BRAIN COMMUNICATIONS

Comparison of subtyping methods for neuroimaging studies in Alzheimer's disease: a call for harmonization

Rosaleena Mohanty,[1,*] Gustav Mårtensson,[1,*] Konstantinos Poulakis,[1] J-Sebastian Muehlboeck,[1] Elena Rodriguez-Vieitez,[1] Konstantinos Chiotis,[1] Michel J. Grothe,[2,3] Agneta Nordberg,[1,4] Daniel Ferreira[1,†] and Eric Westman,[1,5,†] for the Alzheimer's Disease Neuroimaging Initiative[‡]

* These authors are shared first authors.

† These authors are shared senior authors.

‡ Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.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 the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Biological subtypes in Alzheimer's disease, originally identified on neuropathological data, have been translated to *in vivo* biomarkers such as structural magnetic resonance imaging and positron emission tomography, to disentangle the heterogeneity within Alzheimer's disease. Although there is methodological variability across studies, comparable characteristics of subtypes are reported at the group level. In this study, we investigated whether group-level similarities translate to individual-level agreement across subtyping methods, in a head-to-head context. We compared five previously published subtyping methods. Firstly, we validated the subtyping methods in 89 amyloid-beta positive Alzheimer's disease dementia patients (reference group: 70 amyloid-beta negative healthy individuals) using structural magnetic resonance imaging. Secondly, we extended and applied the subtyping methods to 53 amyloid-beta positive prodromal Alzheimer's disease and 30 amyloid-beta positive Alzheimer's disease dementia patients (reference group: 200 amyloid-beta negative healthy individuals) using structural magnetic resonance imaging and tau positron emission tomography. Subtyping methods were implemented as outlined in each original study. Group-level and individual-level comparisons across methods were performed. Each individual subtyping method was replicated, and the proof-of-concept was established. At the group level, all methods captured subtypes with similar patterns of demographic and clinical characteristics, and with similar cortical thinning and tau positron emission tomography uptake patterns. However, at the individual level, large disagreements were found in subtype assignments. Although characteristics of subtypes are comparable at the group level, there is a large disagreement at the individual level across subtyping methods. Therefore, there is an urgent need for consensus and harmonization across subtyping methods. We call for the establishment of an open benchmarking framework to overcome this problem.

1 Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden
2 Unidad de Trastornos del Movimiento, Servicio de Neurología y Neurofisiología Clínica, Instituto de Biomedicina de Sevilla, Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville, Spain
3 Clinical Dementia Research Section, German Center for Neurodegenerative Diseases (DZNE), Rostock, Germany
4 Theme Aging, Karolinska University Hospital, Stockholm, Sweden
5 Department of Neuroimaging, Centre for Neuroimaging Sciences, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

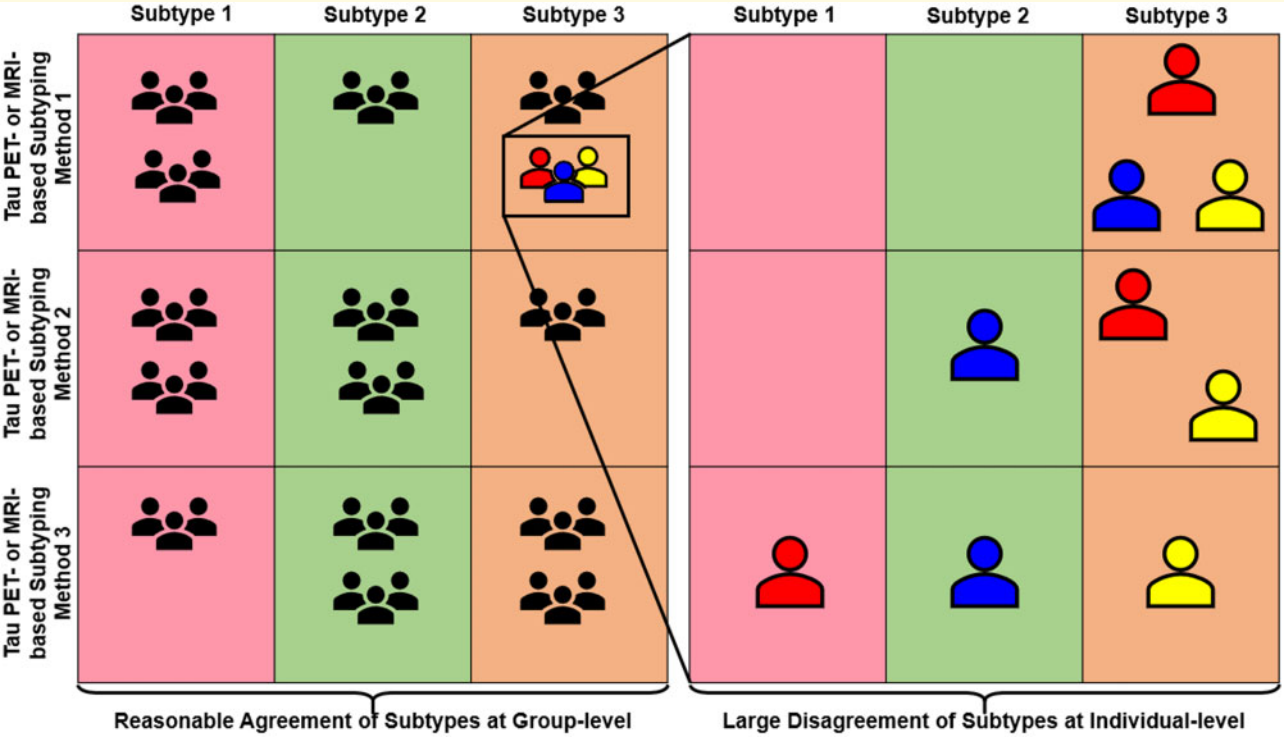
Received July 07, 2020. Revised September 17, 2020. Accepted October 05, 2020. Advance Access publication November 9, 2020

