**MANUSCRIPT DRAFT**

**Modality Matters: Prediction of Cognitive Outcome Using Brain Age Derived from Different Modalities**

E. Doering1,2, G. Antonopoulos3,5, M. Hönig1,4, T. van Eimeren1,2, M. Daamen², H. Boecker², S. B. Eickhoff3,5, K. R. Patil5,5, A. Drzezga1,2,4 for the Alzheimer’s Disease Neuroimaging Initiative[[1]](#footnote-1)

1University Hospital Cologne, Clinic and Policlinic for Nuclear Medicine, Cologne, Germany, 2German Center for Neurodegenerative Diseases, Positron Emission Tomography, Bonn, Germany, 3Research Center Juelich, Brain and Behavior (INM-7), Juelich, Germany, 4 Research Center Juelich, Molecular Organization of the Brain (INM-2), Juelich, Germany, 5Heinrich-Heine-University, Institute of Systems Neuroscience, Duesseldorf, Germany

Keywords: biological age, machine learning, Alzheimer’s disease

Objectives: Brain aging is characterized by anatomical and molecular changes. Deviations from the normal aging trajectory in the form of advanced brain aging relative to chronological age (“brain age gap”, *BAG*) is associated with various neurological abnormalities. Such normal aging trajectories are typically estimated from magnetic resonance imaging (MRI), however, changes in neuronal glucose metabolism, visible on 18F-Fluorodeoxyglucose positron emission tomography (FDG-PET), likely precede anatomical changes observed on MRI. Here, we compare the accuracy (mean absolute error, MAE) of brain age estimation from FDG-PET and MRI, and we associate BAG derived from both modalities with cognitive impairment, and Alzheimer’s disease biomarkers. Furthermore, we present cutoffs for the prediction of cognitive outcome after two years. Analyses were conducted in individuals without (CN), with subjective (SCI) and with mild cognitive impairment (MCI).

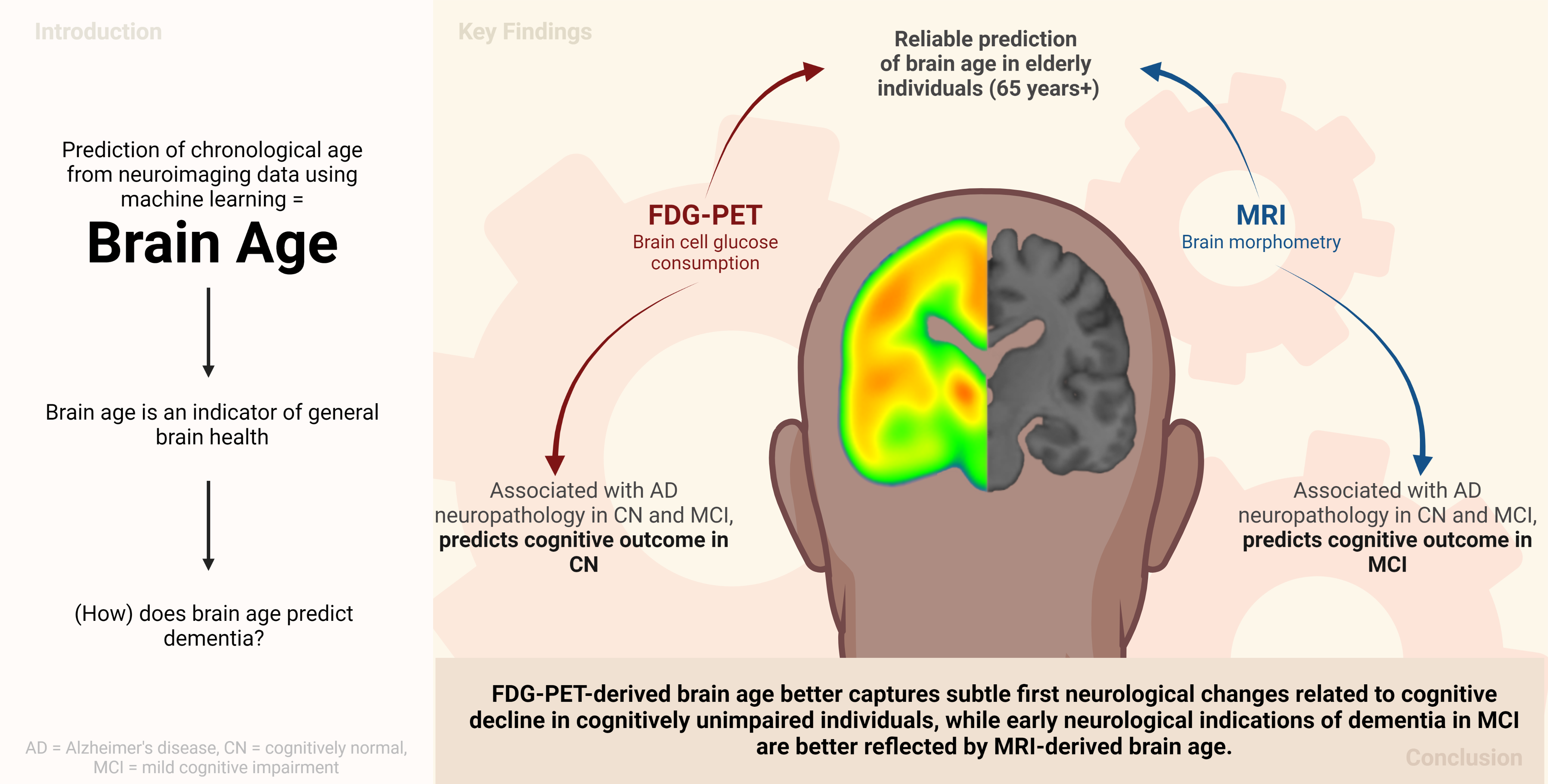
Methods: Machine learning algorithms were trained to estimate brain age from 376 matched MRI or FDG-PET scans of CN from the Alzheimer’s Disease Neuroimaging Initiative using a nested cross-validation approach. BAG was computed and correlated with measures of amyloid and tau pathology in CN and MCI (n=596). Finally, BAG was used to predict cognitive outcome after two years using cross-validated logistic regression. Cutoffs for cognitive decline were estimated from the logistic regression output.

Results: FDG-PET (MAE=2.46 years) and MRI (MAE=1.96 years) both predicted chronological age well. Both, FDG-PET- and MRI-derived BAG are correlated with amyloid load across groups and with cognitive performance in MCI. FDG-PET-derived BAG above 0.85 years is indicative of pending cognitive impairment in CN/SCI, while an MRI-derived BAG above 2.23 years suggested development of dementia in MCI.

Conclusion:

Brain age is reliably estimated from FDG-PET or MRI. FDG-PET-derived BAG better captures early neurological changes related to the development of cognitive impairment in CN/SCI, while early neurological indications of dementia in MCI are better reflected by MRI.

**Graphical Abstract**



**1 Introduction**

Brain aging entails the change in or decline of various physiological brain functions. The age of the brain can be modeled using machine learning algorithms by predicting a person’s chronological age from their neuroimaging data. As such, brain age can be seen as a proxy of overall brain health. Deviations of brain age from chronological age indicate that the brain’s age is either advanced (a positive *brain age gap, “BAG”*), or delayed (a negative BAG), and it is associated with a variety of neurological conditions, including diagnoses across the Alzheimer’s disease (AD) continuum1–5. A recent study3 showed that BAG is substantially different between individuals who will develop cognitive impairment, and those who won’t, thereby motivating further research into BAG as a prognostic biomarker of cognitive decline.

Age-related changes of the brain are most apparent from anatomical changes, such as loss of brain volume (atrophy), and molecular changes, such as a decline of neuronal metabolism (neuronal dysfunction). These two processes can be visualized by T1-weighted magnetic resonance imaging (MRI) and 18F-Fluorodeoxyglucose-PET (FDG-PET), respectively. As atrophy is preceded by neuronal dysfunction, FDG-PET is a slightly earlier indicator of neurodegeneration compared to MRI6,7. So far, estimation of brain age is typically achieved using MRI, rather than FDG-PET. Only one recent study compared FDG-PET to the standard of MRI for brain age prediction, and the study showed slightly better performance of brain age prediction when using FDG-PET3. However, in this study, FDG-PET data was not investigated independently from MRI, as FDG-PET was pre-processed using partial volume correction, thereby possibly biasing the comparison. Together, these findings argue for further exploration of FDG-PET-derived BAG, and its possibly superior performance in delineating earliest deviations from normal aging when cognitive impairment is not yet apparent. Furthermore, no BAG cutoffs for the prognosis of cognitive impairment have yet been published, but are critical for clinicians to make use of the information derived from brain age in a standardized way.

