p<0.001) and AT(N) 352 SVM models (A<sub>SVM</sub>/T<sub>SVM</sub>/N<sub>SVM</sub>: 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 A $\beta$ , provided additional accuracy for modeling PHC<sub>Global</sub> in CU and CI individuals (eFigure 4). Larger benefits were seen for models including nonbinary tau and neurodegeneration in CI individuals relative to CU individuals. For CU individuals (mean RMSE for A<sub>BIN</sub>/T<sub>BIN</sub>/N<sub>BIN</sub>: 0.381), we observed improvements when

It is made available under a CC-BY-NC 4.0 International license .

- 358 including non-binary categorical tau (A<sub>BIN</sub>/T<sub>CAT</sub>/N<sub>BIN</sub>: RMSE=0.366, t=2.8, p<0.05),
- 359 continuous tau (A<sub>BIN</sub>/T<sub>CON</sub>/N<sub>BIN</sub>: RMSE=0.366, t=2.6, p<0.05), or continuous biomarkers
- 360 (A<sub>CON</sub>/T<sub>CON</sub>/N<sub>CON</sub>: RMSE=0.359, t=2.5, p<0.05). Considering CI people (mean RMSE
- 361 for A<sub>BIN</sub>/T<sub>BIN</sub>/N<sub>BIN</sub>: 0.523), we observed increases in accuracy for the categorical tau
- 362 (Авіл/Тсат/Nвіл: RMSE=0.490, t=2.6, p<0.05), continuous tau (Авіл/Тсол/Nвіл:
- 363 RMSE=0.485, t=3.3, p<0.01), continuous neurodegeneration (A<sub>BIN</sub>/T<sub>BIN</sub>/N<sub>CON</sub>:
- 364 RMSE=0.495, t=3.4, p<0.01), continuous AT(N) (A<sub>CON</sub>/T<sub>CON</sub>/N<sub>CON</sub>: RMSE=0.479, t=3.4,
- 365 p<0.01), tau SVM (T<sub>SVM</sub>: RMSE=0.484, t=2.4, p<0.05), and AT(N) SVM
- 366 (A<sub>SVM</sub>/T<sub>SVM</sub>/N<sub>SVM</sub>: RMSE=0.442, t=4.8, p<0.001) models.
- 367 Modeling longitudinal cognition

368 The pattern of results we observed were largely consistent when modeling 369 prospective change in cognition. All model varieties tested were significantly more 370 accurate for predicting longitudinal change in PHC<sub>Global</sub> (range in RMSE reduction: 4.1-371 30.3%, all p<0.05) relative to a covariate-only model (eFigure 5a), with the largest 372 benefits seen for the all-continuous (A<sub>CON</sub>/T<sub>CON</sub>/N<sub>CON</sub>: RMSE=0.386, t=5.83, p<0.001), 373 multimodal SVM ( $A_{SVM}/T_{SVM}/N_{SVM}$ : RMSE=0.368, t=5.44, p<0.001), and tau SVM ( $T_{SVM}$ : 374 RMSE=0.373, t=5.44, p<0.001). No biomarker linear models with non-binary measures 375 improved prediction accuracy relative to an all-binary baseline (eFigure 5b, all p>0.05). 376 However, the tau SVM (T<sub>SVM</sub>: RMSE=0.373, t=3.05, p<0.01) and multimodal SVM

377  $(A_{SVM}/T_{SVM}/N_{SVM}: RMSE=0.368, t=3.54, p<0.01)$  were still significantly more accurate at

378 predicting change in PHC<sub>Global</sub> than a model consisting of all-binary predictors.

# 379 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 SVMbased assessments of tau and neurodegenerative pathology (eFigure 6). Models which included non-binary definitions of A $\beta$  alone did not surpass the all-binary model for any neuropsychological domain. Continuous tau and neurodegeneration measures improved accuracy for prediction of PHC<sub>EF</sub> (p<0.05) and PHC<sub>Visual</sub> (p<0.05), while non-

It is made available under a CC-BY-NC 4.0 International license .

