Published Ahead of Print on July 16, 2020 as 10.1212/WNL.0000000000010256

NEUROLOGY/2019/026054

OPEN

Neurology®
The most widely read and highly cited peer-reviewed neurology journal
The Official Journal of the American Academy of Neurology

AMERICAN ACADEMY OF NEUROLOGY®

Neurology Publish Ahead of Print
DOI: 10.1212/WNL.0000000000010256



# Multi-tracer model for staging cortical amyloid deposition using PET imaging

Lyduine E. Collij, MSc[1*]; Fiona Heeman, MSc[1*]; Gemma Salvadó, MSc[2] ; Silvia Ingala, MD[1]; Daniele Altomare, MSc[3,4,5] ; Arno de Wilde, MD[3]; Elles Konijnenberg, MD[3] ; Marieke van Buchem, BSc[3] ; Maqsood Yaqub, PhD[1] ; Pawel Markiewicz, PhD[6] ; Sandeep S.V. Golla, PhD[1] ; Viktor Wottschel, PhD[1] ; Alle Meije Wink, PhD[1] ; Pieter Jelle Visser, PhD[3]; Charlotte E. Teunissen, PhD[7] ; Adriaan A. Lammertsma, PhD[1] ; Philip Scheltens, PhD[3]; Wiesje M. van der Flier, PhD[3] ; Ronald Boellaard, PhD[1]; Bart N.M. van Berckel, PhD[1]; José Luis Molinuevo, PhD [2, 2.2, 2.3] ; Juan Domingo Gispert, PhD[2, 2.1,2.3] ; Mark E. Schmidt, MD[8]; Frederik Barkhof, PhD[1,6#] ; Isadora Lopes Alves, PhD[1#]; for the ALFA Study; for the Alzheimer's Disease Neuroimaging Initiative[$]; on behalf of the AMYPAD Consortium

The Article Processing Charge was funded by the AMYPAD Consortium, which is a Project funded by the IMI JU 2 initiative (Grant number 115952).

* # Both authors contributed equally to this work

[1]*Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Radiology and Nuclear Medicine, De Boelelaan 1117, Amsterdam, Netherlands*

[2]*Barcelonaβeta Brain Research Center (BBRC), Pasqual Maragall Foundation, Barcelona – Spain*

[2.1]*Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Madrid, Spain*

[2.2]*Centro de Investigación Biomédica en Red de Fragilidad y Envejecimiento Saludable (CIBERFES), Madrid, Spain*

[2.3]*Universitat Pompeu Fabra, Barcelona, Spain*

[3]*Amsterdam UMC, Vrije Universiteit Amsterdam, Alzheimer Center and department of Neurology, De Boelelaan 1117, Amsterdam, Netherlands*

[4]*Laboratory of Neuroimaging of Aging (LANVIE), University of Geneva, Geneva, Switzerland*

[5]*Memory Clinic, University Hospital of Geneva, Geneva, Switzerland*

[6]*Centre for Medical Image Computing, Medical Physics and Biomedical Engineering, UCL, United Kingdom*

[7]*Amsterdam UMC, Vrije Universiteit Amsterdam, Neurochemistry Laboratory, De Boelelaan 1117, Amsterdam, Netherlands*

[8]*Janssen Pharmaceutica NV, Beerse, Belgium*

[$]Data used in precreation 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.

**Corresponding author:** Isadora Lopes Alves, E: i.lopesalves@amsterdamumc.nl

Character count title: 76

Word count abstract: 248

Word count main text: 4785

Number of references: 33

Number of figures and tables: 7

Key words: PET imaging, Amyloid, Alzheimer's disease, Staging, Cognition

## Statistical analysis

Statistical analysis were conducted by L.E.C.[1], F.H.[1], G.E.[2], S.I.[1], A.W.[3], M.B.[3], M.Y.[1], S.S.V.G.[1], V.W.[1], A.M.W.[1], C.E.T.[7], J.D.G.[2,2.1-2.1] and I.L.A.[1]

## Study funding

The project leading to this paper has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115952. This Joint Undertaking receives the support from the European Union's Horizon 2020 research and innovation programme and EFPIA. F. Barkhof is supported by the NIHR UCLH biomedical research centre.

Data collection and sharing for this project was funded by the EU/EFPIA Innovative Medicines Initiative Joint Undertaking EMIF grant agreement (grant No115372). This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115952. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. FB is supported by the NIHR UCLH biomedical research centre. FB, VW and AMW are supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 666992.

Research of the Alzheimer Center Amsterdam is part of the neurodegeneration research program of Amsterdam Neuroscience. The clinical database structure was developed with funding from Stichting Dioraphte. Alzheimer Center Amsterdam is supported by Alzheimer Nederland and Stichting VUmc fonds. The ABIDE project was supported by a ZonMW-Memorabel grant (ABIDE; project No 733050201), in the context of the Dutch Deltaplan

Dementie, and through a grant of Piramal Imaging (PET scan costs) to the Stichting Alzheimer & Neuropsychiatrie, Amsterdam.

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

The ALFA Study is funded "la Caixa" Foundation (LCF/PR/GN17/10300004) and the Alzheimer's Association and an international anonymous charity foundation through the the TriBEKa Imaging Platform project. Authors would like to thank GE Healthcare for kindly

providing $^{18}$F-flutemetamol doses of ALFA participants and Roche Diagnostics International Ltd for kindly providing the kits for the CSF analysis of the ALFA+ participants. Additional funding has been obtained by Project RTI2018-102261-B-I00, funded by European Regional Development Fund (EDRF) / Ministry of Science and Innovation - State Research Agency (Spain). Finally, data were provided by OASIS-3: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P50AG00561, P30NS09857781, P01AG026276, P01AG003991, R01AG043434, UL1TR000448, R01EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly.

## Disclosure

**L. Collij**, **F. Heeman**, **G. Salvadó**, **S. Ingala**, **D. Altomare**, **A. de Wilde**, **E. Konijnenberg**, **M. van Buchem**, **M. Yaqub**, **P. Markiewicz**, **S. Golla**, **V. Wottschel**, and **A. Wink** report no disclosures relevant to the manuscript.

**P. Visser** has served as member of the advisory board of Roche Diagnostics. Dr Visser received nonfinancial support from GE Healthcare, research support from Biogen and grants from Bristol-Myers Squibb, EU/EFPIA Innovative Medicines Initiative Joint Undertaking, EU Joint Programme–Neurodegenerative Disease Research (JPND and ZonMw.

**C. Teunissen** has functioned in advisory boards of Roche, received non-financial support in the form of research consumables from ADxNeurosciences and Euroimmun, performed contract research or received grants from Probiodrug, Biogen, Esai, Toyama, Janssen prevention center, Boehringer, AxonNeurosciences, EIP farma, PeopleBio, Roche. Prof.dr. Teunissen received grants from the European Commission, the Dutch Research Council (ZonMW), Association of Frontotemporal Dementia/Alzheimer's Drug Discovery Foundation, The Weston Brain Institute, Alzheimer Netherlands.

**A. Lammertsma** reports no disclosures relevant to the manuscript.

**P. Scheltens** received grants from GE Healthcare, Piramal, and Merck, paid to his institution; he has received speaker's fees paid to the institution Alzheimer Center, VU University Medical Center, Lilly, GE Healthcare, and Roche.

**W. van der Flier**, **R. Boellaard**, and **B. van Berckel** report no disclosures relevant to the manuscript.

**J. Molinuevo** received advisory honoraria from Alergan, Roche diagnostics, Genentech, Novartis, Lundbeck, Oryzon, Biogen, Lilly, Janssen, Green Valley, MSD, Eisai, Alector and Raman Health. He also received research support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking AMYPAD grant agreement n° 115952; the EU/EFPIA Innovative Medicines Initiative Joint Undertaking EPAD grant agreement n° 115736; the EU/EFPIA Innovative Medicines Initiative Joint Undertaking AETIONOMY grant n° 115568; and 'la Caixa' Foundation.

**J. Gispert** has received speaker's fees from Biogen and Philips. In addition he.holds a 'Ramón y Cajal' fellowship (RYC-2013-13054), has received research support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking AMYPAD grant agreement n° 115952, and from Ministerio de Ciencia y Universidades (grant agreement RTI2018-102261).

**M. Schmidt** received funding from Janssen Pharmaceutica N.V.

**F. Barkhof** received payment and honoraria from Bayer Genzyme, Biogen-Idec, TEVA, Merck, Novartis, Roche, IXICO Ltd, GeNeuro, and Apitope Ltd for consulting; payment from the IXICOLtd, and MedScape for educational presentations; research support via grants from EU/EFPIA Innovative Medicines Initiative Joint Undertaking (AMYPAD consortium), EuroPOND (H2020), UK MS Society, Dutch MS Society, PICTURE (IMDI-NWO), NIHR UCLH Biomedical Research Centre (BRC), ECTRIMS-MAGNIMS.

**I. Lopes Alves** reports no disclosures relevant to the manuscript.

Appendix 2 - Coinvestigators - http://links.lww.com/WNL/B170
Appendix 3 - Coinvestigators - http://links.lww.com/WNL/B171
Appendix 4 - Coinvestigators - http://links.lww.com/WNL/B172

## Abstract

**Objective:** To develop and evaluate a model for staging cortical amyloid deposition using PET with high generalizability.