Here, we aimed to investigate the potential of FDG-PET and MRI, independently, to serve as predictors of chronological age. First, we estimated brain age in cohorts of individuals who were cognitively normal (CN), or had subjective cognitive impairment (SCI), and in cohorts of patients with mild cognitive impairment (MCI) using either FDG-PET, or MRI. Second, we compared associations of FDG-PET- and MRI-derived BAG and cognitive performance/Alzheimer’s disease neuropathology in these cohorts. Finally, we applied machine learning classification to predict cognitive outcome from BAG, and subsequently calculated a cutoff for BAG for the prognosis of cognitive impairment within.

**2 Method**

**2.1 Participants**

Baseline T1-weighted MRI and FDG-PET scans of 376 CN and SCI (*CN+SCIADNI*), and 596 individuals with MCI (*MCIADNI*) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](https://ida.loni.usc.edu/collaboration/access/adni.loni.usc.edu)), 59 MRI and FDG-PET scans of CN from the Open Access of Imaging Studies-3 (OASIS-3, https://www.oasis-brains.org/8, *CNOASIS*) database, as well as 80 MRI scans of MCI (MCIDELCODE)and 88 FDG-PET scans of SCI (SCIDELCODE) from the DZNE-Longitudinal Cognitive Impairment and Dementia Study (DELCODE)9 were acquired for the preparation of this article. To be included, participants had to be older than 60 years at the time of their scan. The primary goal of ADNI has been to test whether biological markers and clinical and neuropsychological assessment can be combined to measure the progression of MCI and dementia. Scans from the ADNI database were selected such that MRI and FDG-PET scans from the same individual were not more than one year apart for unbiased comparison of the modalities (CN: mean = 31 days, SD = 29 days[[2]](#footnote-2), 336 of 376 FDG-PET scans acquired after day of MRI scan; MCI: mean = 29 days, SD = 25 days, 532 of 596 FDG-PET scans acquired after day of MRI scan).

Diagnoses from ADNI, OASIS, and DELCODE followed the current recommendations for these groups10,11. A diagnosis of CN entailed individuals had no significant impairment in memory or cognitive functions or activities of daily living, and no significant memory concern. Individuals with SCI (*SCI,* n = 106) were also included in this cohort. To be considered SCI, either the study participant, an informant, or the clinician (ADNI)/the study participant (DELCODE) reported a significant memory concern in the absence of objective impairment of memory of cognitive function. An MCI diagnosis was provided to individuals with measurable impairment in cognitive function in the absence of dementia or significant impairments of daily living.

**2.2 Acquisition & Preprocessing of Biomarkers of Neurodegeneration**

FDG-PET scans in ADNI and OASIS were acquired dynamically 30-60 minutes (6x5min frames) after injection with an average dose of 185 MBq (5mCi). The DELCODE FDG-PET data were acquired 40-60 minutes (4x5 min frames) after injection with an average dose of 170-180 MBq. In ADNI, we used the available “Co-registered, averaged”-format, and a similar format was available in DELCODE. For OASIS, this format was manually established. Pre-processing was thereafter performed using the Statistical Parametric Mapping 12 toolbox (SPM12; [www.fil.ion.ucl.ac.uk](http://www.fil.ion.ucl.ac.uk)) in MATLAB (r2021b, The MathWorks Inc): All FDG-PET scans were aligned to the anterior commissure/posterior commissure, and subsequently co-registered and normalized to a template in standard MNI152 space. Lastly, standardized uptake value ratios (SUVR) were calculated (reference: pons12).

T1-weighted MRI scans in were acquired on according to published MRI acquisition protocols8,9,13. Scans were pre-processed using denoising (spatial-adaptive Non-Local Means), spatial registration, bias-correction, and skull-striping. Subsequently, scans were segmented by an adaptive maximum a posteriori approach14 with a partial volume model15. For non-linear transformation, the Geodesic Shooting Algorithm16 was used based on SPM12.

**2.3 Calculation of brain age**

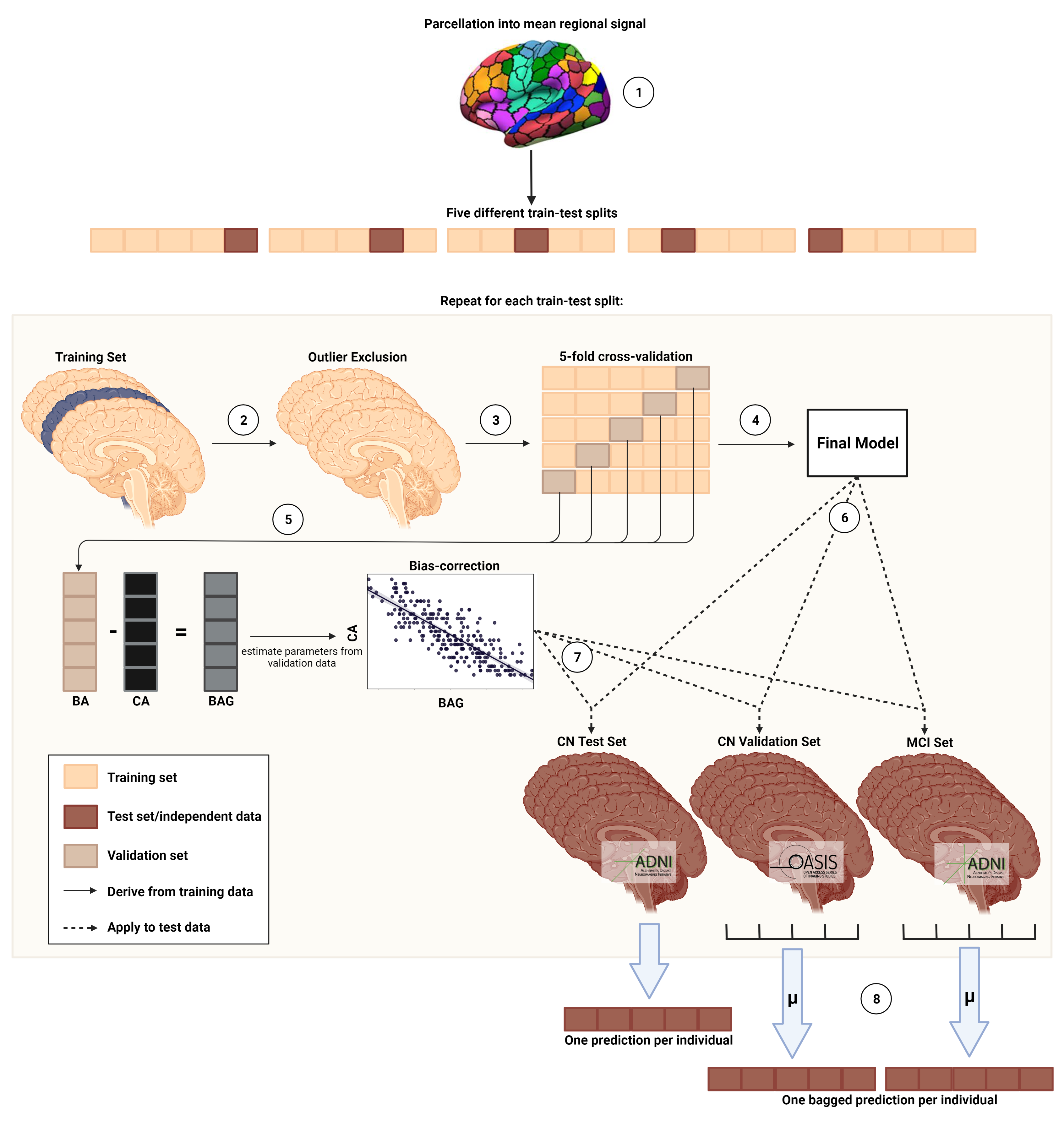
To estimate brain age, we implemented a pipeline (**Fig 1)** in Python 3.8.5 using the Julearn library (<https://juaml.github.io/julearn/main/index.html>), which in turn is based on scikit-learn17. The same pipelines were run independently for FDG-PET and MRI. First, regional averages of the signal of interest was extracted for the respective modality (FDG-PET: SUVR, MRI: gray matter volume) using a composite atlas containing 200 cortical18 and 16 sub-cortical regions19. Next, we applied a nested cross-validation approach: We repeatedly (five times) split the CN sample into different train and test sets, such that each individual occurred in a test set exactly once. Through stratification, the original proportions of young-old (65 - 74 years, ~52% of our sample), middle-old (75 - 84 years, ~40% of our sample) and oldest-old individuals (85 years+, ~8% of our sample)20 in the CN+SCIADNI sample were maintained in each train and test set. Each outer cross-validation loop consisted of outlier exclusion, an inner cross-validation yielding of a final model, estimation of parameters for bias correction, estimation of brain age in the test sets, and application of bias correction to the respective CN+SCIADNI test set, as well as the other cohorts.