- 387 binary categorical measures only improved prediction of PHC<sub>Visual</sub> (p<0.05). The AT(N)
- 388 SVM had significantly higher accuracy for modeling PHC<sub>EF</sub> (p<0.01), PHC<sub>Visual</sub> (p<0.05),
- 389 and PHC<sub>Memory</sub> (p<0.05). No significant differences were observed for PHC<sub>Language</sub>
- 390 prediction accuracy (all p>0.05).

## 391 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 imagingbased 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), nonbinary categorical (t=3.68, p<0.01), and continuous (t=3.96, p<0.01) biomarker definitions.

# 399 **Discussion**

400 AD biomarkers differ from each other along various axes such as the underlying 401 pathology they measure (e.g.  $A\beta$ , tau), the modality (e.g., imaging, CSF, blood), and 402 measurement characteristics (e.g., variable type). Our analyses indicate that 403 differences along these dimensions result in considerable variability when imaging 404 biomarkers are utilized in downstream tasks. We show that even biomarkers which 405 assess the same pathology exhibit a range in accuracy when applied for modeling 406 cognition. Additionally, we demonstrate that multivariate machine learning approaches 407 can surpass traditional biomarker definitions for cognitive prediction in AD. Careful 408 consideration should be applied when selecting biomarker definitions for predictive 409 tasks, as certain operationalizations may be relatively less informative than other 410 variants. While we specifically focus on cognitive prediction, our results may be 411 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

It is made available under a CC-BY-NC 4.0 International license .

415 associations with cognition than assessments of A $\beta$  or neurodegeneration, a well-416 documented finding.<sup>42–44</sup> Feature importance analyses indicated that tau predictors 417 were weighted higher than A $\beta$  or neurodegeneration measures in models which 418 incorporated all three AT(N) categories. However, combined AT(N) models generally 419 outperformed unimodal ones, even when tau was the single biomarker included. Thus, 420 while measures of tau are important indicators of cognitive decline, incorporation of 421 measures spanning other pathologies is warranted for enhancing predictive accuracy.

422 We observed that SVM models outperformed more traditional linear models of 423 cognition. Tau and multimodal SVMs were the best predictors overall in many 424 experiments. These models were also the only models which outperformed binary 425 biomarker models for prediction of longitudinal cognitive decline. A $\beta$  and 426 neurodegeneration SVMs were also relatively stronger than other individual biomarker 427 definitions of these pathologies. The superior performance of SVM models suggest that 428 there may be key predictive signal occurring in brain regions external to the manually 429 defined meta-ROIs which are utilized in most of the biomarker definitions we tested. 430 However, the SVMs also allowed for non-linear transformations of input features, 431 making them relatively more powerful models.

432 Inclusion of non-binary tau and neurodegeneration predictors led to small but 433 consistent improvements in accuracy relative to binary alternatives. However, binary Aß 434 measures performed equally to non-binary ones. As such, our findings indicate that 435 binarization along dimensions of tau and neurodegeneration (e.g., labeling individuals 436 as T+/- or N+/-) may obfuscate information relevant to the prediction of cognitive 437 decline. On the other hand, dichotomization of AB status may be sufficient. These 438 notion agrees with revised criteria for diagnosis and staging of AD<sup>2</sup>: their proposed PET 439 staging system includes binary assessment of A<sup>B</sup> (i.e., has AD or not) and multi-level 440 staging of tau based on the extent of progression outside the medial temporal lobe. 441 Interestingly, the specific benefits for tau and neurodegeneration (and not  $A\beta$ ) were 442 consistent when considering only CU or CI individuals and when modeling some 443 neuropsychological domains (executive functioning and visuospatial performance). 444 These results agree with a previous study which found similar non-dichotomized tau

It is made available under a CC-BY-NC 4.0 International license .