**Methods:** 3027 subjects (1763 Cognitively Unimpaired (CU), 658 Impaired, 467 Alzheimer's disease (AD) dementia, 111 non-AD dementia, and 28 with missing diagnosis) from six cohorts (EMIF-AD, ALFA, ABIDE, ADC, OASIS-3, ADNI) who underwent amyloid PET were retrospectively included; 1049 subjects had follow-up scans. Applying dataset-specific cut-offs to global Standard Uptake Value ratio (SUVr) values from 27 regions, single-tracer and pooled multi-tracer regional rankings were constructed from the frequency of abnormality across 400 CU subjects (100 per tracer). The pooled multi-tracer ranking was used to create a staging model consisting of four clusters of regions as it displayed a high and consistent correlation with each single-tracer ranking. Relationships between amyloid stage, clinical variables and longitudinal cognitive decline were investigated.

**Results:** SUVr abnormality was most frequently observed in cingulate, followed by orbitofrontal, precuneal, and insular cortices, then the associative, temporal and occipital regions. Abnormal amyloid levels based on binary global SUVr classification were observed in 1.0%, 5.5%, 17.9%, 90.0%, and 100.0% of stage 0-4 subjects, respectively. Baseline stage predicted decline in MMSE (ADNI: $N$=867, $F$=67.37, $p$<0.001; OASIS: ($N$=475, $F$=9.12, $p$<0.001) and faster progression towards an MMSE≤25 (ADNI: $N$=787, $HR_{stage1}$=2.00, $HR_{stage2}$=3.53, $HR_{stage3}$=4.55, $HR_{stage4}$=9.91, $p$<0.001; OASIS: $N$=469, $HR_{stage4}$=4.80, $p$<0.001).

**Conclusion**: The pooled multi-tracer staging model successfully classified the level of amyloid burden in >3000 subjects across cohorts and radiotracers, and detects pre-global

amyloid burden and distinct risk profiles of cognitive decline within globally amyloid-positive subjects.

## Introduction

Positron emission tomography (PET) can identify amyloid-β (Aβ) plaques *in vivo* with high sensitivity and specificity in clinical populations.[1] The clinical standard for determining amyloid abnormality is a dichotomous visual assessment, while classification based on thresholds of global standardized uptake value ratios (SUVr) are generally used for research purposes. However, both procedures might miss emerging amyloid pathology, and the extent of the pathological burden is generally disregarded.[2 3]

Recently, staging of amyloid burden with PET imaging has showed promising results to address the limitations of signal dichotomization.[4-6] Hanseeuw et al. identified three stages of amyloid deposition where cortical and striatal PET signal is classified separately, corresponding to distinct risk of cognitive decline.[4] Grothe and colleagues used [$^{18}$F]florbetapir PET data from cognitively unimpaired (CU) subjects and constructed a four-stage model, which successfully staged 410 (98%) subjects from the same cohort.[5] More recently, Mattson et al. used cerebrospinal fluid (CSF) to develop a longitudinally valid PET staging model, which could also be applied to an independent cohort using a different radiotracer.[6] However, the model construction could not be replicated and contrasting CSF/PET groups could have hampered the ability to identify more fine-grained stages.[6] As such, further evidence is required regarding the generalizability of single-tracer models to multi-tracer studies.

The present work aimed to construct and apply a data-driven and generalizable PET-based model for staging cortical amyloid burden. Cross-sectional and longitudinal relationships were assessed between amyloid stages and cognitive status, age, genetic risk, CSF and risk of future cognitive decline.

## Materials and Methods

### Cohorts

From six cohorts, all participants with available amyloid PET scans of sufficient quality for quantification were retrospectively included in this study (**Table 1**). In detail, [$^{18}$F]flutemetamol scans of 226 CU subjects from the Alzheimer's and Family (ALFA) cohort of the Barcelonaβeta Brain Research center,[7] 145 memory clinic patients from the Amsterdam Dementia Cohort (ADC) of the Amsterdam University Medical Center,[8] and 190 CU subjects from the Innovative Medicine Initiative European Medical Information Framework for AD (EMIF-AD)[9] project (http://www.emif.eu/emif-ad-2/) were included. [$^{18}$F]florbetaben scans of 353 memory clinic patients were obtained from the Alzheimer's biomarkers in daily practice (ABIDE) study.[10] Regional SUVr values were obtained from [$^{18}$F]florbetapir scans of 360 subjects and [$^{11}$C]PiB scans of 572 subjects from the Open Access Series of Imaging Studies (OASIS)-3 dataset (https://www.oasis-brains.org/).[11] Regional SUVr values from [$^{18}$F]florbetapir scans of 1179 subjects were included from the ADNI database (http://adni.loni.usc.edu/). The ADNI study was launched in 2003 as a public-private partnership, led by principal investigator Michael W. Weiner, MD. The primary goal of ADNI is to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD.

Across cohorts, 3027 subjects (1763 CU, 658 Impaired, 467 AD dementia, 111 non-AD dementia, and 28 with missing diagnosis at time of baseline PET) were included. In total, 4783 PET scans were included, as 1049 subjects had at least two amyloid PET scan sessions (OASIS and ADNI), with an average of 2.7 PET scans per subject (range 2 – 5). For each study, missing data was assumed to be missing at random.

**Standard Protocol Approvals, Registrations, and Patient Consents**

The protocol, patient information, consent form, and other relevant study documentation were approved by the Ethics Committees or Institutional Review Boards of each site before study initiation. The studies were performed in accordance with the Declaration of Helsinki and consistent with Good Clinical Practice. Before enrollment, all patients provided written informed consent.

## Image acquisition and processing

[$^{18}$F]Flutemetamol scans from the ALFA cohort consisted of four frames (4x5 minutes) acquired 90-110 minutes post-injection (p.i.). Images were checked for motion, and PET and accompanying structural T1-weighted MR images were warped into MNI space using SPM12. [$^{18}$F]Flutemetamol scans from ADC studies and [$^{18}$F]florbetaben scans from ABIDE were processed as described previously,[12] with static scans consisting of four frames (4x5 minutes) acquired 90-110 minutes p.i. In addition, [$^{18}$F]flutemetamol EMIF-AD scans were acquired using a dual-time-window protocol (0-30 minutes p.i., 60 minute break, 90-110 minutes p.i.).[13] Images were checked for motion, accompanying structural T1-weighted MR images were co-registered to PET using the Vinci software (Max Planck Institute for Neurological Research, Cologne, Germany) and then warped into MNI using SPM12. [$^{18}$F]Florbetapir (50-70 minutes p.i.) and [$^{11}$C]PiB (30-60 minutes p.i.) data from the OASIS platform were processed with FreeSurfer and the PET Unified Pipeline (https://github.com/ysu001/PUP).[14]

Acquisition and standardized preprocessing steps of MR and PET data in ADNI have been reported previously and are described in detail on the ADNI website (http://adni.loni.usc.edu/methods/pet-analysis-method/pet-analysis). Briefly, [$^{18}$F]florbetapir

PET scans consisted of four frames (4x5 minutes), acquired 50-70 minutes p.i. and were processed using FreeSurfer.

The Desikan-Killiany (DK) atlas was used to extract regional SUVr values normalized to cerebellar grey matter [15]. Small regions enclosed within high-signal areas were excluded due to risk for spill-in and misregistration errors (pericalcarine and banks superior temporal), and small/intermediate regions were merged (volume-weighted average) to reduce noise-related variability. The resulting 27 regions were anterior, isthmus, and posterior cingulate cortices, inferior frontal gyrus (pars opercularis, pars triangularis, and pars orbitalis), lateral- and medial orbitofrontal cortex, middle frontal gyrus (caudal and rostral middle frontal), superior frontal gyrus, frontal pole, inferior temporal gyrus, middle temporal gyrus, superior temporal gyrus (superior temporal and transverse temporal), fusiform gyrus, entorhinal cortex, parahippocampal gyrus, lingual gyrus, lateral occipital gyrus temporal pole, insula, inferior parietal gyrus, supramarginal gyrus, precuneus, superior parietal gyrus, precentral gyrus, postcentral gyrus, paracentral gyrus, and cuneus. Global SUVr was determined as a volume-weighted average of the 27 regions above.

## Model development

To construct a PET-based cortical amyloid staging model, regions were ranked as suggested in Grothe et al., assuming those most frequently abnormal in a CU population reflect early events in the pathological process, similar to how pathology studies are performed.[5 16 17]

Gaussian Mixture Modelling (mixtools and AdaptGauss packages from R statistical software program) was used to identify two distributions from global cortical SUVr values and based on each complete cohort, tracer and cohort specific cut-offs were determined based on the mean plus two standard deviations (SD) from the Gaussian (i.e. ‘normal’) distribution

corresponding to the amyloid-negative group (Data available from Dryad (Figure 1) https://doi.org/10.5061/dryad.7wm37pvp9).

Using an in-house algorithm for random optimal allocation subsampling, 400 baseline scans balanced across tracers (i.e. 100 per tracer) were randomly selected from the CU group for model construction to minimize demographic differences (i.e. APOE ε4 carriership, age, and sex) (**Table S1 available from Dryad doi:10.5061/dryad.7wm37pvp9**).

**Single-tracer regional ranking**

First, four different regional rankings were computed for each radiotracer (*N*=100 per tracer) in order to assess comparability. The cohort and tracer specific cut-offs were applied to each region and these were ranked according to the frequency of abnormality (% subjects for which a region's SUVr > cut-off) across each radiotracer group. Correlations between the four rankings ([$^{11}$C]PiB, [$^{18}$F]flutemetamol, [$^{18}$F]florbetaben, and [$^{18}$F]florbetapir) was assessed by Spearman ranking correlation analysis.

**Multi-tracer regional ranking**

Next, the complete selection of 400 baseline scans from CU subjects was pooled to create a multi-tracer regional ranking as per above. Agreement between the multi-tracer and each of the single-tracer rankings was also assessed.

**Staging model**

To overcome differences between single-tracer rankings identified in the previous steps, the model construction was performed based on the pooled multi-tracer data from 400 CU subjects.