**2.3.1 Outlier exclusion**

Outlier exclusion was performed in the outer cross-validation loop to ensure data quality in an automated manner. The interquartile range (IQR) was inferred from the CN+SCIADNI training set. Subjects outside 6xIQR were removed from the train and respective CN+SCIADNI test, OASISCN and SCIDELCODE sets. Importantly, as previous works have shown, MCI subjects show an advanced brain age, which translates to a reduced signal in age-relevant brain regions5. Thus, outlier exclusion was not applied to the MCI samples.

**2.3.2 Inner cross-validation**

The inner cross-validation procedure was performed for hyperparameter tuning and yielded five ‘final models’, i.e., one that was optimized for each train set. Two types of algorithms previously recommended for small sample sizes21 were implemented for brain age prediction: support vector regression (SVR) and relevance vector regression (RVR). Hyperparameter tuning was performed using five-fold stratified cross-validation (for a list of hyperparameters, see Supplementary Materials Table 1). During each iteration of the inner cross-validation, four parts of the training data were first scaled (by removing the median and scaling the data according to the IQR, “robust scaler” from the scikit-learn library) and then used to fit the models. The respective scaling parameters and fitted models were subsequently applied to the fifth part of training data, i.e., the validation set. As a result of the inner cross-validation, one optimal RVR and one optimal SVR was yielded, where “optimal” refers to the respective hyperparameter configuration that allowed for the smallest average mean absolute error (MAE) between CA and BPA across the validation set. The final model was the model with the smallest average MAE on the validation data across the remaining two optimal models.



**Fig 1. Nested cross-validation approach for brain age prediction.** Five different train-test splits were used to train and test the models. (1) Mean regional gray matter volume or SUVr were inferred from a composite atlas. (2) Outlier exclusion ranges were inferred from the training data, and applied to both the training and test data. (3) Models were trained using five-fold cross-validation. (4) The model with the smallest MAE on the validation folds was chosen as the final model. (5) BA and CA from the validation folds was used to derive bias correction parameters. (6) The final model was subsequently applied to the test sets. (7) Bias correction parameters were applied to predictions in the test set. (8) Mean of predictions across five models is considered as final prediction for CN Validation and MCI set. BA = brain age; CA = chronological age; BAG = brain age gap

**2.3.3 Bias correction**

Brain age is subject to a frequently reported bias, in which brain age of older individuals is under- and brain age of younger individuals is overestimated22, regardless of the data or method under consideration23. Here, bias correction parameters were estimated using a linear model in the validation set, and subsequently applied to all test sets. The final brain age was calculated using slope (ɑ) and an intercept (β) as follows:

**2.3.4 Precision of brain-predicted age and brain age gap**

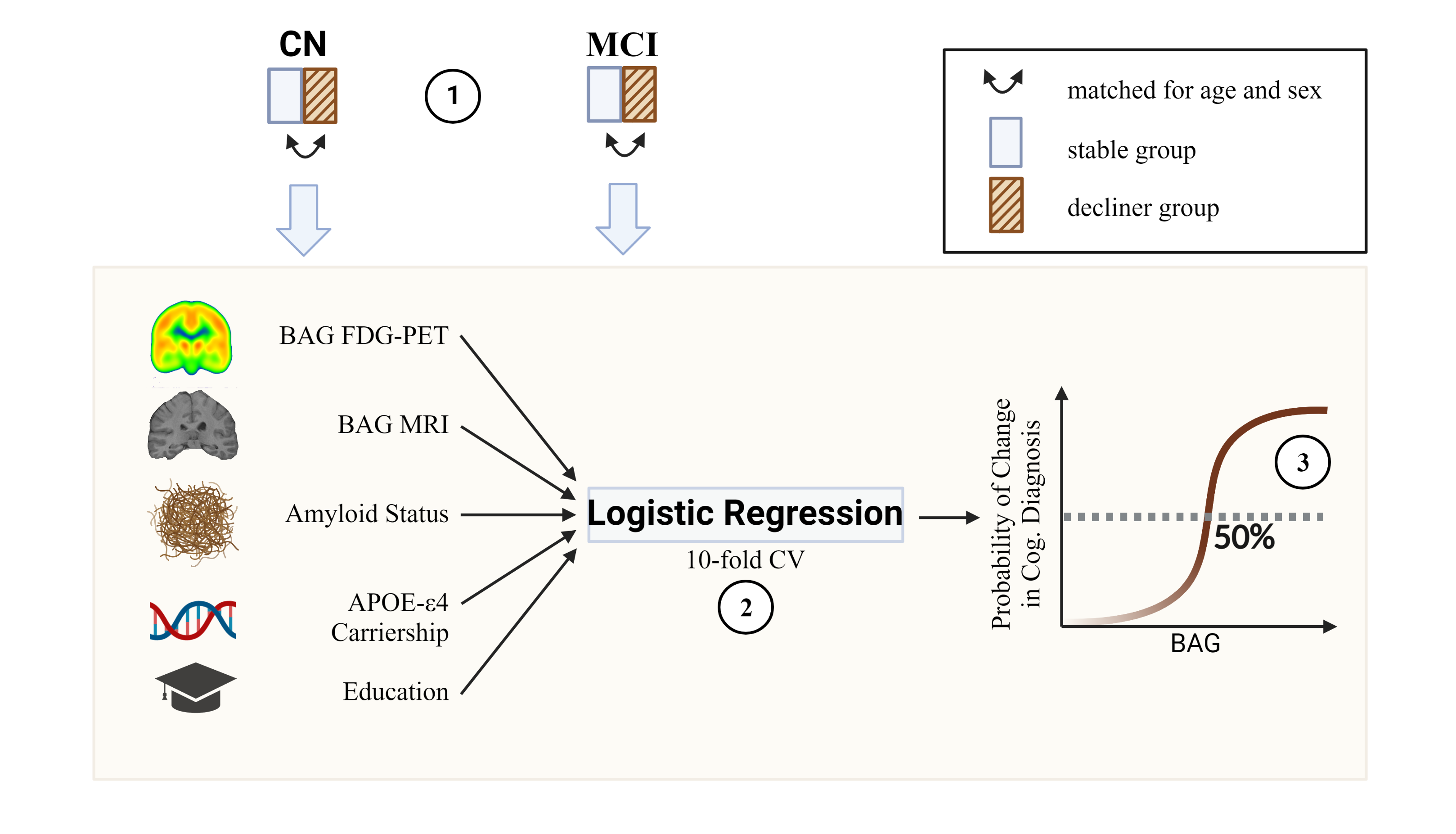
As a result of the above described nested cross-validation approach, we obtained five final models per modality. Thus, per modality, we obtained one prediction per (non-outlier) subject in the CN sample (n = 357), and five predictions per (non-outlier) subject in the CNOASIS (n = 52), SCIDELCODE (n = 88), MCIADNI (n = 596), and MCIDELCODE sample (n = 80). Feature importance of single brain regions towards this regression task was assessed by considering the learned weight coefficients of models with linear kernels. For non-linear kernels, weight coefficients are not possible to infer. Finally, for each individual, BAG was calculated as the difference between brain age and chronological age, such that a positive BAG indicated higher brain age compared to chronological age.

**2.4 Associations of brain age gap with cognitive performance, Alzheimer’s disease neuropathology, and cognitive decline**

To assess whether BAG is associated with cognitive performance, we calculated partial correlations between BAG and composite scores of memory (ADNI-MEM24), and executive function (ADNI-EF25), while correcting for age, sex, years of education and APOE-ε4 carriership status. The correlations were tested against a Bonferroni-corrected α-level of .025 (0.05/2). ADNI-MEM combines several scores used to evaluate individuals’ memory performance from the Rey Auditory Verbal Learning Test, Alzheimer’s Disease Assessment Scale and Mini Mental State Exam. ADNI-EF is a summary score of several executive function tasks, including: Category Fluency, Trails, Digit span backwards, Wechsler Adult Intelligence Scale-R Digit Symbol Substitution, Number Cancellation, and Clock Drawing items.

To assess whether BAG is associated with AD neuropathology, we calculated partial correlations between BAG and PET amyloid load (AV45-PET), as well as amyloid, tau and phosphor-tau accumulation in cerebrospinal fluid (CSF), while correcting for age, sex, years of education and APOE-ε4 carriership. The Bonferroni-corrected α-level was set to 0.0125 (0.05/4). For AV45-PET, mean SUVR are publicly available from previous analyses26–29. CSF measures of amyloid, tau and phospho-tau were acquired via lumbar puncture and analyzed using the Roche Elecsys® beta-Amyloid1-42, Total Tau and Phospho Tau181p immunoassays30. The number of tau PET scans already evaluated for SUVr in the current cohorts was too small to include this biomarker into the current analyses.