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

448 While nearly all the imaging biomarkers we tested improved prediction of 449 cognitive impairment, the same was not true for CSF counterparts. Models which 450 incorporated CSF biomarkers as predictors did not perform better than baseline models 451 which only included standard covariates, regardless of the analyte or its 452 operationalization. Previous findings have similarly demonstrated stronger associations 453 for imaging biomarkers and cognitive scores in AD, relative to fluid biomarkers<sup>45</sup>. 454 Importantly, the CSF analytes tested largely reflect earlier pathological cascades which 455 likely develop and saturate prior to the onset of neurodegeneration and cognitive 456 decline<sup>2,46</sup>. As such, they may be less suited for providing direct associations with 457 cognitive decline, and more suited for diagnosis or prediction of future decline.

458 Our study has limitations which should be considered. First, while we performed 459 a comprehensive and rigorous evaluation of multiple biomarkers, biomarkers such as 460 fluorodeoxyglucose-PET, cortical thickness, and functional imaging were not included in 461 this study. Future studies are warranted to examine them. Second, this study relied 462 only on ADNI because inclusion of other sources posed issues of harmonization and 463 biomarker availability. While ADNI is one of few databases which can enable the 464 analyses we conducted, it is also relatively limited in its inclusion of demographic 465 diversity<sup>47</sup>. As such, our results warrant replication in other datasets.

466 The growth of large data initiatives has led to an explosion of approaches for 467 biomarker assessment of AD. While picking from the myriads of methods for 468 quantification of AD pathology, it is important for researchers to mind the biological, 469 statistical, and practical characteristics of each approach. Our results demonstrate that 470 different operationalizations of the same pathology can result in variable performance 471 for downstream predictive tasks. More complex indices of pathology may be superior to 472 dichotomous alternatives, particularly for measures of neurodegeneration and tau. 473 Finally, data-driven, machine learning approaches may be preferable for identifying 474 biomarker contributions to cognitive decline.

# 475 Acknowledgement

476 The authors thank the staff for the Washington University Center for High

477 Performance Computing who helped enable this work. Computations were performed

478 using the facilities of the Washington University Research Computing and Informatics

479 Facility (RCIF). The RCIF has received funding from NIH S10 program grants:

480 1S10OD025200-01A1 and 1S10OD030477-01.

# 481 Study Funding

482 This work was supported by the National Institutes of Health (NIH) (R01483 AG067103) and the BrightFocus Foundation (ADR A2021042S).

484 Data collection and sharing for this project was funded by the Alzheimer's 485 Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 486 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-487 0012). ADNI is funded by the National Institute on Aging, the National Institute of 488 Biomedical Imaging and Bioengineering, and through generous contributions from the 489 following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; 490 Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, 491 Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; 492 EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; 493 Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & 494 Development, LLC.: Johnson & Johnson Pharmaceutical Research & Development 495 LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx 496 Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; 497 Piramal Imaging: Servier: Takeda Pharmaceutical Company: and Transition 498 Therapeutics. The Canadian Institutes of Health Research is providing funds to support 499 ADNI clinical sites in Canada. Private sector contributions are facilitated by the 500 Foundation for the National Institutes of Health (www.fnih.org). The grantee 501 organization is the Northern California Institute for Research and Education, and the 502 study is coordinated by the Alzheimer's Therapeutic Research Institute at the University

- 503 of Southern California. ADNI data are disseminated by the Laboratory for Neuro
- 504 Imaging at the University of Southern California.

# 505 **Disclosures**

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

# 509 **References**

- 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
- Jack CR, Andrews JS, Beach TG, et al. Revised criteria for diagnosis and staging of
   Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & amp; Dementia*. Published online June 27, 2024:alz.13859. doi:10.1002/alz.13859
- 516 3. Hansson O. Biomarkers for neurodegenerative diseases. *Nat Med*. 2021;27(6):954 517 963. 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
- 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
- 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
- 529 7. Abdelnour C, Agosta F, Bozzali M, et al. Perspectives and challenges in patient
  530 stratification in Alzheimer's disease. *Alzheimers Res Ther*. 2022;14:112.
  531 doi:10.1186/s13195-022-01055-y
- 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