In the construction step, we assessed the impact of strategies for cut-off and stage definition on the applicability of the model. Therefore, two additional data-driven cut-offs were defined as 1) the Bayes' optimal classifier (i.e. intersection between the two distributions) and 2) mean plus three SD from the Gaussian distribution corresponding to the amyloid-negative group (**Figure S1 available from Dryad doi:10.5061/dryad.7wm37pvp9**). Subsequently, these additional cut-offs were also used to rank regions based on the frequency of abnormality across the 400 CU subjects. Finally, four stages of amyloid burden were defined based on three independent strategies:

1) *Equal frequencies*: regions were grouped in four categories, each with equal frequencies[5]
2) *Equal number of regions*: regions were grouped so that each stage had an equal number of regions, and
3) *Equal volume*: regions were grouped such that each stage contained ¼ of the total cortical brain volume.

The steps described above (**Fig. 1**) resulted in nine candidate staging models (Data available from Dryad (Figure 2) https://doi.org/10.5061/dryad.7wm37pvp9). Further, in order to assess the impact of known confounding factors (e.g. spill-in, PVE and cerebral blood flow effects) on the regional ranking, the regional ranking was repeated for available PVC SUVr (EMIF-AD) and distribution volume ratio (DVR) values (EMIF-AD and OASIS) and reported in the supplementary results in Dryad.[18]

## Model application and assessment

All scans ($N$=4783) were classified into five stages (0-4) according to all nine models. A stage was attributed when more than 50% of the encompassed regions displayed SUVr > cut-off.[5] In addition, the model was hierarchical by design, i.e. higher stages are only achieved

once the staging conditions are also met for previous stages; otherwise the scan was considered unclassifiable.

The optimal model used for all subsequent analyses was defined as the one minimizing the number of unclassifiable subjects across cohorts. This performance indicator was chosen to assess the applicability of the model to a heterogeneous dataset.

## Statistical analysis

Statistical Package for the Social Sciences (SPSS) version 22 was used for all statistical analyses.

## Cross-sectional relationships

Chi-squared tests were used to assess the distribution of stages across dichotomous amyloid PET status (based on global SUVr), syndromic diagnosis, and APOE ε4 alleles. One-way ANCOVA was performed to assess the relationship between stages and age, Mini Mental State Examination (MMSE), and cerebrospinal fluid (CSF) measures, with cohort as a covariate and excluding subjects with a clinical diagnosis of non-AD dementia. A Bonferroni post-hoc test was used to assess between-stage differences, and effect-size was reported according to Cohen's guidelines ($\eta^2$=0.01 small, $\eta^2$=0.06 medium, and $\eta^2$=0.14 large), with significance set at $p$<0.05.

CSF measures of $A\beta_{40}$, $A\beta_{42}$, p-Tau, and t-Tau were available in a sub-set of subjects (**Table 1**). These were standardized to z-scores (tau CSF after log transformation due to skewedness) within each dataset based on the mean and SD of CU subjects to correct for cross-cohort differences in assays.

## Longitudinal PET

Based on first and last PET scans, subjects were considered *Stable* (no changes in stage or global PET status), *Progressors* (increase in stage or global PET status conversion from normal to abnormal) or *Reversed* (decrease in stage or global PET status conversion from abnormal to normal). Based on all PET scans, the monotonicity of the trajectory was determined. Differences in follow-up time between stage *Stable* and *Progressors* were tested using one-way ANOVA.

## Longitudinal Cognition

Longitudinal cognitive data was only available for subsets of ADNI and OASIS cohorts, and all non-AD dementia subjects were excluded for these analyses. The effect of 1) baseline stage and 2) global SUVr classification on cognitive decline were investigated using a linear mixed model (LMM) analysis with MMSE as outcome measure (corrected for age, sex, clinical diagnosis, and time between repeated measures), and Kaplan-Meier (KM) survival analysis with an MMSE score $\leq 25$ as event (excluding subjects with MMSE $\leq 25$ at baseline). In parallel, a Cox Regression analysis was performed to obtain stage- and global status-related Hazard Ratios. For both the LMM and KM survival analyses, pairwise comparison with Bonferroni correction was performed to assess between-stage differences.

## Data Availability

The data that support the findings of this study can be made available upon request with the study-specific principal investigator (i.e. ABIDE, EMIF-AD, ADC, ALFA) or are openly available (i.e. ADNI & OASIS open-source databases).

## Results

### Demographics

Single- and cross-cohort baseline demographics are shown in **Table 1**.

### Model development

Single-tracer regional rankings can be found in **Fig. 2A**. Correlation of the regional rankings was highest between [$^{18}$F]florbetaben and [$^{18}$F]flutemetamol ($\rho$=0.89), followed by [$^{11}$C]PiB and [$^{18}$F]flutemetamol ($\rho$=0.81), and then by [$^{11}$C]PiB and [$^{18}$F]florbetaben ($\rho$=0.79). Lower correlations ($\rho$ ranged from 0.58 to 0.63) were observed between [$^{18}$F]florbetapir and all other tracers (**Fig. 2B**). The multi-tracer regional ranking showed improved correlations ($\rho$ ranged from 0.69 to 0.94) with each single-tracer ranking compared to the correlations of the single-tracer rankings between themselves (**Fig. 2C**) and was used to construct the nine potential staging models (Data available from Dryad (Figure 2) https://doi.org/10.5061/dryad.7wm37pvp9).

The model defined based on equal frequencies using the mean plus 2SD cut-offs of the Gaussian distribution corresponding to the amyloid-negative group (**Table 1**) resulted in the smallest amount of unclassifiable subjects *(N*=23, 0.8%) and was therefore selected for all subsequent analyses. Of note, 2/23 unclassifiable subjects were represented in model construction. The results for the application of all nine models can be found in Data available from Dryad (Table 2) https://doi.org/10.5061/dryad.7wm37pvp9.

### Model application and assessment

According to the final 5-stage model (**Fig. 3**), the ranking of regional frequency of abnormality starts with the cingulate regions (~60% of subjects), followed by precuneus,

paracentral gyrus, lateral orbital cortex and insula (~40% of subjects), basal temporal, frontal, and more associative cortices (~20-30% of subjects), and ends with other temporal and occipital regions (~10% of subjects).

The model successfully classified virtually all 4783 scans, with only 48 (1.0%) unclassifiable scans. These scans were not from any particular cohort, tracer, diagnostic group, or age group (data not shown). Most unclassifiable scans ($N$=38, 79.2%) met the requirements for stages 1 and 3, but failed to fulfill stage 2. The remaining ($N$=10, 21.8%) displayed sufficient abnormal stage 2 regions without achieving stage 1.

## Cross-sectional relationships

At baseline, 1058 (35.0%) subjects were classified as stage 0, 694 (22.9%) as stage 1, 336 (11.1%) as stage 2, 319 (10.5%) as stage 3, and 597 (19.7%) subjects as stage 4. The distribution of stages per cohort is depicted in **Fig. 4A**. The association between stages and age was small ($N$=2866, $F$=30.55, $\eta^2$=0.04). Global SUVr classification identified positivity in 1.0%, 5.5%, 17.9%, 90.0%, and 100.0% in stage 0-4 subjects, respectively ($N$=3004, $\chi^2$=2439.67) (**Fig. 4B**). There was a significant association between stage and syndromic diagnosis ($N$=2977, $\chi^2$=679.00), with most (~60%) CU and non-AD dementia subjects classified as stage 0-1, ~40% the impaired subjects were classified as stage 0, while the remaining subjects were distributed across stages, and AD dementia subjects mostly classified as stages 3 and 4 (**Fig. 4C**). Baseline stage was related to number of APOE ε4 alleles ($N$=2887, $\chi^2$=400.16) (**Fig. 4D**). Higher stages were associated with lower baseline MMSE scores across diagnostic groups ($N$=2852, $F$=129.64, $\eta^2$=0.15), within CU subjects ($N$=1743, $F$=3.27, $\eta^2$=0.01), and within impaired subjects ($N$=654, $F$=14.12, $\eta^2$=0.08) with distinct effect-sizes. Increasing stages were associated with decreases in $A\beta_{42}$ levels ($N$=1462, $F$=235.35, $\eta^2$=0.39). More specifically, stage 3 and 4 subjects had lower $A\beta_{42}$ levels

compared to all other stages, stage 2 subjects had lower levels compared to stage 0/1, and there stage 1 subjects had lower levels compared to stage 0 (**Fig. 4E**). Finally, cortical amyloid stage was related to p-Tau levels, with stages 3 and 4 associated with higher z-scores compared to the other stages, but not differently from each other ($N$=1461, $F$=108.58, $\eta^2$=0.22) (**Fig. 4F**).

## Longitudinal PET

In total, 1049 subjects (ADNI: $N$=741 and OASIS: $N$=308) had longitudinal PET available, with a mean follow-up period of 3.97 years ($SD$=1.84, range=0.86 – 9.61 years). Based on stage at first and last PET scans, 65.5% of subjects were *Stable*, 24.9% subjects were *Progressors*, 8.1% *Reversed*, and 1.5% had an unclassifiable scan at follow-up. In comparison, based on global amyloid PET status, 89.0% of subjects were *Stable*, 6.7% subjects were *Progressors*, and 3.9% *Reversed*. Importantly, progression and reversal rates were relatively consistent across stages of amyloid burden(**Table 2**). Considering all scans, 14.8% of subjects displayed a non-monotonic behavior based on the staging model, compared to 6.8% based on global PET.