Finally, we trained a logistic regression classifier to predict cognitive outcome within two years from baseline (where BAG was assessed). Cognitive outcome was a binary variable (“stables” vs. “decliners”), based on the final diagnosis at the two year follow-up visit. Thus, CN who received a diagnosis of MCI or AD *within two years* were cognitive *decliners*, while CN who maintained the CN diagnosis *until 24 months after BAG assessment* yielded the group of *stables*. All other individuals belonging to the CN group were excluded from this analysis. For MCI, decliners were those individuals who progressed to dementia within two years, while individuals who maintained the MCI diagnosis until 24 months after BAG assessment were considered stable. Again, all other individuals were excluded this analysis. In ADNI, a diagnosis of dementia at follow-up entailed presence of dementia symptoms, abnormal memory and cognitive function and fulfillment of NINCDS/ADRDA criteria for probable AD. In DELCODE, a diagnosis of dementia at follow-up consisted of XX. MCI patients who were diagnosed as CN after two years were disregarded in the current analyses (n=29). Both in CN and MCI, we extracted a random sample of stables, matched in number, age and sex to the complete cohort of decliners. MRI- and FDG-derived BAG in these samples, together with amyloid status (CSF amyloid1-42 <= 1100 pg/ ml31), APOE-ε4 carriership and years of education, were used as input to predict cognitive outcome using a 10-fold cross-validated logistic regression classifier, as depicted in **Fig 2**. As amyloid status was not available for all individuals, analyses were conducted in two ways: once including individuals with missing amyloid information (NA values coded as 0, and amyloid negativity coded as reference; “whole samples”), and once excluding individuals with missing amyloid information (“complete samples”). Significant predictors (α=.05) of cognitive outcome were recorded. To derive a BAG cutoff for elevated risk of a change in cognitive diagnosis, a logistic regression was fitted to model the relationship between BAG from the significant BAG, and cognitive outcome. The intercept of this curve at 50% probability was set as a cutoff and validated in the current (ADNI), as well as the DELCODE sample.



**Fig 2. Estimation of a BAG cutoff for cognitive decline.** (1) A stable group was created matched in age and sex to the group of all decliners in CN or MCI. (2) 10-fold cross-validated prediction of cognitive decline within two years was conducted with FDG-PET and MRI BAG, as well as amyloid status, APOE-ε4 carriership and education as predictors. (3) The BAG cutoff for increased risk of cognitive decline was inferred from 50% probability of a change in cognitive diagnosis in the cross-validated classification output.

**3 Results**

**3.1 Participants**

This study included 972 MRI and FDG-PET scans (respectively) from the ADNI database (CN+SCIADNI: n = 376; MCIADNI: n = 596), as well as data from two validation cohorts. To validate brain age estimation, we used 59 MRI and FDG-PET scans (respectively) of CN from the OASIS-3. To validate BAG cutoffs for the prediction of cognitive outcome, we used 88 FDG-PET scans of SCI and 80 MRI scans of MCI patients from DELCODE. An overview of participant characteristics is presented in **Table 1**. In the cognitively unimpaired cohorts, CNOASIS and SCIDELCODE were significantly younger than CN+SCIADNI (t = 3.44, *p* < .001), especially in the MRI cohort, and the MMSE of SCIDELCODE was higher compared to CN+SCIADNI (t = -2.30, *p* = .03). Among MCI samples, MCIDELCODE had significantly less years of education (t = 6.01, *p* < .001)

**3.2 Accuracy and demographic profile of brain-predicted age**

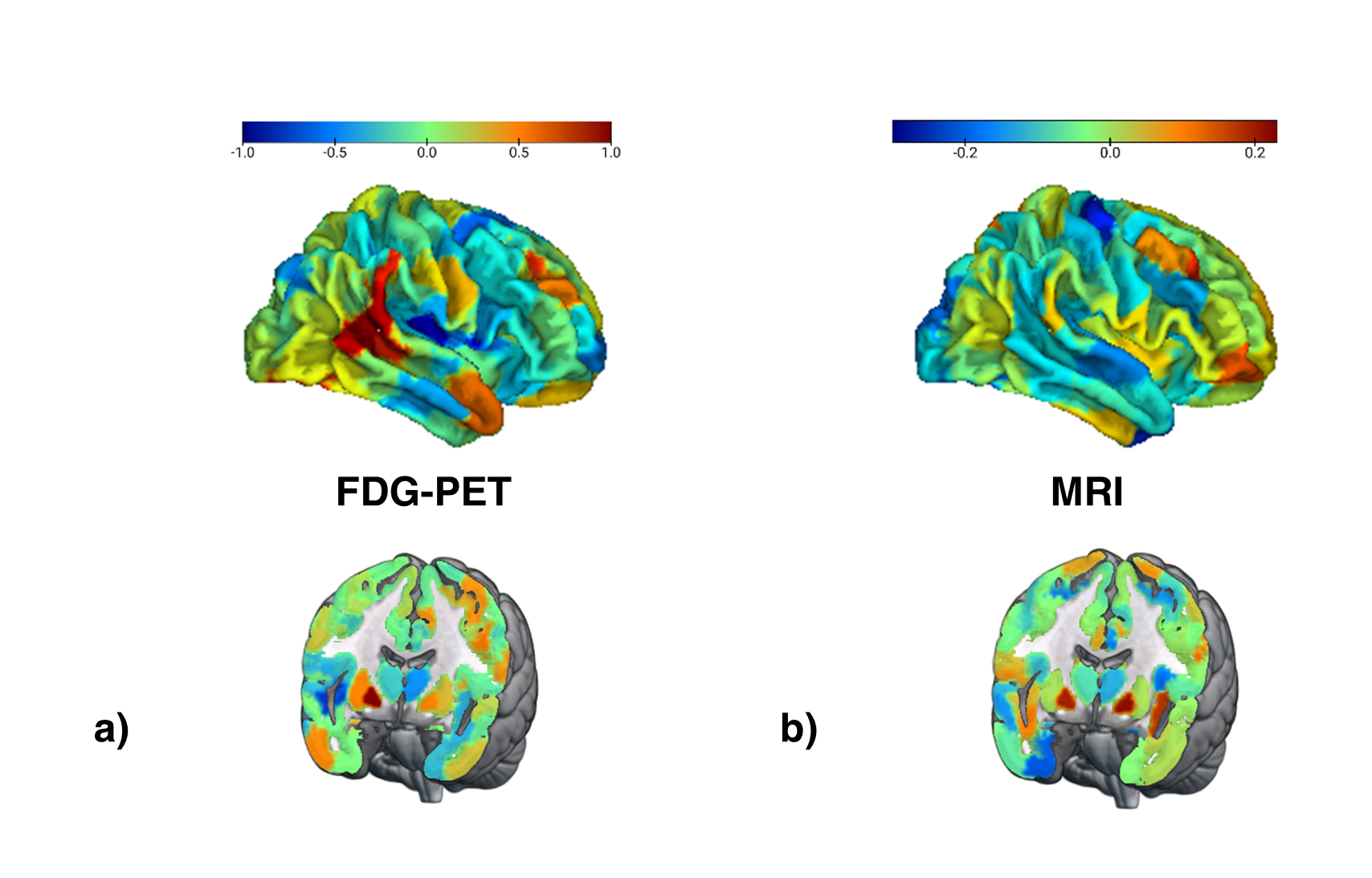
In the CN+SCIADNI group (n = 357 after outlier exclusion, test predictions), MRI and FDG-PET predicted chronological age with an MAE of 1.96 and 2.63 years, and a broad range of BAG spanning 16 and 18.7 years, respectively (see **Table 2**). MAE of MRI-derived brain age was significantly lower (paired\_t = -6.69, p = .026). Both, MRI- and FDG-PET-derived BAG were normally distributed and moderately strongly correlated across modality (r = .288, *p* < 0.001). In the external CNOASIS sample (n=52 after outlier exclusion), bagged predictions of chronological age had an MAE of 2.23 and 2.03 years for MRI and FDG-PET, respectively, thus showing high generalization performance of the models to external datasets. SCIDELCODE individuals’ brains were estimated from FDG-PET to be, on average, 2.07 years advanced in age compared to their chronological age. MCIADNI individuals’ brains were estimated to be, on average, 1.51 or 1.07 years advanced in age when predicted from MRI or FDG-PET, and, consistently, MCIDELCODE individuals’ brains were estimated to be 1.42 years older than their chronological age. Bias correction eliminated the correlation between BAG and chronological age in the CN+SCIADNI, CNOASIS, both DELCODEsamples, although a marginal (α=0.1) correlation remained between BAG and chronological age in the CNOASIS sample (MRI: r = -.242, *p* = .08, FDG-PET: r = .266, *p* = 0.06).