- 535 9. Collij LE, Heeman F, Salvadó G, et al. Multitracer model for staging cortical amyloid
  536 deposition using PET imaging. *Neurology*. 2020;95(11):e1538-e1553.
  537 doi:10.1212/WNL.00000000010256
- 538 10. Klunk W, Cohen A, Bi W, et al. Why we need two cutoffs for amyloid imaging: Early
  539 versus Alzheimer's-like amyloid-positivity. *Alzheimer's & Dementia*. 2012;8(4,
  540 Supplement):P453-P454. doi:10.1016/j.jalz.2012.05.1208
- 541 11. Jack CR, Wiste HJ, Weigand SD, et al. Defining imaging biomarker cut-points for
  542 brain aging and Alzheimer's disease. *Alzheimers Dement*. 2017;13(3):205-216.
  543 doi:10.1016/j.jalz.2016.08.005
- 544 12. Weigand AJ, Maass A, Eglit GL, Bondi MW. What's the cut-point?: a systematic
  545 investigation of tau PET thresholding methods. *Alzheimers Res Ther*. 2022;14(1):49.
  546 doi:10.1186/s13195-022-00986-w
- 547 13. Mattsson-Carlgren N, Leuzy A, Janelidze S, et al. The implications of different
  548 approaches to define AT(N) in Alzheimer disease. *Neurology*. 2020;94(21):e2233549 e2244. doi:10.1212/WNL.00000000009485
- 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/s00259020-05152-8
- 557 16. Bucci M, Chiotis K, Nordberg A. Alzheimer's disease profiled by fluid and imaging
  558 markers: tau PET best predicts cognitive decline. *Mol Psychiatry*. 2021;26(10):5888559 5898. doi:10.1038/s41380-021-01263-2
- 560 17. Morris JC. Clinical Dementia Rating: A Reliable and Valid Diagnostic and Staging
  561 Measure for Dementia of the Alzheimer Type. *International Psychogeriatrics*.
  562 1997;9(S1):173-176. doi:10.1017/S1041610297004870
- 563 18. Mukherjee S, Choi SE, Lee ML, et al. Cognitive Domain Harmonization and
  564 Cocalibration in Studies of Older Adults. *Neuropsychology*. 2023;37(4):409-423.
  565 doi:10.1037/neu0000835
- 566 19. ADNI Study Documents. Alzheimer's Disease Neuroimaging Initiative. 2024.
   567 Accessed May 15, 2024. https://adni.loni.usc.edu/methods/documents/
- 568 20. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of
  569 neuroanatomical structures in the human brain. *Neuron*. 2002;33(3):341-355.
  570 doi:10.1016/s0896-6273(02)00569-x

- 571 21. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for
  572 subdividing the human cerebral cortex on MRI scans into gyral based regions of
  573 interest. *NeuroImage*. 2006;31(3):968-980. doi:10.1016/j.neuroimage.2006.01.021
- 574 22. Diedrichsen J. A spatially unbiased atlas template of the human cerebellum. 575 *Neuroimage*. 2006;33(1):127-138. doi:10.1016/j.neuroimage.2006.05.056
- 576 23. Baker SL, Maass A, Jagust WJ. Considerations and code for partial volume
  577 correcting [18F]-AV-1451 tau PET data. *Data in Brief*. 2017;15:648-657.
  578 doi:10.1016/j.dib.2017.10.024
- 579 24. Rousset OG, Ma Y, Evans AC. Correction for partial volume effects in PET: principle 580 and validation. *J Nucl Med*. 1998;39(5):904-911.
- 581 25. Mormino EC, Kluth JT, Madison CM, et al. Episodic memory loss is related to
  582 hippocampal-mediated beta-amyloid deposition in elderly subjects. *Brain*.
  583 2009;132(Pt 5):1310-1323. doi:10.1093/brain/awn320
- 584 26. Jagust WJ, Landau SM, Shaw LM, et al. Relationships between biomarkers in aging
  585 and dementia. *Neurology*. 2009;73(15):1193-1199.
  586 doi:10.1212/WNL.0b013e3181bc010c
- 587 27. Klunk WE, Koeppe RA, Price JC, et al. The Centiloid Project: standardizing
  588 quantitative amyloid plaque estimation by PET. *Alzheimers Dement*. 2015;11(1):1589 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
- 593 29. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. *Acta* 594 *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/s13195021-00880-x