Further, stage *Progressors* ($M_{years}$=4.61, $SD_{years}$=1.92, 95% CI: 4.38 to 4.85) had a significantly longer follow-up period compared to stage *Stable* ($M_{years}$=3.77, $SD_{years}$=1.75, 95% CI: 3.64 to 3.91) subjects ($N$=941, $F$=40.68, $\eta^2$=0.04). Global *Progressors* ($M_{years}$=4.45, $SD_{years}$=1.98, 95% CI: 3.98 to 4.92) also had a significantly longer follow-up period compared to global *Stable* ($M_{years}$=3.94, $SD_{years}$=1.82, 95% CI: 3.82 to 4.06) subjects ($N$=1008, $F$=5.00, $\eta^2$=0.01).

## Longitudinal cognition

Linear mixed model analysis showed that baseline stage, corrected for age, sex, clinical diagnosis, and time between follow-up visits (repeated measures), predicts longitudinal MMSE decline (ADNI: $N$=867, $F$=67.37; OASIS: ($N$=475, $F$=9.12). More specifically in ADNI, post-hoc analyses showed that stage 1-4 declined faster compared to 0 (95% CI: -1.13 to 0.42, -1.49 to -0.44, -2.25 to 1.47 and -2.74 to -2.01 respectively), those in stage 4 declined faster than stage 1/2 (95% CI: -2.05 to -1.14, -2.00 to -0.81, respectively) subjects, and those in stage 3 declined faster than stage 1/2 subjects (95% CI: -1.56 to -0.60, -1.50 to -0.28, respectively ). In OASIS, stage 4 subjects declined faster than stage 0/1/2 subjects (95% CI: -1.13 to -0.45, -1.09 to -0.39 and 1.35 to -0.60 respectively). No other significant differences between stages were observed (**Fig. 5A&B**).

Furthermore, in ADNI, higher baseline stage predicted faster decline towards an MMSE≤25 ($N$=787) (**Fig. 5C**). Post-hoc analyses showed a number of stage-wise differences in time to reach the event (**Table 2).** This effect was mainly driven by the impaired subgroup (**Figure S7 available from Dryad doi:10.5061/dryad.7wm37pvp9**). In comparison, a global SUVr dichotomization showed amyloid-positive subjects ($N$ = 242) reached the event faster than amyloid-negative subjects ($N$=552).) (**Table 2**).

In OASIS, a faster decline towards an MMSE≤25 was observed only for subjects with baseline stages 3 and 4 ($N$=469) (**Fig. 5D**). In comparison, global SUVr dichotomization showed amyloid-positive subjects ($N$=122) reached the event faster compared to amyloid-negative ones ($N$=351) . Mean estimated time to event and hazard ratio's for all survival analyses can be found in **Table 2**.

## Discussion

Our pooled-analysis of >3000 subjects demonstrates that a PET-based multi-tracer model can reliably assign cortical amyloid stage across six cohorts. More specifically, the 5-stage model classified >99% of 4783 scans from a large heterogeneous sample, across diagnostic groups, and scanned with four different amyloid PET radiotracers. The model provides information on amyloid burden beyond global dichotomized classification and shows a clear relationship with genetic risk, CSF values, and clinical variables, thereby supporting its value.

The excellent cross-cohort performance of the presented model can be attributed to its multi-tracer aspect. While previous groups achieved high applicability either within a single-cohort [4 5] or between two studies,[6] neither fully replicated the model in an independent sample. In this work, we were able to perform a head-to-head comparison of the same methodology applied to a balanced sample of subjects across four different tracers, which showed marked differences in regional rankings (**Fig. 2A**) especially for [$^{18}$F]florbetapir. This particular tracer effect could be due to the choice of target and reference region in this work, which has been previously deemed suboptimal for ADNI scans.[19] However, a consistent ROI definition across cohorts was prioritized in this work to allow the pooling of the data-sets. This finding suggests the direct translation of a single-tracer model to multi-tracer studies could bring challenges. As the field moves towards large-scale integrative studies, the need for harmonizing results across radiotracers and cohorts greatly increases, as can already be appreciated by initiatives such as the Centiloid Project.[20]

The observed regional ordering of the 5-stage model is in somewhat disagreement to established neuropathological literature.[16 17] The main difference would be the early appearance of the cingulate in our model, while it is reported only after neocortical involvement by Thal,[17] and after basal portions of the frontal, temporal and occipital lobe by

Braak.[16] In addition, Braak reports an early involvement of basal temporal regions, a pattern which is not consistently observed in PET literature. Except for the work of Grothe and colleagues,[5] several PET studies fail to identify early temporal involvement despite distinct methodologies, and instead also point to the cingulate and precuneus as early regions, including the most recent and extensive staging work to date.[6 21 22] This consistent discrepancy between neuropathology and PET-only studies has been described previously,[23] and is likely to be influenced by signal distortion present in PET imaging. Due to the proximity of medial regions to white matter and the additional grey matter signal spill-in from the contralateral hemisphere, medial regions such as those in stage 1 and 2 in this work are more frequently classified as abnormal in PET imaging compared to lateral counterparts, even when levels of pathology are comparable..[24]

While the discrepancy with pathological studies might be due to sample size, resolution limitations and differences in technique such as the intrinsic signal distortion of PET imaging, remaining differences across PET studies could also be explained by methodological choices including the use of partial volume correction (PVC). However, our results from PVE-corrected SUVr values did not substantially differ from the main model (Data available from Dryad (Figure 3) https://doi.org/10.5061/dryad.7wm37pvp9), suggesting that spill-in from white matter (of special concern for $^{18}$F-labeled radiotracers [25]) has only a limited effect on the regional ordering. In addition, the similar ordering observed with DVR indicates that the overall impact of SUVr confounders on this method is limited (Data available from Dryad (Figure 4) https://doi.org/10.5061/dryad.7wm37pvp9).[26 27]

The observed relationships between stages and several demographic, clinical, and CSF measures support the validity of the staging model. In particular, the lower CSF $A\beta_{42}$ values observed in stage 2 compared to stage 0/1 subjects suggests that regional amyloid PET may be more sensitive for early pathological detection compared to standard dichotomization of

global amyloid burden. In addition, the comparable levels of CSF $A\beta_{42}$ between stage 3 and stage 4 subjects indicate that an amyloid-PET staging system can differentiate pathological burden even after a plateau is reached in CSF levels. Furthermore, the cortical amyloid staging model seems valuable for prognosis as baseline amyloid stage relates to distinct risk profiles of subsequent cognitive decline as measured with MMSE (**Fig. 5**). In particular, distinct patterns can be observed per cohort, where a step-wise effect of stage on cognitive decline is apparent mostly in the impaired subjects of ADNI, while only later stages related to faster decline in CU individuals. Further, a staging system captures differences in cognitive decline between stage 3 and 4 subjects, where the latter decline faster than a globally positive group, indicating the extent of the 'global' burden provides further prognostic information. Importantly, these results demonstrate the risk related to amyloid pathology alone, and in order to fully assess an individual's risk profile, it is necessary to incorporate additional biomarkers.[2]

In view of these results, the main added value of this cortical amyloid staging model compared to current approaches in PET quantification is twofold. First, it is able to detect regional abnormality before global positivity is assigned, as more than 90% of stage 1, approximately 80% of stage 2 and 10% of stage 3 subjects are classified as negative based on global SUVr. This points to the limitations of dichotomizing amyloid PET and provides further evidence to the possibility for early pathological identification with amyloid PET imaging. Second, even after global positivity is assigned the model allows for a detailed assessment of the extent of amyloid burden and corresponding risk of cognitive decline (**Fig. 5, stage 3 vs. stage 4 subjects**). Therefore, independent of global PET status, the stage classifications can help identify individuals with Alzheimer's pathological changes who have greater risk of amyloid-related long-term cognitive decline, a relevant population for secondary prevention trials (Ritchie 2016). In addition, the staging's intrinsic grouping of

individuals with similar levels of amyloid pathology ensures treatment arms are appropriately balanced [28] and may be relevant for modeling expected treatment effects on cognition.[29]

Importantly, the staging model better captures the accumulation process in comparison to a simple conversion from negative to positive. In this study, the model identified 24.7% of subjects with longitudinal PET as *Progressors* compared to 6.7% of converters based on global SUVr. On the other hand, the staging model relies on classifying regional SUVr, which might be more sensitive than composite or global signal to local blood flow changes, segmentation and co-registration errors, all common challenges in longitudinal PET analyses.[19] This can be appreciated by the model's identification of 14.8% non-monotonic trajectories, of which approximately half are also observed using dichotomous global SUVr. This suggests that creating a composite ROI per stage, as has been done by Mattsson and colleagues,[6] could reduce the variability observed in our results, although likely not below the effect already present with global SUVr. Since the choice of reference region could also have affected these results, future work could assess the performance of different regions in order to improve longitudinal plausibility.

One potential solution to minimize the impact of regional sensitivity in a staging system while maintaining a more fine-grained description of the accumulation process would be to create stages based on global SUVr alone. In order for such a model to be applicable across tracers, a harmonization tool such as the Centiloid scale would be required, which would avoid single-tracer cut-offs.[20] Conversion to Centiloid was out of the scope of this work, and in fact, the staging model can be considered an alternative and simple method for cross-tracer analysis of amyloid PET images. Nonetheless, a supplementary analysis was performed where ADNI data was divided in quintiles to create 5 stages based on global SUVr. Although such a global system showed a low agreement (Kappa = 0.32) to the stages defined in this work (Data available from Dryad (Table 3, Figure 8)

https://doi.org/10.5061/dryad.7wm37pvp9), in general a similar pattern of increased risk per level of amyloid burden was observed. However, hazard-ratios are consistently higher for the 5-stage model compared to a quintiles-based approach, and only the 5-stage model associates the first stage with increased risk compared to amyloid-negative subjects. These results suggest that while global analyses are able to provide risk stratification, the use of regional information is particularly relevant for the earliest changes in amyloid burden and better distinguishes risk profiles across the spectrum.