Women showed lower BAG compared to men in CN+SCIADNI (tMRI = -6.98, *pMRI* < .0001, tFDG-PET = -1.98, *pFDG-PET* = .05), SCIDELCODE (tFDG-PET = -2.13, *p* = 0.04), MCIADNI (tFDG-PET = -3.85, *pFDG-PET* < .001; tMRI = -5.58, *pMRI* < .0001), and MCIDELCODE (tMRI = -2.73, *pMRI* < .008), especially on MRI-derived BAG. Carriership of the APOE-ε4 allele, genetically predisposing for Alzheimer’s disease, was associated with higher BAG in MCIADNI (only MRI: tMRI = 2.72, *pMRI*= 0.007). Years of education was not correlated with BAG in any of the samples.

Model selection returned linear SVMs five out of five and four out of five times for MRI and FDG-PET, respectively. To assess the feature importance (δ) of brain age estimation in the two modalities, we extracted all brain region’s weight coefficients as learned by these models. Regional weight coefficients were strongly correlated within modalities (MRI: r = [.79, 0.89]), FDG-PET: r = [.74, .79], but average weight coefficients were not correlated between the two modalities (r = .048, *p* = .48), i.e. the regions used for brain age prediction in the two modalities were substantially different (see **Fig 3**). For FDG-PET (range: [-0.99 (right globus pallidus), 1.04 (right caudate nucleus)]), important regions (very low or very high weight coefficient) included temporal, pre-frontal, and sub-globus pallidus, nucleus accumbens, and caudate nucleus). Notably, SUVR in all highly important regions in FDG-PET (mean(δ) + 2\*SD(δ) > δ < mean(δ) - 2\*SD(δ)) were right hemispheric, and negatively correlated with chronological age. For MRI (range: [-0.30 (right hippocampus), 0.22 (right visual network)]), important regions included parietal, pre-frontal, occipital, and sub-cortical regions (e.g., hippocampus, nucleus accumbens, globus pallidus, and caudate nucleus). Again, the majority of highly important regions was right hemispheric and negatively correlated with age. A full list of highly important regions with correlation results can be found in the Supplementary materials, section Feature Importance.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 1.** Overview of samples | | | | | |
|  | CN+SCIADNI | CNOASIS | SCIDELCODE | MCIADNI | MCIDELCODE | |
| *n* total | 376 | 59 | 88 | 596 | 80 | |
| Age at PET scan [avg. years (SD)] | 73.9 (5.94) | 71.7**-** (4.22) | 70.9**-** (5.57) | 73.2 (6.93) | NA | |
| Age at MRI scan [avg. years (SD)] | 73.8 (5.92) | 70.36**-** (4.17) | NA | 73.2 (6.92) | 73.4 (5.87) | |
| Sex [%female (nNA)] | 51 (0) | 59 (0) | 41 (0) | 42 (2) | 45 (0) | |
| CSFAβ1-42 Status [%positive (nNA)] | 39 (85) | NA | 43 (28) | 65 (139) | 52 (38) | |
| MMSE [avg. score] | 29 (1.23) | 29 (1.01) | 29**+** (1.03) | 28 (1.75) | 28 (1.67) | |
| Education [avg. years (SD)] | 16 (2.71) | 16 (2.70) | 16**-** (3.00) | 16 (2.67) | 14**-** (3.06) | |
| Notes. Percentage of CSFAβ1-42 Status indicates percentage of amyloid positive individuals among all who received lumbar puncture (excluding NA). Number of individuals who did not receive lumbar puncture for Aβ1-42 is in parentheses. +significantly higher than CN+SCIADNI, -significantly lower than CN+SCIADNI. Comparisons done within modality and group, via t-test for numeric, and χ² for categorical variables, with α = 0.05. | | | | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2.** Precision of predicting chronological age from FDG-PET and MRI scans. | | | | | | |  |  |
|  | CN+SCIADNI | | CNOASIS | | MCIADNI | | SCIDELCODE | MCIDELCODE |
|  | MRI | FDG | MRI | FDG | MRI | FDG | FDG | MRI |
| *n* total | 357⁺ | 357⁺ | 52⁺ | 52⁺ | 596 | 596 | 88 | 80 |
| MAE | 1.96 | 2.63 | 2.23 | 2.03 | 2.62 | 2.51 | 2.64 | 2.62 |
| Range | [-6.91, 9.09] | [-8.72, 9.98] | [-5.44, 7.40] | [-6.70, 5.46] | [-9.07, 9.55] | [-10.91, 9.23] | [-5.49, 7.92] | [-4.64, 9.60] |
| ME | 0.05 | -0.001 | -0.71 | -0.04 | 1.51 | 1.07 | 2.07 | 1.42 |
| R² | .816 | .693 | .593 | .617 | .773 | .794 | .641 | .676 |
| *Notes.* +After outlier exclusion using CN train set (IQR > 6). Precision of bagged predictions are shown for all but the CN+SCIADNI samples. | | | | | | | | |



**Fig 3** **Feature importance for brain age prediction.** a) Average weights of support vector regression across three linear SVR for brain age prediction using FDG-PET. Weights were highly correlated across models (r > 0.7). b) Average weights of support vector regression across five linear SVR for brain age prediction using MRI. Weights were highly correlated across models (r > 0.7).

**3.3 BAG and Cognitive Performance**

Partial correlations between BAG and composite scores for memory (ADNI-MEM) and executive function scores (ADNI-EF) were calculated to evaluate whether BAG is associated with cognitive performance in the two modalities using the ADNI samples.

In CN, ADNI-MEM (*p* = .64) and ADNI-EF (*p* = .13) scores were normally distributed as assessed with the Shapiro Wilk test, thus, Pearson correlations were computed. Higher MRI-derived BAG was associated with lower ADNI-EF scores (r = -.137, *p*= .01). A marginally significant correlation was observed between FDG-PET-derived BAG and ADNI-MEM scores (r = -.100, *p* = .06).

In MCI, ADNI-MEM (*p* = .004) and ADNI-EF (*p* = .05) were not normally distributed as assessed with the Shapiro Wilk test, thus, Spearman-rank correlations were assessed. Both, MRI- and FDG-PET-derived BAG were significantly negatively correlated with ADNI-MEM (rhoMRI = -.379, *pMRI <* .0001; rhoFDG-PET = -.237, *pFDG-PET* < .0001) and ADNI-EF (rhoMRI = -.272, *pMRI <* .0001; rhoFDG-PET = -.243, *pFDG-PET*  < .0001).