- 33. Landau SM, Mintun MA, Joshi AD, et al. Amyloid deposition, hypometabolism, and
  longitudinal cognitive decline. *Annals of Neurology*. 2012;72(4):578-586.
  doi:10.1002/ana.23650
- 34. Su Y, Flores S, Wang G, et al. Comparison of Pittsburgh compound B and
  florbetapir in cross-sectional and longitudinal studies. *Alzheimers Dement (Amst)*.
  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 Amyloid616 PET to Predict Future Cognitive Decline and Amyloid Accumulation. *Neurology*.
  617 2021;96(4):e619-e631. doi:10.1212/WNL.000000000011214
- 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
- 38. Nadeau C, Bengio Y. Inference for the Generalization Error. *Machine Learning*.
  2003;52(3):239-281. doi:10.1023/A:1024068626366
- 39. Bouckaert RR, Frank E. Evaluating the Replicability of Significance Tests for
  Comparing Learning Algorithms. In: Dai H, Srikant R, Zhang C, eds. Advances in *Knowledge Discovery and Data Mining*. Lecture Notes in Computer Science.
  Springer; 2004:3-12. doi:10.1007/978-3-540-24775-3\_3
- 40. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and
  Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1995;57(1):289-300. doi:10.1111/j.25176161.1995.tb02031.x
- 41. Haufe S, Meinecke F, Görgen K, et al. On the interpretation of weight vectors of
  linear models in multivariate neuroimaging. *NeuroImage*. 2014;87:96-110.
  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
- 43. La Joie R, Visani AV, Baker SL, et al. Prospective longitudinal atrophy in
  Alzheimer's disease correlates with the intensity and topography of baseline tauPET. *Sci Transl Med.* 2020;12(524):eaau5732. doi:10.1126/scitranslmed.aau5732
- 44. Ossenkoppele R, Smith R, Mattsson-Carlgren N, et al. Accuracy of Tau Positron
   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
- 45. Lu J, Ma X, Zhang H, et al. Head-to-head comparison of plasma and PET imaging
  ATN markers in subjects with cognitive complaints. *Transl Neurodegener*.
  2023;12(1):34. doi:10.1186/s40035-023-00365-x
- 46. Tissot C, Therriault J, Kunach P, et al. Comparing tau status determined via plasma
  pTau181, pTau231 and [18F]MK6240 tau-PET. *eBioMedicine*. 2022;76:103837.
  doi:10.1016/j.ebiom.2022.103837
- 47. Weiner MW, Veitch DP, Miller MJ, et al. Increasing participant diversity in AD
  research: Plans for digital screening, blood testing, and a community-engaged
  approach in the Alzheimer's Disease Neuroimaging Initiative 4. *Alzheimer's & Dementia*. 2023;19(1):307-317. doi:10.1002/alz.12797

It is made available under a CC-BY-NC 4.0 International license .

# 657 **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 <sub>Global</sub>	0.81 (0.36)	0.32 (0.46)	-0.42 (0.54)	<0.001

658 **Table 1**: Sample characteristics. The last column shows p-values for significance tests

659 comparing distributions of variables across dementia status groups (CDR). Chi-squared

tests of association were used for categorical variables (sex, APOE status) and one-

661 way ANOVAs were used for continuous variables (age, Centiloid, PHC<sub>Global</sub>).

It is made available under a CC-BY-NC 4.0 International license .