In this work, a total of nine model definitions were assessed without an available standard of truth to determine the underlying regional levels of amyloid burden. While the results add to the evidence of a temporal evolution of the topography of amyloid pathology, this methodology is not sufficient to shed light into the potential underlying mechanism. In fact, our results reflect the ordering of regions as detected by PET, which are likely to be affected by the distortions in the PET signal between medial and lateral regions.[24] Therefore, this staging model should be seen as a tool, applicable across tracers and cohorts, with which one is able to assess a subject's level of brain amyloidosis and its associated risks in terms of cognitive decline. As such, the model is optimized to classify as many subjects as possible, and, while defining additional stages could provide a more detailed picture of the pathological process, it would likely increase the number of unclassifiable scans due to between-subject variability. In addition, when utilizing this model within a particular application, image processing and cut-off definitions can be optimized accordingly, e.g. using more lenient cut-offs to detect risk of subsequent amyloid accumulation, or more conservative and neuropathology-grounded cut-offs for trial inclusion.

Main limitations of this study are the lack of pathological confirmation, the use of a global cut-off for regional SUVr values, the omission of sub-cortical structures, and the atlas-based definition of regions of interest. First, no pathological evaluation has been made at the level of

detail available from PET scans used in this analysis, thereby limiting the availability of a standard of truth to this work. However, current studies find similar relationships between global and regional SUVr and post-mortem data.[30] Second, the use of a global SUVr cut-off to define regional abnormality has not been validated, and such a validation would require a reference method (regional visual assessment or regional histopathological data), with its application across datasets potentially posing major challenges. In addition, recent work suggests that while regional cut-offs would differ from global ones, the final regional ordering would be highly consistent with results from global cut-off.[31-33]Third, including sub-cortical structures in the model would likely require region-specific cut-offs, as indicated by previous work.[4] However, our model could be combined with previous work by Hanseeuw et al. for a complete cortical and subcortical staging of amyloid deposition. Finally, atlas-based approaches could be limiting the power to detect signal increases that do not follow anatomical boundaries, and a voxel-wise model could be considered to extend the current work.[34]

## Conclusion

A multi-tracer 5-stage cortical amyloid staging model developed solely based on amyloid PET scans from 400 CU subjects is able to classify the level of amyloid burden in >3000 subjects across cohorts and radiotracers. Amyloid stage is strongly related to demographic and clinical measures, can detect pre-global amyloid burden and distinguish cognitive decline risk profiles within amyloid-positive subjects. The model is easily applicable to any cohort after a cohort-specific global SUVr cut-off is defined and the DK atlas is applied, and it has the potential to improve targeted subject trial inclusion and support routine clinical work. In relation to identification of pre-global amyloid pathology, further comparison of this model with other established methods is warranted to minimize false positives.

## Acknowledgements

The authors would like to acknowledge all data donors and collaborators (ADNI, EMIF-AD, OASIS-3, ABIDE, ALFA, and ADC). This research would not have been possible without the time and dedication of every participant included in these studies.

## Appendix 1: Authors

| Name | Location | Contribution |
|---|---|---|
| Lyduine E. Collij, MSc | Amsterdam UMC, VUmc, Amsterdam, NL | Literature search, design, data collection, analysis, interpretation, drafted manuscript |
| Fiona Heeman, MSc | Amsterdam UMC, VUmc, Amsterdam, NL | Literature search, design, analysis, interpretation, drafted manuscript |
| Gemma Salvadó, MSc | Barcelonaβeta Research Center, Barcelona, ES | Literature search, design, data collection, analysis, interpretation, drafted manuscript |
| Silvia Ingala, MD | Amsterdam UMC, VUmc, Amsterdam, NL | Data collection, analysis, revised manuscript |
| Daniele Altomare, MSc | University (Hospital) of Geneva, Geneva, Switzerland | Design, interpretation, revised manuscript |
| Arno de Wilde, MD | Amsterdam UMC, VUmc, Amsterdam, NL | Data collection, analysis, revised manuscript |
| Elles Konijnenberg, MD | Amsterdam UMC, VUmc, Amsterdam, NL | Data collection, interpretation, revised manuscript |
| Marieke van Buchem, BSc | Amsterdam UMC, VUmc, Amsterdam, NL | Data collection, analysis, revised manuscript |
| Maqsood Yaqub, PhD | Amsterdam UMC, VUmc, Amsterdam, NL | Design, data collection, analysis, interpretation, revised manuscript |
| Pawel Markiewicz, PhD | University College London, London, UK | Design, interpretation, revised manuscript |
| Sandeep S.V. Golla, PhD | Amsterdam UMC, VUmc, Amsterdam, NL | Analysis, interpretation, revised manuscript |
| Viktor Wottschel, PhD | Amsterdam UMC, VUmc, Amsterdam, NL | Design, analysis, interpretation, revised manuscript |
| Alle Meije Wink, PhD | Amsterdam UMC, VUmc, Amsterdam, NL | Design, analysis, interpretation, revised manuscript |
| Pieter Jelle Visser, PhD | Amsterdam UMC, VUmc, Amsterdam, NL | Design, interpretation, revised manuscript |
| Charlotte E. Teunissen, PhD | Amsterdam UMC, VUmc, Amsterdam, NL | Data collection, analysis, revised manuscript |
| Adriaan A. Lammertsma, PhD | Amsterdam UMC, VUmc, Amsterdam, NL | Design, interpretation, revised manuscript |

| | | |
|---|---|---|
| Philip Scheltens, PhD | Amsterdam UMC, VUmc, Amsterdam, NL | Design, data collection, interpretation, revised manuscript |
| Wiesje M. van der Flier, PhD | Amsterdam UMC, VUmc, Amsterdam, NL | Design, data collection, interpretation, revised manuscript |
| Ronald Boellaard, PhD | Amsterdam UMC, VUmc, Amsterdam, NL | Design, data analysis, interpretation, revised manuscript |
| Bart N.M. van Berckel, PhD | Amsterdam UMC, VUmc, Amsterdam, NL | Design, data collection, interpretation, revised manuscript |
| José Luis Molinuevo, PhD | Barcelonaβeta Research Center, Barcelona, ES; | Design, data collection, interpretation, revised manuscript |
| Juan Domingo Gispert, PhD | Barcelonaβeta Research Center, Barcelona, ES; | Design, data collection, interpretation, revised manuscript |
| Mark E. Schmidt, MD | Janssen Pharmaceutica NV, Beerse, BE | Design, interpretation, revised manuscript |
| Frederik Barkhof, PhD | Amsterdam UMC, VUmc, Amsterdam, NL; University College London, London, UK | Literature search, design, data collection, interpretation, revised manuscript |
| Isadora Lopes Alves, PhD | Amsterdam UMC, VUmc, Amsterdam, NL | Literature search, design, analysis, interpretation, drafted manuscript |

## References


1. Morris E, Chalkidou A, Hammers A, et al. Diagnostic accuracy of (18)F amyloid PET tracers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging* 2016;43(2):374-85. doi: 10.1007/s00259-015-3228-x
2. Jack CR, Jr., Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* 2018;14(4):535-62. doi: 10.1016/j.jalz.2018.02.018
3. Thal DR, Beach TG, Zanette M, et al. [(18)F]flutemetamol amyloid positron emission tomography in preclinical and symptomatic Alzheimer's disease: specific detection of advanced phases of amyloid-beta pathology. *Alzheimers Dement* 2015;11(8):975-85. doi: 10.1016/j.jalz.2015.05.018
4. Hanseeuw BJ, Betensky RA, Mormino EC, et al. PET staging of amyloidosis using striatum. *Alzheimers Dement* 2018 doi: 10.1016/j.jalz.2018.04.011
5. Grothe MJ, Barthel H, Sepulcre J, et al. In vivo staging of regional amyloid deposition. *Neurology* 2017;89(20):2031-38. doi: 10.1212/WNL.0000000000004643
6. Mattsson N, Palmqvist S, Stomrud E, et al. Staging beta-Amyloid Pathology With Amyloid Positron Emission Tomography. *JAMA Neurol* 2019 doi: 10.1001/jamaneurol.2019.2214 [published Online First: 2019/07/18]
7. Molinuevo JL, Gramunt N, Gispert JD, et al. The ALFA project: A research platform to identify early pathophysiological features of Alzheimer's disease. *Alzheimers Dement (N Y)* 2016;2(2):82-92. doi: 10.1016/j.trci.2016.02.003
8. van der Flier WM, Scheltens P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. *J Alzheimers Dis* 2018;62(3):1091-111. doi: 10.3233/JAD-170850
9. Konijnenberg E, Carter SF, Ten Kate M, et al. The EMIF-AD PreclinAD study: study design and baseline cohort overview. *Alzheimers Res Ther* 2018;10(1):75. doi: 10.1186/s13195-018-0406-7
10. de Wilde A, van Maurik IS, Kunneman M, et al. Alzheimer's biomarkers in daily practice (ABIDE) project: Rationale and design. *Alzheimers Dement (Amst)* 2017;6:143-51. doi: 10.1016/j.dadm.2017.01.003
11. Marcus DS, Fotenos AF, Csernansky JG, et al. Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. *J Cogn Neurosci* 2010;22(12):2677-84. doi: 10.1162/jocn.2009.21407
12. Collij L, Konijnenberg E, Reimand J, et al. Assessing Amyloid Pathology in Cognitively Normal Subjects using [(18)F]Flutemetamol PET: Comparing Visual Reads and Quantitative Methods. *J Nucl Med* 2018 doi: 10.2967/jnumed.118.211532
13. Heeman F, Yaqub M, Lopes Alves I, et al. Optimized dual-time-window protocols for quantitative [(18)F]flutemetamol and [(18)F]florbetaben PET studies. *EJNMMI Res* 2019;9(1):32. doi: 10.1186/s13550-019-0499-4
14. Su Y, D'Angelo GM, Vlassenko AG, et al. Quantitative analysis of PiB-PET with FreeSurfer ROIs. *PLoS One* 2013;8(11):e73377. doi: 10.1371/journal.pone.0073377
15. Desikan RS, Segonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 2006;31(3):968-80. doi: 10.1016/j.neuroimage.2006.01.021
16. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* 1991;82(4):239-59.
17. Thal DR, Rub U, Orantes M, et al. Phases of A beta-deposition in the human brain and its relevance for the development of AD. *Neurology* 2002;58(12):1791-800.
18. Alves IL. Supplementary Material for "Multi-tracer model for staging cortical amyloid deposition using PET imaging", v3, Dryad, Dataset, . 2020 doi: https://doi.org/10.5061/dryad.7wm37pvp9