**3.4 BAG and AD Neuropathology**

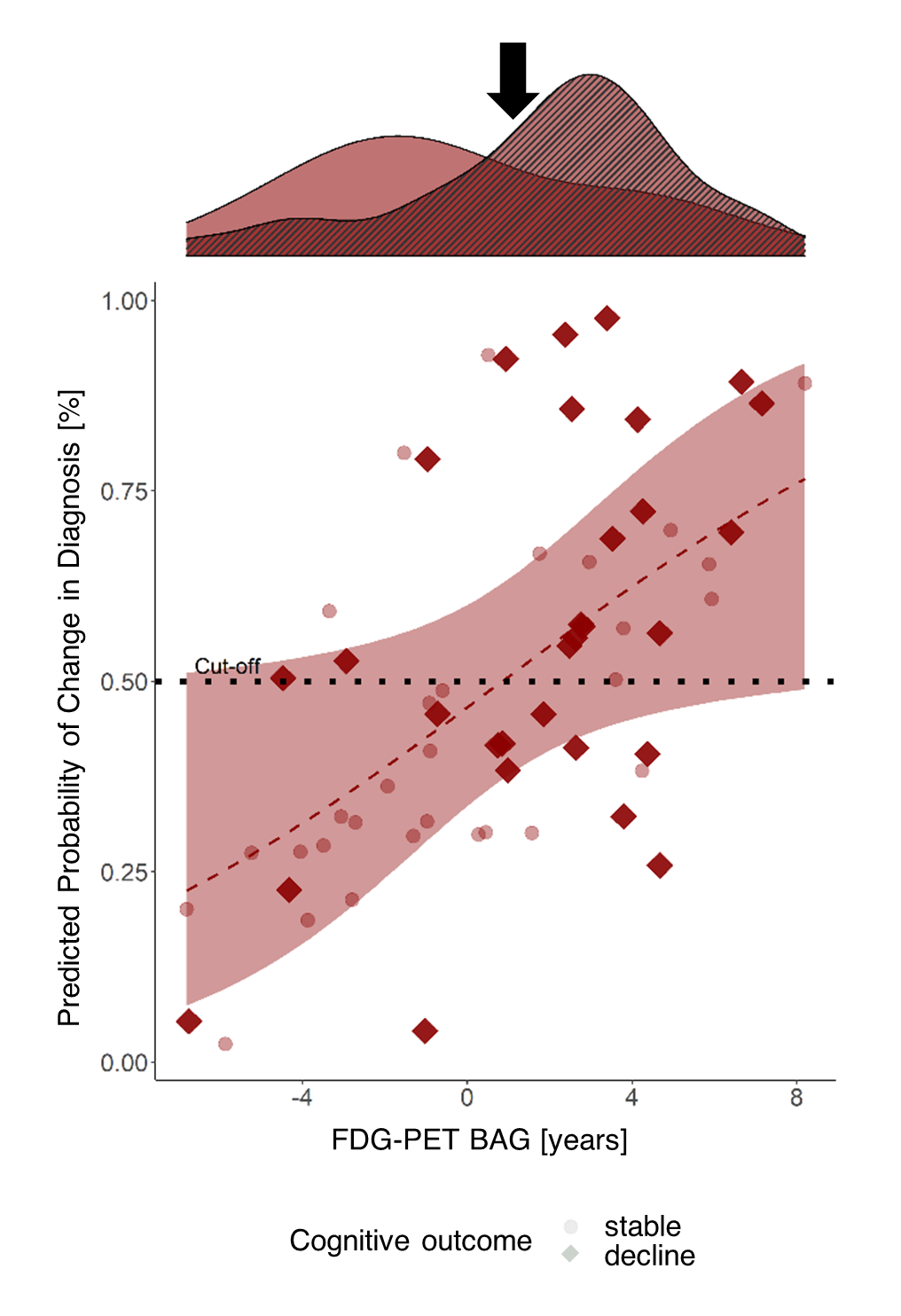
Partial correlations between BAG and PET amyloid status (global AV45), CSF β-amyloid1–42 (CSF Aβ1-42), CSF total-tau (CSF Tau) and CSF phospho-tau181 (CSF pTau181) were calculated to evaluate whether BAG is associated with AD neuropathology in the two modalities using the ADNI sample.

In CN+SCIADNI, higher MRI- and FDG-PET-derived BAG were both correlated with a decrease in CSF Aβ1-42 (rhoMRI = -.152, *pMRI* .012; rhoFDG-PET = -.152, *pFDG-PET* = .012), but none of the other pathological variables. In MCIADNI, both, higher MRI- and FDG-PET-derived BAG were associated with enhanced amyloid accumulation, as reflected by significant negative correlations with CSF Aβ1-42 (rhoMRI = -.232, *pMRI <* .0001; rhoFDG-PET = -.144, *pFDG-PET* = .002), and (marginal) positive correlations with global AV45 (rhoMRI = -.272, *pMRI <* .0001; rhoFDG-PET = -.1, *pFDG-PET*  = .026). Moreover, higher MRI-derived BAG was marginally associated with higher levels of CSF pTau181 (rhoMRI = .093, *pMRI* = .048).

**3.5 BAG and Cognitive Outcome**

To assess the potential of BAG from the two modalities to serve as an indicator of cognitive outcome, and to calculate cutoffs for elevated risk of cognitive decline in CN and MCI, cognitive outcome after two years was predicted from MRI- and FDG-PET-derived BAG, APOE-e4 carriership, CSF amyloid status and years of education. We applied 10-fold cross-validated logistic regression in a subsample containing all decliners, and an equal number of stables matched in age and sex for both, CN and MCI.

In the CN+SCIADNI sample, 278 individuals remained stable until year two, while 30 obtained a diagnosis of cognitive impairment (MCI or dementia) within two years. Consequently, a subsample of 30 stables and all 30 decliners constituted the subsample for prediction of cognitive outcome in CN. We found that, holding all other predictor variables constant, FDG-PET-derived BAG and APOE-ε4 carriership significantly predicted cognitive outcome after two years. The odds of a cognitive impairment diagnosis within two years were increased by 29% (95% CI [1.079, 1.604], *p* = .010) for every FDG-PET-derived BAG year. Moreover, the odds of developing cognitive impairment were increased by eight-fold with a positive APOE-ε4 carriership status (95% CI [1.496, 71.814], *p* = .028). To obtain a cutoff for prognoses of cognitive impairment, we fit a logistic regression model on cognitive outcome by BAG on FDG-PET. The intersection of the curve with 50% probability of receiving such a diagnosis was at 0.85 years FDG-PET BAG (see **Fig. 4**). This suggests that cognitively unimpaired individuals with a brain age advanced by 0.85 years have an elevated risk of converting to cognitive impairment. In the current CN+SCIADNI subsample, stratification by this cutoff yielded a specificity of 70.0% and a sensitivity of 66.7%. To validate the cutoff in an external dataset, we applied it to the SCIDELCODE cohort



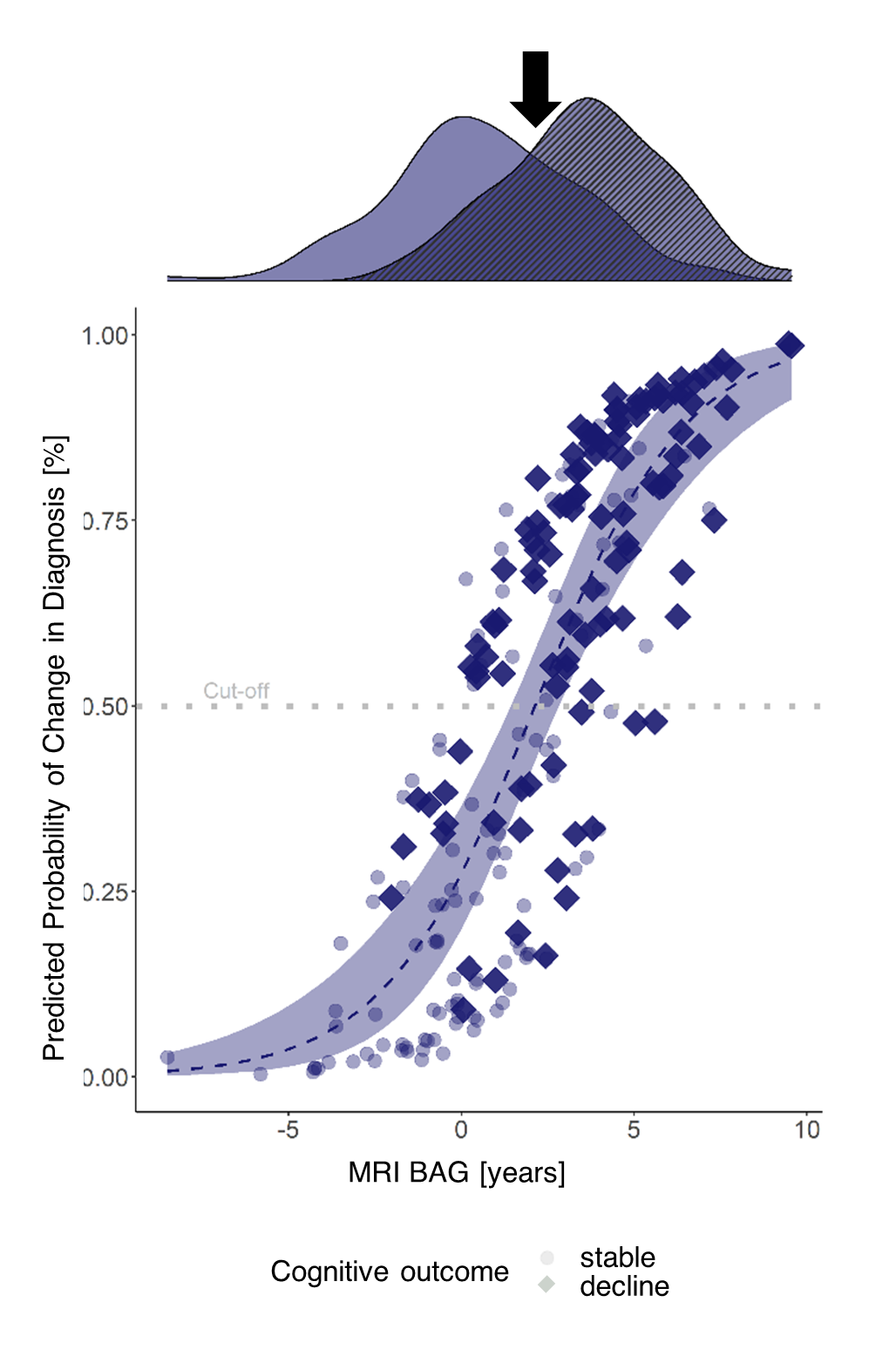
**Fig. 4** **Cross-validated probability of a change in diagnosis from CN/SCI to MCI/Dementia within two years by FDG-PET BAG.** Probability was estimated by fitting a logistic regression model on cognitive outcome by FDG-PET BAG, MRI BAG, amyloid status, apoe-e4 carriership and years of education. The red line shows the logistic regression on cognitive outcome by FDG-PET BAG, with the shaded area representing standard error. The red line intersects 50% at 0.85 years BAG on FDG-PET. The density plot above shows FDG-PET BAG distribution of stables (clear) and decliners (striped) in the subsample and the black error points to the cutoff.