Correspondence to: Rosaleena Mohanty
Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet
Stockholm, Sweden
E-mail: rosaleena.mohanty@ki.se

Keywords: Alzheimer's disease; subtypes; structural MRI; tau PET; method comparison

Abbreviations: Aβ = amyloid-beta; ADNI = Alzheimer's disease neuroimaging initiative; APOE = apolipoprotein E; HC = healthy control; ICV = intracranial volume; MMSE = mini-mental state examination; NFT = neurofibrillary tangle; PET = positron emission tomography; PVC = partial volume correction; sMRI = structural magnetic resonance imaging; SUVR = standardized uptake value ratio.

Graphical Abstract



Introduction

The study of biological subtypes has opened a great opportunity to unravel the heterogeneity within Alzheimer's disease. The topic was rekindled in 2011 by the seminal study from Murray *et al.* (2011), and during the last 5 years, it has exploded with numerous structural magnetic resonance imaging (sMRI) subtyping studies [see Ferreira *et al.* (2020) for a review]. In 2018, the first tau positron emission tomography (PET) subtyping study was published (Whitwell *et al.*, 2018), and more are expected to come in the near future.

However, studies investigating Alzheimer's disease subtypes differ considerably, with almost no methodological consensus. Murray *et al.* (2011) based subtyping on postmortem tau neurofibrillary tangle (NFT) counts in the hippocampus and three cortical regions. All patients were at Braak's stage V or VI (Braak and Braak, 1995) and were classified into three subtypes according to the 25th and 75th percentiles in the hippocampus-to-cortex index: *typical Alzheimer's disease*, *limbic-predominant Alzheimer's disease* and *hippocampal-sparing Alzheimer's disease*. Byun *et al.* (2015) translated this subtyping method to sMRI data using volumes of the same brain regions as in Murray's method, but defined abnormality as −1 standard deviation from age-, sex- and intracranial volume (ICV)-adjusted normative data of healthy controls. This method identified a fourth subtype: *minimal atrophy Alzheimer's disease* (Byun *et al.*, 2015). In contrast, Risacher *et al.* (2017) followed the 25th and 75th percentiles procedure using the hippocampus-to-cortex index but extended the three cortical regions used by Murray *et al.* (2011) to seven regions. Risacher *et al.* (2017) also corrected for age, sex and ICV, but they based this correction on a reference group of amyloid-beta negative (Aβ−) healthy controls and used a different correction method additionally including the MRI field strength. Ferreira *et al.* and follow-up studies from our

lab used visual rating scales of brain atrophy in medial temporal, frontal and posterior cortices (Ferreira *et al.*, 2017, 2018, 2019; Persson *et al.*, 2017; Ekman *et al.*, 2018; Oppedal *et al.*, 2019; Machado *et al.*, 2020) and determined clinical cut points for abnormality (Ferreira *et al.*, 2015). We also used unsupervised clustering in another cross-sectional study by Poulakis *et al.* (2018), which was recently extended for subtyping on longitudinal data (Poulakis *et al.*, 2020). Other groups used different unsupervised clustering methods (Noh *et al.*, 2014; Dong *et al.* 2015, 2017; Hwang *et al.* 2016; Na *et al.*, 2016; Zhang *et al.* 2016; Park *et al.* 2017; Varol *et al.* 2017), highlighting the methodological variability across studies. Additionally, Charil *et al.* (2019) recently translated Murray's method to tau PET while Whitwell *et al.* (2018) applied a clustering method on tau PET data.