19. Landau SM, Fero A, Baker SL, et al. Measurement of longitudinal beta-amyloid change with 18F-florbetapir PET and standardized uptake value ratios. *J Nucl Med* 2015;56(4):567-74. doi: 10.2967/jnumed.114.148981
20. Klunk WE, Koeppe RA, Price JC, et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimers Dement* 2015;11(1):1-15 e1-4. doi: 10.1016/j.jalz.2014.07.003
21. Palmqvist S, Scholl M, Strandberg O, et al. Earliest accumulation of beta-amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nat Commun* 2017;8(1):1214. doi: 10.1038/s41467-017-01150-x
22. Farrell ME, Chen X, Rundle MM, et al. Regional amyloid accumulation and cognitive decline in initially amyloid-negative adults. *Neurology* 2018 doi: 10.1212/WNL.0000000000006469
23. Fantoni E, Collij L, Alves IL, et al. The spatial-temporal ordering of amyloid pathology and opportunities for PET imaging. *J Nucl Med* 2019 doi: 10.2967/jnumed.119.235879 [published Online First: 2019/12/15]
24. Smith A, Buckley C. [18F]flutemetamol PET image representation of Ab pathology; differences between lateral and medial image intensity for equivalent levels of pathology. 10th Human Amyloid Imaging. Miami, FL, USA, 2016.
25. Catafau AM, Bullich S. Amyloid PET imaging: applications beyond Alzheimer's disease. *Clin Transl Imaging* 2015;3(1):39-55. doi: 10.1007/s40336-014-0098-3
26. van Berckel BN, Ossenkoppele R, Tolboom N, et al. Longitudinal amyloid imaging using 11C-PiB: methodologic considerations. *J Nucl Med* 2013;54(9):1570-6. doi: 10.2967/jnumed.112.113654
27. Schmidt ME, Chiao P, Klein G, et al. The influence of biological and technical factors on quantitative analysis of amyloid PET: Points to consider and recommendations for controlling variability in longitudinal data. *Alzheimers Dement* 2015;11(9):1050-68. doi: 10.1016/j.jalz.2014.09.004
28. Guo T, Dukart J, Brendel M, et al. Rate of β-amyloid accumulation varies with baseline amyloid burden: Implications for anti-amyloid drug trials. *Alzheimer's & Dementia* 2018 doi: 10.1016/j.jalz.2018.05.013
29. Liu E, Schmidt ME, Margolin R, et al. Amyloid-beta 11C-PiB-PET imaging results from 2 randomized bapineuzumab phase 3 AD trials. *Neurology* 2015;85(8):692-700. doi: 10.1212/WNL.0000000000001877
30. Seo SW, Ayakta N, Grinberg LT, et al. Regional correlations between [(11)C]PIB PET and post-mortem burden of amyloid-beta pathology in a diverse neuropathological cohort. *Neuroimage Clin* 2017;13:130-37. doi: 10.1016/j.nicl.2016.11.008
31. Lopes-Alves I, Collij L, Heeman F, et al. EVENT-BASED MODELING OF THE TEMPORAL ORDERING OF REGIONAL β-AMYLOID DEPOSITION IN THE BRAIN. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 2018;14(7):P887-P88. doi: 10.1016/j.jalz.2018.06.1137
32. Salvadó G, Collij L, Niñerola-Baizán A, et al. VOXEL-BASED AMYLOID PET STAGING FOR THE WHOLE ALZHEIMER'S DISEASE CONTINUUM. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 2019;15(7):P425-P26.
33. Jelistratova I, Bugla K, Teipel S, et al. Reconsidering in-vivo models of regional amyloid pathology progression: region-specific thresholds and longitudinal data2019.
34. Bilgel M, Prince JL, Wong DF, et al. A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. *Neuroimage* 2016;134:658-70. doi: 10.1016/j.neuroimage.2016.04.001

**Table 1. Baseline Demographics Cohorts**

| **Tracer** | | **[$^{18}$F]Flutemetamol** | | | | | | **[$^{18}$F]Florbetaben** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Cohort** | **ALL (*N* = 3027)** | **ALFA (*N* = 226)** | **EMIF-AD (*N* = 190)** | **ADC (*N*= 145)** | | | | **ABIDE (*N* = 353)** | | | |
| **Diagnostic group** | **N/A** | **CU** | **CU** | **CU (*N* = 3)** | **Impaired (*N* = 10)** | **AD Dementia (*N* = 84)** | **Non AD Dementia (*N* = 48)** | **CU (*N* = 128)** | **Impaired (*N* = 66)** | **AD Dementia (*N* = 106)** | **Non AD Dementia (*N* = 53)** |
| **Age (SD)** | 68.71 (9.08) | 61.17 (4.83) | 70.44 (7.55) | 59.76 (1.78) | 62.80 (7.65) | 62.47 (5.41) | 61.77 (5.54) | 60.70 (7.97) | 66.13 (7.09) | 66.69 (7.47) | 66.84 (7.16) |
| **Gender (F)** | 1551 (51.30%) | 142 (62.8%) | 112 (58.9%) | 0 (0.0%) | 3 (30.0%) | 45 (53.6%) | 18 (37.5%) | 53 (41.4%) | 24 (36.4%) | 52 (49.1%) | 15 (28.3%) |
| **MMSE (SD)** | 27.61 (3.14) | 29.01 (2.25) | 28.99 (1.14) | 25.00 (3.46) | 25.80 (1.75) | 22.78 (3.10) | 23.96 (3.72) | 27.76 (2.42) | 26.94 (2.01) | 22.41 (4.24) | 24.45 (3.40) |
| **APOE ε4 carriership +[#]** | 1230 (42.30%) | 98 (43.3%) | 62 (33.3%) | 1 (33.3%) | 5 (50.0%) | 63 (75.0%) | 18 (37.5%) | 50 (39.4%) | 31 (47.7%) | 63 (59.4%) | 21 (42.0%) |
| **Global PET +[$]** | 1007 (33.30%) | 31 (13.7%) | 36 (18.9%) | 0 (0.0%) | 2 (20.0%) | 81 (96.4%) | 10 (20.8%) | 25 (19.5%) | 26 (39.4%) | 71 (67.0%) | 16 (30.2%) |
| **PET cut-off*** | N/A | 1.34 | 1.27 | 1.22 | | | | 1.28 | | | |
| **CSF $A\beta_{42}$ (SD)** | N/A | 1347.12 (360.98) | 892.28 (317.98) | 1277.00 (131.52) | 918.80 (318.69) | 539.75 (92.56) | 939.48 (215.34) | 1081.74 (283.94) | 906.45 (319.25) | 681.61 (276.33) | 892.94 (295.90) |
| **CSF T-Tau (SD)** | N/A | 204.44 (77.13) | N/A | 192.67 (180.94) | 303.71 (397.04) | 569.99 (339.82) | 357.38 (213.89) | 366.23 (286.48) | 476.39 (274.50) | 651.77 (368.39) | 387.72 (178.33) |
| **CSF P-Tau (SD)** | N/A | 16.78 (8.25) | 76.34 (44.38) | 37.00 (35.50) | 58.40 (46.83) | 77.09 (27.36) | 44.05 (18.55) | 53.92 (29.84) | 65.74 (27.82) | 79.04 (35.37) | 53.47 (17.88) |
| **CSF Essay** | N/A | Elecsys | Adx Euroimmune | Innotest | | | | Innotest | | | |