(ndecliners= 8, nstables = 80), where we obtained a specificity of 87.5% and a sensitivity of 33.8%.

After removing those individuals who did not have information on amyloid status available (n=6), a “complete” sample of 24 decliners remained, thus constituting a sample size of 48. Results from the complete samples were largely consistent with results obtained from the whole samples and can be found in the Supplementary Materials, section “Prediction of Cognitive Outcome”.

Three hundred forty eight MCI patients remained stable until year two, while 113 MCI patients converted to dementia. Consequently, a subsample of 113 stables and all 113 decliners constituted the subsample for prediction of cognitive outcome in MCI. Holding all other predictor variables constant, MRI-derived BAG, a positive amyloid status in CSF, and APOE-ε4 carriership significantly predicted cognitive outcome after two years. With every one-year increase in BAG on MRI, the odds of converting to MCI or dementia were increased by 52% (95% CI [1.304, 1.788]), while a positive amyloid status and APOE-ε4 carriership increased those odds by 411% (95% CI [1.632, 11.255]) and 303% (95% CI [1.495, 6.254]), respectively. To obtain a cutoff for cognitive outcome, we fit a logistic regression model on cognitive outcome by MRI BAG. The intersection of the curve with 50% probability of receiving such a diagnosis was at 2.14 years of MRI BAG (see **Fig. 5**). Again, this suggests that if the brain age of an MCI patient is advanced by at least this cutoff value, they have an elevated risk of converting to dementia. In the current MCIADNI subsample, stratification by this cutoff yielded a specificity of 73.5% and a sensitivity of 71.7%. In the MCIDELCODE sample (nstables = 41, ndecliners = 28), the cutoff had a comparable specificity of 67.9% and a sensitivity of 73.2%.

Three hundred sixty nine MCI patients had full information for all considered variables, thus constituting the decliner group of the complete samples. Results from the complete samples were consistent with results obtained from the whole samples and can be found in the Supplementary Materials. Finally, given the correlation observed between FDG-PET- and MRI BAG, we assessed logistic regression models with unimodal BAG32. Considered in separate models, both MRI- and PET-BAG significantly predicted Cognitive Outcome in MCIADNI, while only FDG-PET predicted cognitive outcome in CN+SCIADNI (see Supplementary Tables S2 and S3 for estimates of logistic regression in whole samples).



**Fig. 5 Cross-validated probability of cognitive decline within two years after a baseline diagnosis of MCI by MRI-BAG.** Probability was estimated by fitting a logistic regression model on cognitive outcome by FDG-PET BAG, MRI BAG, amyloid status, apoe-e4 carriership and years of education. The blue line shows the logistic regression on cognitive outcome by MRI BAG, with the shaded area representing standard error. The blue line intersects 50% at 2.14 years BAG on FDG-PET. The density plot above shows FDG-PET BAG distribution of stables (clear) and decliners (striped) in the subsample and the black error points to the cutoff.

**4 Discussion**

Previous studies have mostly used MRI to estimate brain age. A recent study showed for the first time that FDG-PET, which is an earlier indicator of neurodegeneration-associated cerebral changes, could be successfully used to estimate brain age3. Here, we compared the accuracy of FDG-PET and MRI-estimated brain age, and provided a comprehensive overview of the cognitive and neuropathological profile of FDG-PET and MRI-derived BAG in different cognitive groups. We showed that 1) MRI and FDG-PET both predict chronological age accurately, 2) MRI and FDG-PET-derived BAG both reflect neuropathological abnormality in CN/SCI and MCI, as well as cognitive dysfunction in MCI, and 3) BAG derived from FDG-PET yields prognoses of cognitive outcome in CN/SCI, while MRI does the same in MCI. We further computed and validated cutoffs for the prognosis of cognitive impairment. While the potential of both MRI and FDG-PET to estimate brain age has been shown previously3, the current study confirms this finding under consideration of the two modalities in independence (i.e., without using information from one modality for pre-processing of the other, e.g. for partial volume correction). However, the better accuracy of FDG-PET-based brain age estimation reported by Lee and colleagues did not replicate in our analyses with the two modalities being considered independently.

Congruent with previous work, our findings suggest that FDG-PET shows greater and more consistent changes related to early and subtle neurodegeneration-associated changes in the brain. MRI, on the other hand, was superior in delineating dementia-related changes with an MCI7. Among the CN population, our results are likely most relevant to individuals experiencing SCI, who are 1) assumed to recognize cognitive deficits before they become clinically measurable10, 2) more likely to develop MCI or AD compared to CN33, and 3) likely to be seen by a physician compared to CN given their subjective symptoms. Prediction of cognitive outcome in this cohort based on FDG-PET BAG was moderately to highly specific, while sensitivity was low. Consistently, Lee and colleagues showed that FDG-PET BAG is significantly increased in CN converting to MCI or AD at baseline3. Together, these findings deliver strong evidence that FDG-PET BAG could complement the identification of at-risk individuals (i.e., those who have a BAG above our proposed cutoff), while further information must be gathered for individuals with SCI with a BAG below the cutoff.

The inclusion of BAG into clinical trials of AD could have several advantages. Numerous anti-amyloid therapies are currently under assessment or have recently been approved for the treatment of MCI and early AD. Since cognitive decline is often an outcome factor of these trials, a BAG cutoff could help identify those individuals who are most at risk of cognitive decline, thereby aiding in reducing the number of participants and thus cost and time of treatment trials. Moreover, BAG is an established summary marker of brain health34 and reflects various neurological abnormalities beyond AD35,36. In line with that, beyond its prognostic value, we have shown that BAG is associated with amyloid load, regardless of modality or group, and that in MCI, both MRI and FDG-PET BAG depict impairment of memory and executive function. The inclusion of BAG as a summary measure of overall brain health as a secondary outcome variable could provide useful additional information on drug efficacy.

Similar to previous studies3,7, We found a spatial disconnect between brain regions displaying aging as observed on FDG-PET and MRI. Thus, different aging processes may be observed depending on the choice of modality, which underlines the importance of consideration of the appropriate modality for a research question. The (mostly right hemispheric) regions deemed most important by our MRI and FDG-PET models have previously been described to be substrates heavily affected by aging7, and most of the regions we identified as highly important were strongly correlated with aging in our cohort. Notably, most of the highly age-affected brain regions on MRI and all on FDG-PET were right hemispheric, thus substantiating the idea that the right hemisphere exhibits greater age-related decline37,38. Moreover, temporal, parietal, and pre-frontal regions are strongly affected by AD39. Since these regions were also highly relevant in estimating brain age by our current models, our work supports the claim that AD-related neurodegeneration is a form of advanced brain aging.

Some limitations should be acknowledged. First, the sensitivity and especially specificity values of FDG-PET BAG for prognoses in cognitively unimpaired individuals, as well as of MRI BAG for prognoses in MCI patients, are not high enough for these measures to be used as a stand-alone biomarker of cognitive outcome. However, accurate prognoses, especially for cognitively unimpaired individuals, are rare and we believe that brain age with a group-dependent choice of modality can support this process. Future work should assess the combined potential of e.g., FDG-PET BAG and APOE-ε4 carriership as a prognostic biomarker of cognitive outcome. Such analysis was not possible with our current data, as it would have introduced feature leakage, and thus, biased predictions. A second limitation is the different scan protocol of DELCODE FDG-PET data (acquisition time: 40-60 min post injection) compared to ADNI and OASIS (acquisition time: 30-60 min post injection). While this could possibly have influenced the generalization performance of our models to the DELCODE cohort, we believe that the difference would not be substantial with an equal acquisition time, given that we averaged time frames over the whole acquisition time. Third, it is not a straightforward task to acquire FDG-PET scans from a CN/SCI population, as it requires logistic availability, comparably high cost, and the injection of a radioactive tracer, in the lack of an objective indication of cognitive impairment. However, especially for an SCI population, where prognoses are not yet otherwise available, BAG assessment via FDG-PET might be useful to deliver a first indicator of cognitive outcome. Whether the multi-dimensional feature space of FDG-PET can be replaced by easier accessible fluid biomarkers of neurodegeneration, and whether they would accurately reflect brain age, is debatable and an intriguing matter for future research. Moreover, the average BAG (ME) of SCIDELCODE exceeds previously reported BAG on MRI (1.1 years2). These differences possibly are possibly driven by a combination of factors, including different choice of modality, as we used FDG-PET to estimate brain age of SCIDELCODE. Whether the FDG-PET BAG is abnormally high, or whether higher FDG-PET BAG in SCI reflects very early neurological dysfunction needs further investigation. Finally, our definition of CN only required the absence of objective cognitive impairment, but not normality according to specific biomarkers. Thus, participants with and without underlying amyloid pathology were included into our training sample, which possibly introduced a confound. However, we chose to train our models to capture potential pathological heterogeneity of cognitively normal individuals to obtain a realistic estimate of brain aging.