Despite this variability, all these studies tend to identify subtypes with similar characteristics, arguing for validation [see Ferreira *et al.* (2020) for a review]. However, this validation is reported at the group level. The ultimate goal of investigating heterogeneity in Alzheimer's disease is to understand individual variability, hence, necessitating individual-level validation. Surprisingly, no head-to-head comparison of subtyping methods has been published so far. Such a comparison arises as an urgent and important step towards facilitating consistent progress in this field, especially with the current surge in subtyping studies using sMRI investigating subtype or disease progression (Young *et al.*, 2018; Marinescu *et al.*, 2019; Poulakis *et al.*, 2020) and tau PET (Whitwell *et al.*, 2018; Charil *et al.*, 2019; Jeon *et al.*, 2019). To illustrate this problem, in the present study, we applied different subtyping methods reported in five previous studies (Murray *et al.*, 2011; Byun *et al.*, 2015; Ferreira *et al.*, 2017; Risacher *et al.*, 2017; Poulakis *et al.*, 2018; Charil *et al.*, 2019) on sMRI and tau PET data from the same cohort. Thereby, we substantiated our claim for the need for harmonizing subtyping methods, which aims at achieving consensus at group- *and* individual-levels despite methodological differences. In our primary analyses, we performed a head-to-head comparison and report subtypes' frequencies, characteristics and cortical thickness and tau PET uptake maps from the different methods. In our secondary analyses, we investigated how methodological variations influence the performance of the different subtyping methods. We hypothesized that across subtyping studies, the comparability of subtypes at the group level may not translate to the individual level.

Materials and methods

Participants

All participants were selected from the Alzheimer's Disease Neuroimaging Initiative (ADNI; http://adni.loni.usc.edu/). The goal of the ADNI (launched in 2003, principal investigator: Michael W. Weiner; Mueller *et al.*, 2005) is to measure the progression of prodromal Alzheimer's disease and early Alzheimer's disease using MRI, PET, biomarkers and clinical and neuropsychological assessments. We included two separate ADNI cohorts:

Firstly, since subtypes have been predominantly identified in Alzheimer's disease dementia, we validated the previously published subtyping methods using sMRI in a cohort of 89 Alzheimer's disease dementia patients (Aβ+) from ADNI-1. We also included a control group of 70 Aβ− healthy individuals (HC). Amyloid status was determined by cerebrospinal fluid biomarkers ($A\beta_{1\text{-}42}$ cut-off = 192 pg/ml) (Shaw *et al.*, 2009).

Secondly, subtyping was applied to a cross-sectional cohort of 84 patients (54 Aβ+ prodromal Alzheimer's disease patients, 30 Aβ+ Alzheimer's disease dementia patients) subsampled from ADNI-2 and -3 using sMRI and tau PET. The control group comprised 200 Aβ− HC. Amyloid status was determined through amyloid PET (florbetapir SUVR cut-off = 1.11; Joshi *et al.*, 2012 or florbetaben SUVR cut-off = 1.08, following ADNI's current recommendation, http://adni.loni.usc.edu/).

We will refer to these two cohorts as the *sMRI cohort* (ADNI-1, Alzheimer's disease dementia patients) and the *sMRI-tauPET cohort* (ADNI-2 and -3, prodromal Alzheimer's disease and Alzheimer's disease dementia patients). We validated the previously published methods (Byun *et al.*, 2015; Ferreira *et al.*, 2017; Risacher *et al.*, 2017; Poulakis *et al.*, 2018; Charil *et al.*, 2019) in the sMRI cohort and extended our analyses to the sMRI-tauPET cohort. The study protocol followed by all participating centres within the ADNI was approved by their respective institutional review board. Informed and written consent was obtained from all the participants.

MRI and PET imaging

MRI acquisition and processing

3D accelerated T1-weighted sequences were acquired with sagittal slices and voxel size $1.1 \times 1.1 \times 1.2$ mm^3. MRI data for the ADNI-1 were acquired on 1.5 T scanners, and MRI data for ADNI-2 and -3 were acquired on 3.0 T scanners.