**Table 1. Baseline Demographics Cohorts (continued)**

| **Tracer** | **[$^{11}$C]PiB** | | | | **[$^{18}$F]Florbetapir** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Cohort** | **OASIS (*N* = 572)** | | | | **OASIS (*N* = 360)** | | | | **ADNI (*N* = 1181)** | | | |
| **Diagnostic group** | **CU (*N* = 482)** | **Impaired (*N* = 32)** | **AD Dementia (*N* = 49)** | **Non AD Dementia (*N* = 9)** | **CU (*N* = 304)** | **Impaired (*N* = 25)** | **AD Dementia (*N* = 30)** | **Non AD Dementia (*N* = 1)** | **Missing (*N* = 28)** | **CU (*N* = 430)** | **Impaired (*N* = 525)** | **AD Dementia (*N* = 198)** |
| **Age (SD)** | 64.63 (9.32) | 70.44 (8.35) | 75.43 (7.72) | 66.81 (0.35) | 66.78 (8.53) | 70.92 (6.30) | 73.70 (6.86) | 67.61 | 72.72 (9.84) | 73.97 (6.78) | 72.87 (7.96) | 75.02 (7.75) |
| **Gender (F)** | 292 (60.6%) | 18 (56.3%) | 20 (40.8%) | 4 (44.4%) | 163 (53.6%) | 17 (68.0%) | 18 (60.0%) | 1 (100%) | 14 (51.9%) | 233 (54.2%) | 222 (42.3%) | 82 (41.4%) |
| **MMSE (SD)** | 29.13 (1.15) | 27.97 (1.94) | 23.60 (3.57) | 24.67 (5.66) | 29.04 (1.25) | 28.56 (1.53) | 24.63 (4.00) | 19.00 | 26.00 (2.55) | 29.06 (1.19) | 28.02 (1.78) | 22.49 (3.28) |
| **APOE ε4 carriership +**[#] | 161 (33.5%) | 15 (46.9%) | 31 (63.3%) | 4 (44.4%) | 101 (34.7%) | 6 (30.4%) | 23 (76.7%) | 1 (100.0%) | 8 (36.4%) | 102 (27.9%) | 237 (46.2%) | 129 (66.8%) |
| **Global PET +**[$] | 114 (23.7%) | 15 (46.9%) | 44 (89.8%) | 6 (66.7%) | 65 (21.4%) | 12 (48.0%) | 25 (83.3%) | 1 (100.0%) | 11 (39.43%) | 94 (21.9%) | 192 (36.6%) | 131 (66.2%) |
| **PET cut-off*** | 1.16 | | | | 1.28 | | | | 1.41 | | | |
| **CSF $A\beta_{42}$ (SD)** | N/A | | | | N/A | | | | N/A | 1246.92 (433.15) | 1016.95 (431.53) | 696.12 (338.48) |
| **CSF T-Tau (SD)** | N/A | | | | N/A | | | | N/A | 241.58 (91.21) | 277.50 (128.44) | 374.06 (156.31) |
| **CSF P-Tau (SD)** | N/A | | | | N/A | | | | N/A | 22.13 (9.35) | 26.60 (14.32) | 36.71 (16.37) |
| **CSF Essay** | N/A | | | | N/A | | | | Elecsys | | | |

CU = cognitively unimpaired subjects, which includes both controls and subjective cognitive decliners.
Impaired subjects either had a clinical diagnosis of MCI or a Clinical Dementia Rating of 0·5 (in the absence of a clinical diagnosis)

AD and non-AD dementia subjects all had a clinical diagnosis of AD or non-AD dementia, respectively.
[#] Subject carries at least 1 APOE ε4 allele
[$] Based on global cortical amyloid burden
* Cut-off derived from Gaussian Mixture Modelling and used for final model

**Table 2. Longitudinal Results**

| **Longitudinal PET** | | | | | | |
|---|---|---|---|---|---|---|
| | | | **Baseline Stage** | | | |
| | **Stage 0** | **Stage 1** | **Stage 2** | **Stage 3** | **Stage 4** | **Total** |
| **Stable** | 409 (78.8%) | 61 (36.1%) | 35 (39.3%) | 43 (39.4%) | 134 (86.5%) | 682 (65.5%) |
| **Progressed** | 107 (20.6%) | 68 (40.2%) | 35 (39.3%) | 49 (45.0%) | - | 259 (24.9%) |
| **Reversed** | - | 33 (19.5%) | 18 (20.2%) | 15 (13.8%) | 18 (11.6%) | 84 (8.1%) |
| **Unclassifiable at follow-up** | 3 (0.6%) | 7 (4.1%) | 1 (1.1%) | 2 (1.8%) | 3 (1.9%) | 16 (1.5%) |
| **Total** | 519 (49.9%) | 169 (16.2%) | 89 (8.5%) | 109 (10.5%) | 155 (14.9%) | 1041[Ψ] |

| **Longitudinal Cognition** | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **ADNI cohort** | | | | | | **OASIS cohort** | | | | | |
| | | **Kaplan Meier** | | | | **Cox Regression** | | | **Kaplan Meier** | | | **Cox Regression** | |
| **PET status** | **# Subjects** | **% subjects who reached the event** | **Estimated time to event (years) Mean (SE)** | **95% Confidence Interval** | | **Hazard Ratio's** | **95% Confidence Interval** | **# Subjects** | **% subjects who reached the event** | **Estimated time to event (years) Mean (SE)** | **95% Confidence Interval** | **Hazard Ratio's** | **95% Confidence Interval** |
| **Global PET -** | 552 | 19.5 | 6.71 (0.12) | 6.47 | 6.95 | Reference group | | 351 | 6.0 | 11.73 (0.16) | 11.43 | 12.01 | Reference group |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Global PET +** | 242 | 54.9 | 4.17 (0.18) | 3.81 | 4.52 | 3.45* | 2.63 | 4.53 | 122 | 23.1 | 9.29 (0.48) | 8.35 | 10.24 | 4.42* | 2.50 | 7.83 |
| **No amyloid** | 462 | 24.9 | 6.99 (0.12) | 6.76 | 7.23 | Reference group | | | 152 | 7.2 | 10.79 (0.21) | 10.37 | 11.20 | Reference group | | |
| **Stage 1** | 97 | 25.8 | 5.65 (0.28) | 5.10 | 6.20 | 2.00# | 1.26 | 3.16 | 117 | 7.7 | 11.53 (0.30) | 10.93 | 12.12 | 1.10 | 0.46 | 2.65 |
| **Stage 2** | 41 | 41.5 | 4.65 (0.39) | 3.88 | 5.42 | 3.53#0,$1 | 2.07 | 6.00 | 86 | 1.2 | 11.54 (0.14) | 11.27 | 11.82$ | 0.17$0,1 | 0.02 | 1.33 |
| **Stage 3** | 82 | 45.1 | 4.10 (0.29) | 3.52 | 4.68 | 4.55#0,1 | 3.04 | 6.81 | 36 | 16.7 | 9.30 (0.54) | 8.24 | 10.35 | 2.32#2 | 0.86 | 6.26 |
| **Stage 4** | 105 | 59.0 | 3.30 (0.24) | 2.82 | 3.78 | 9.91#0-3 | 4.86 | 9.81 | 78 | 27.9 | 8.53 (0.58) | 7.40 | 9.67 | 4.80#0,1,2 | 2.31 | 9.98 |

Ψ8 scans were unclassifiable at baseline and therefore not included in these analyses.
*$p < 0.001$ compared to Global PET –
#Significantly higher risk compared to stage number(s) in superscript ($p < 0.05$)
$Higher risk compared to stage number(s) in superscript at trend level ($p = 0.05$-$0.10$)

**Figure 1. Construction of cortical amyloid staging models Schematic representation of the work-flow for constructing nine staging models.** First, three different data-driven cut-offs per cohort were determined using Gaussian Mixture Modelling and each entire cohort. Then, 400 out of all CU subjects were selected for model construction using a iterative algorithm, following by regional ranking of the cortical regions in their frequency of abnormality across these subjects. Finally, three definitions of stages were applied to the ranking.

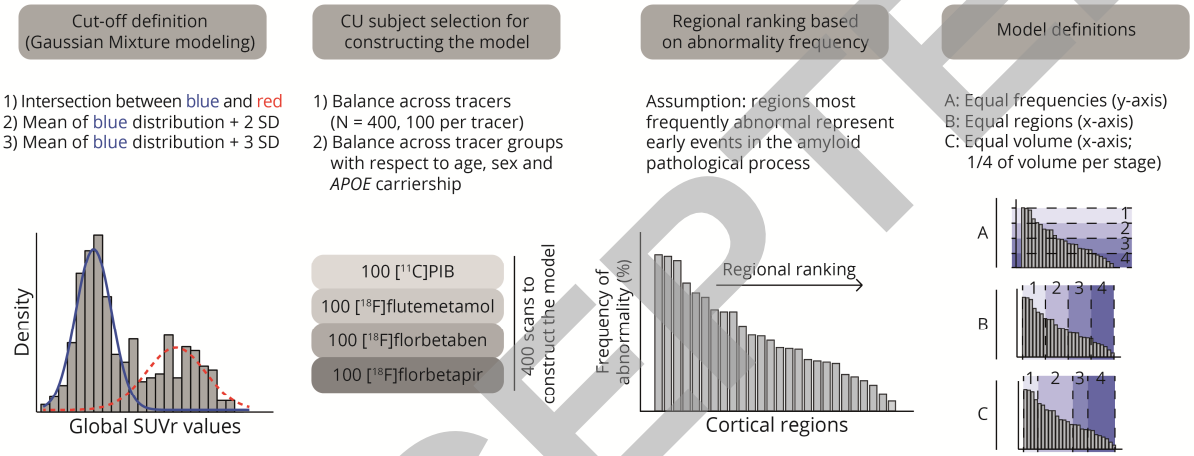

**Figure 2. Regional ranking per radiotracer** A) Frequency of regional abnormality across 100 CU subjects per radiotracer. B) Heatmap showing Spearman's rank correlation (rho) between the tracer-specific regional rankings. C) Correlation matrix displaying the correlation between each single-tracer regional ranking and the multi-tracer regional ranking based on the pooled data of 400 CU subjects.