# 662 Figures

	Amyloid	Tau	Neurodegeneration
Baseline -			
Hippocampus (z<-2.5)			
Tau SUVR>1.20 -			
Centiloid>15			
Hippocampus (z<-2.0)			
Aβ SUVR>1.42			•
Tau SUVR>1.21			
Tau SUVR>1.23 -			
Aβ Composite (z>2.0)			
MT volume (z<-2.5)			
Centiloid>20			•
Aβ SUVR>1.11 -			•
Aβ Composite (GMM)			
Čentiloid (BIZ)			
Aβ Composite (BIZ) +			
Čentiloid>25			
Aβ Composite (z>2.5)			
Hippocampus (Quartiles) +			
Centiloid>30			•
Hippocampus -			
Mattsson Staging			•
Centiloid -			
Aβ Composite -			
Aβ Composite (Quartiles) -			
Centiloid (Quartiles)			
Aβ SÙVR>1.30 -			•
MT volume (z<-2.0)			
Collij Staging -			
Aβ SUVR>1.24 -			
MTV (Quartiles) +			
MT tau (Quartiles) 🕂			
MT tau (z>2.0) +			
MT tau (z>2.5) +			
Tau SUVR>1.33 -			
Amyloid SVM +			
MT volume			
MT tau (GMM) 🕂			
MT taù (BIZ) 🕂			
Braak I SUVR -			
Braak V/VI SUVR			
GM SVM -			
Braak staging (6)			
Braak III/IV SŰVŔ -			
Braak staging (3)			• •
MT tau SŮVŘ -			•
Tau SVM -			
F C	2 04	0.5	
0.	5 0.4	0.5	0.0 0.7
		RMSE	

#### 663

Figure 1: Boxplots showing performance of individual biomarkers for predicting 664 PHC<sub>Global</sub>. Values plotted are RMSEs taken from out of sample predictions for 100 665 666 cross-validation instances (lower value is more accurate). The baseline model only included covariates as independent variables (mean performance for this model 667 668 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 669 the right indicate the variable type of each biomarker (BIN=binary, CAT=non-binary 670 671 categorical, CON=continuous, SVM=support vector machine). All models exhibited 672 significantly lower RMSE than the baseline model, except for hippocampal volume 673 binarized at -2.5 Z-scores [Hippocampus (z<-2.5)].

It is made available under a CC-BY-NC 4.0 International license .



675

676 Figure 2: Boxplots showing RMSE performance of combination biomarkers for predicting global neuropsychological performance (PHC<sub>Global</sub>). A. Individual and 677 678 combination biomarker models are compared against a baseline model using only 679 covariates (mean performance indicated by dotted line) to predict PHC<sub>Global</sub>. **B.** 680 Combination biomarker models with non-binary variable types are compared against a baseline model with binary biomarker definitions (mean performance indicated by dotted 681 682 line). In both panels, colors are used to indicate the variable type of included 683 biomarkers (yellow: binary, purple: non-binary categorical, red: continuous, blue: SVM). 684 Lighter coloring indicates models which only have a single pathology assessment, while darker coloring indicates models which have  $A\beta$ , tau, and neurodegeneration 685 686 biomarkers. Gold stars indicate a significant improvement in accuracy relative to the topmost model. Gray stars and bars highlight significant pairwise differences between 687 individual models. Statistical results are derived from Nadeau-Bengio t-tests with 688 689 correction for multiple comparisons (\*p<0.05, \*\*p<0.01, \*\*\*p<0.001). 690

691

692

#### It is made available under a CC-BY-NC 4.0 International license .



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<sub>BIN</sub>/T<sub>BIN</sub>/N<sub>BIN</sub>) and all continuous (A<sub>CON</sub>/T<sub>CON</sub>/N<sub>CON</sub>) models. C. Brain maps showing average regional feature importance derived from the A<sub>SVM</sub>/T<sub>SVM</sub>/N<sub>SVM</sub> model.

It is made available under a CC-BY-NC 4.0 International license .





**Figure 4**. Estimated cutoffs for A $\beta$ - and tau-PET binarization. **A.** The kernel density estimation of selected cutoffs for A $\beta$  (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

706 lines show other pre-defined cutoff values that were tested.

It is made available under a CC-BY-NC 4.0 International license .



719

720 **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

727 correction for multiple comparisons (\*p<0.05, \*\*p<0.01, \*\*\*p<0.001).