In conclusion, we have shown that FDG-PET and MRI can both be used for brain age prediction and show different advantages depending on the analyzed group: While both, MRI- and FDG-PET BAG accurately reflect neuropathological burden across groups and cognitive performance in MCI, FDG-PET BAG can support the prognosis of cognitive outcome in cognitively unimpaired individuals. BAG on MRI, on the other hand, proves better estimation for risk of dementia in MCI. Estimation of cognitive outcome by means of our BAG cutoffs, could complement the identification of patients in need of frequent monitoring at an early time point of cognitive decline, as well as support clinical trials, both methodologically, and financially.

**References**

1. Beheshti, I., Mishra, S., Sone, D., Khanna, P. & Matsuda, H. T1-weighted MRI-driven brain age estimation in Alzheimer’s disease and Parkinson’s disease. *Aging Dis.* **11**, (2020).

2. Rokicki, J. *et al.* Multimodal imaging improves brain age prediction and reveals distinct abnormalities in patients with psychiatric and neurological disorders. *Hum. Brain Mapp.* **42**, (2021).

3. Lee, J. *et al.* Deep learning-based brain age prediction in normal aging and dementia. *Nat. Aging* (2022) doi:10.1038/s43587-022-00219-7.

4. Löwe, L. C., Gaser, C. & Franke, K. The effect of the APOE genotype on individual BrainAGE in normal aging, Mild cognitive impairment, and Alzheimer’s Disease. *PLoS One* **11**, (2016).

5. Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N. & Sauer, H. BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer’s Disease. *PLoS One* **8**, (2013).

6. Jack, C. R. *et al.* Hypothetical model of dynamic biomarkers of the Alzheimer’s pathological cascade. *The Lancet Neurology* vol. 9 119–128 (2010).

7. Dukart, J. *et al.* Generative FDG-PET and MRI Model of Aging and Disease Progression in Alzheimer’s Disease. *PLoS Comput. Biol.* **9**, e1002987 (2013).

8. LaMontagne, P. J. *et al.* OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. *medRxiv* (2019) doi:10.1101/2019.12.13.19014902.

9. Jessen, F. *et al.* Design and first baseline data of the DZNE multicenter observational study on predementia Alzheimer’s disease (DELCODE). *Alzheimer’s Res. Ther.* **10**, (2018).

10. Jessen, F. *et al.* A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer’s disease. *Alzheimer’s Dement.* **10**, (2014).

11. Albert, M. S. *et al.* The diagnosis of mild cognitive impairment due to Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease. *Alzheimer’s Dement.* **7**, 270–279 (2011).

12. Verger, A., Doyen, M., Campion, J. Y. & Guedj, E. The pons as reference region for intensity normalization in semi-quantitative analysis of brain 18FDG PET: application to metabolic changes related to ageing in conventional and digital control databases. *EJNMMI Res.* **11**, 1–7 (2021).

13. Jack, C. R. *et al.* The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging* vol. 27 (2008).

14. Rajapakse, J. C., Giedd, J. N. & Rapoport, J. L. Statistical approach to segmentation of single-channel cerebral mr images. *IEEE Trans. Med. Imaging* **16**, (1997).

15. Tohka, J., Zijdenbos, A. & Evans, A. Fast and robust parameter estimation for statistical partial volume models in brain MRI. *Neuroimage* **23**, (2004).

16. Ashburner, J. & Friston, K. J. Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation. *Neuroimage* **55**, (2011).

17. Pedregosa, F. *et al.* *Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot*. *Journal of Machine Learning Research* vol. 12 http://scikit-learn.sourceforge.net. (2011).

18. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex* **28**, (2018).

19. Tian, Y., Margulies, D. S., Breakspear, M. & Zalesky, A. Hierarchical organization of the human subcortex unveiled with functional connectivity gradients. *bioRxiv* (2020) doi:10.1101/2020.01.13.903542.

20. Suzman, R. & Riley, M. W. Introducing the ‘oldest old’. *Milbank Mem. Fund Q. Health Soc.* **63**, (1985).

21. Beheshti, I. *et al.* Predicting brain age using machine learning algorithms: A comprehensive evaluation. *IEEE J. Biomed. Heal. Informatics* (2021) doi:10.1109/JBHI.2021.3083187.

22. Beheshti, I., Nugent, S., Potvin, O. & Duchesne, S. Bias-adjustment in neuroimaging-based brain age frameworks: A robust scheme. *NeuroImage Clin.* **24**, (2019).

23. Liang, H., Zhang, F. & Niu, X. Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders. *Hum. Brain Mapp.* **40**, (2019).

24. Crane, P. K. *et al.* Development and assessment of a composite score for memory in the Alzheimer’s Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav.* **6**, (2012).

25. Gibbons, L. E. *et al.* A composite score for executive functioning, validated in Alzheimer’s Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging Behav.* **6**, (2012).

26. Landau, S., Koeppe, R. & Jagust, W. *Florbetaben processing and positivity threshold derivation Motivation for changing the threshold*. (2011).

27. Landau, S. & Jagust, W. *Florbetapir processing methods*. (2011).

28. Jagust, W. J. *et al.* Relationships between biomarkers in aging and dementia. *Neurology* **73**, 1193–1199 (2009).

29. Jagust, W. J. *et al.* The Alzheimer’s Disease Neuroimaging Initiative positron emission tomography core. *Alzheimer’s Dement.* **6**, (2010).

30. Blennow, K. *et al.* Predicting clinical decline and conversion to Alzheimer’s disease or dementia using novel Elecsys Aβ(1–42), pTau and tTau CSF immunoassays. *Sci. Rep.* **9**, (2019).

31. Hansson, O. *et al.* CSF biomarkers of Alzheimer’s disease concord with amyloid-β PET and predict clinical progression: A study of fully automated immunoassays in BioFINDER and ADNI cohorts. *Alzheimer’s Dement.* **14**, (2018).

32. Ranganathan, P., Pramesh, C. & Aggarwal, R. Common pitfalls in statistical analysis: Logistic regression. *Perspect. Clin. Res.* **8**, (2017).

33. Parfenov, V. A., Zakharov, V. V., Kabaeva, A. R. & Vakhnina, N. V. Subjective cognitive decline as a predictor of future cognitive decline a systematic review. *Dement. e Neuropsychol.* **14**, (2020).

34. Cole, J. H., Marioni, R. E., Harris, S. E. & Deary, I. J. Brain age and other bodily ‘ages’: implications for neuropsychiatry. *Molecular Psychiatry* vol. 24 (2019).

35. Eickhoff, C. R. *et al.* Advanced brain ageing in Parkinson’s disease is related to disease duration and individual impairment. *Brain Commun.* **3**, (2021).

36. Wrigglesworth, J. *et al.* Factors associated with brain ageing - a systematic review. *BMC Neurol.* **21**, (2021).

37. Goldstein, G. & Shelly, C. Does the Right Hemisphere Age More Rapidly than the Left? *J. Clin. Neuropsychol.* **3**, (1981).

38. Dolcos, F., Rice, H. J. & Cabeza, R. Hemispheric asymmetry and aging: Right hemisphere decline or asymmetry reduction. *Neurosci. Biobehav. Rev.* **26**, (2002).

39. Bischof, G. N. *et al.* Impact of tau and amyloid burden on glucose metabolism in Alzheimer’s disease. *Ann. Clin. Transl. Neurol.* (2016) doi:10.1002/acn3.339.

**Statements & Declarations**

**Author Contributions**

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by ED and GA, in support of KP and MH. KP, TvE, SE and AD jointly supervised this work. DELCODE data was provided by MD and HB. The first draft of the manuscript was written by ED and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**COMPETING INTERESTS**

MH reports no conflict of interest.

**Funding**

Maybe mention SFB?

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](https://ida.loni.usc.edu/collaboration/access/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.

1. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: <http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf> [↑](#footnote-ref-1)
2. Mean and standard deviation indicated from absolute difference in days [↑](#footnote-ref-2)