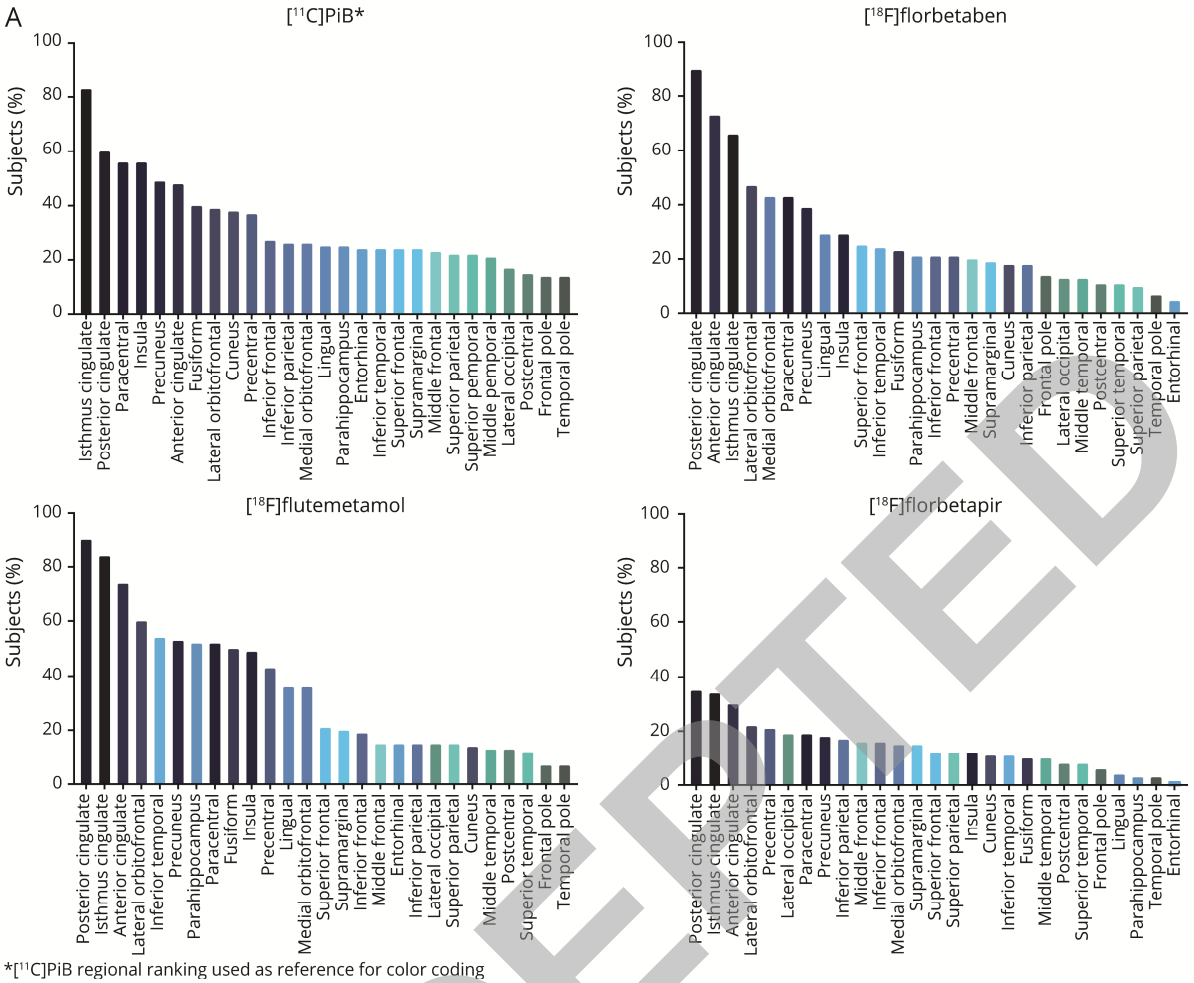

**Figure 3. The 5-stage (mean + 2SD equal frequencies) cortical amyloid staging model** A) Frequency of regional abnormality across 400 CU subjects used to construct the model. The colours represent the four different stages as defined by the equal frequency approach. B) Anatomical image displaying the brain regions involved in each stage. A stage was attributed when more than 50% of the encompassed regions displayed SUVr > cut-off (mean + 2SD cohort- and tracer-specific cutoff). Higher stages were only achieved once the staging conditions are also met for previous stages.

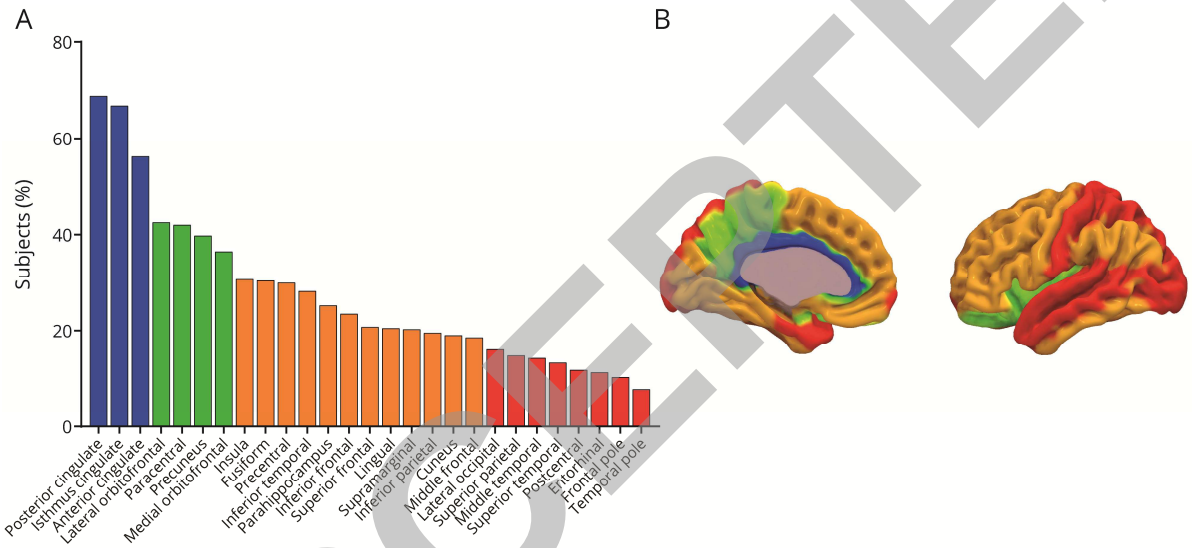

**Figure 4. Cross-sectional analyses** A) Baseline distribution of staging classification per cohort. Classification based on the amyloid staging model versus B) global amyloid PET classification, C) syndromic diagnosis, D) genetic risk, E) z-scored CSF Aβ42 levels, and F) log-transformed z-scored p-Tau values.

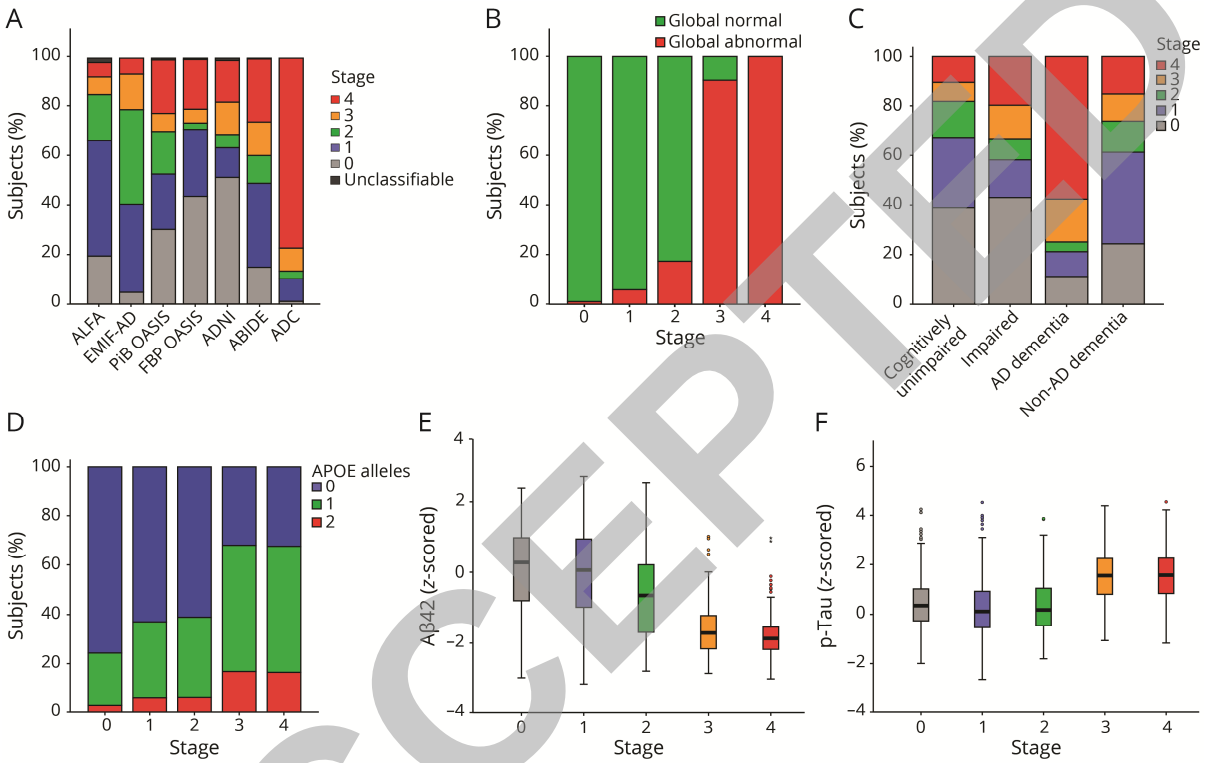

**Figure 5. Longitudinal analyses** Results from the linear mixed model analyses, displaying the effect of baseline amyloid stage on subsequent MMSE scores with coloured bands representing the 95% C.I. for ADNI (A) and OASIS (B) separately. Kaplan- 2 Meier survival plot displaying risk of progression to MMSE ≤ 25 per baseline amyloid stage for ADNI (C) and OASIS (D) separately.