For the sMRI cohort, processed data were already available from our previous studies (Ferreira *et al.*, 2017; Poulakis *et al.*, 2018). For methods from other labs (Risacher *et al.*, 2017; Byun *et al.*, 2015) and for all the methods in the sMRI-tauPET cohort, data were unavailable, so we processed the sMRI through TheHiveDB system (Muehlboeck *et al.*, 2014) with FreeSurfer 6.0.0 (http://freesurfer.net/). Following the cross-sectional stream, quality control of the output from FreeSurfer was conducted visually. Automatic region of interest parcellation yielded volumetric measures for cortical and subcortical brain structures (Fischl *et al.*, 2002; Desikan *et al.*, 2006; Destrieux *et al.*, 2010). For the subtyping method using visual rating scales (Ferreira *et al.*, 2017), the

ratings were computed automatically using Automatic Visual Ratings of Atrophy v0.8 (https://github.com/gsmartensson/avra_public) (Mårtensson *et al.*, 2019), a deep learning model trained on over 3000 MRI scans rated by an expert neuroradiologist with excellent inter-rater agreement (Mårtensson *et al.*, 2020).

Tau PET acquisition and processing

Tau PET scans were collected using PET/CT scanners. [^{18}F]AV-1451 was injected with a dosage of 370 MBq (10.0 mCi) ± 10% and scans were acquired between 75 and 105 min post-injection. The dynamic acquisition was 30 min long and comprised 6 × 5 min frames. For each tau PET scan, a sMRI was available within 90 days (except in three Alzheimer's disease dementia and five prodromal Alzheimer's disease patients, >90 days).

For subtyping methods using tau PET (Murray *et al.*, 2011; Byun *et al.*, 2015; Risacher *et al.*, 2017; Charil *et al.*, 2019), processing was performed using the PetSurfer Toolbox (Greve *et al.*, 2016) within FreeSurfer 6.0.0. AV-1451 images were co-registered onto the corresponding FreeSurfer-processed sMRI. The regions (cortical and subcortical grey matter) estimated for each individual were consistent with those used for sMRI-based subtypes (Desikan *et al.*, 2006). Partial volume correction (PVC) was applied using the symmetric geometric matrix method (ROUSSET and OG 1998). AV-1451 signal was quantified in each region as the standardized uptake value ratio (SUVR), computed with the cerebellum grey matter as the reference region with PVC.

Subtyping methods

Based on two recent systematic reviews (Ferreira *et al.*, 2020; Habes *et al.*, 2020), we identified four sources of methodological variation in subtyping studies:

i. Type of method (hypothesis-driven versus data-driven).
ii. Definition of subtype (dependent on the sample of study versus dependent on an external reference group).
iii. Modality (postmortem NFT versus sMRI versus tau PET).
iv. Measure (regional NFT count versus automated regional volumes/SUVR values versus gross visual ratings).

The method proposed by Murray *et al.* (2011) is the only one based on postmortem NFT count and motivated subsequent neuroimaging studies. In this study, we focused on neuroimaging-based methods based on five subtyping studies, covering all these levels of methodological variation: Risacher *et al.* (Risacher *et al.*, 2017), Byun *et al.* (Risacher *et al.*, 2017), Ferreira *et al.* (Ferreira *et al.*, 2017), Poulakis *et al.* (Poulakis *et al.*, 2018) and Charil *et al.* (Charil *et al.*, 2019). Each subtyping method was implemented to replicate the original method as closely as possible, as elaborated further in Table 1, Fig. 1 and Supplementary Table 1. We also translated some sMRI-based methods to tau PET to test subtyping based on tau pathology (Byun *et al.*, 2015; Risacher *et al.*, 2017). For Byun's method on tau PET, we identified a minimal tau subtype that is not captured by Charil's or Risacher's methods.

Quantification of AV-1451 signal in the hippocampus, a key region for subtyping in many studies (Byun *et al.*, 2015; Risacher *et al.*, 2017; Charil *et al.*, 2019), is contentious (Lee *et al.*, 2018; Lemoine *et al.*, 2018). Hence, we additionally applied subtyping using the entorhinal cortex instead of the hippocampus, also facilitating comparability with the study by Whitwell *et al.* (2018).

Methodological variations

As a secondary objective, we implemented the following methodological variations, evaluating their potential impact on agreements among subtyping methods:

i. *The effect of using three versus seven cortical regions in Risacher's method*
Although Risacher *et al.* (2017) translated Murray's method (Murray *et al.*, 2011) to sMRI, Risacher's method included seven cortical regions instead of the original three regions in Murray's method. Here, we compared these two versions of Risacher's method: with three versus seven cortical regions.
ii. *The effect of statistical corrections for ICV and age on sMRI methods*
In our primary analysis, we evaluated the method by Risacher *et al.* (2017) (seven cortical regions) by adjusting for ICV and age using a single regression model for both covariates. Here, we evaluated the impact of adjusting for ICV only, or adjusting for ICV and age using separate regression models for each covariate. We also performed these comparisons for Risacher's method using three cortical regions.
iii. *The effect of statistical corrections for age on tau PET methods*
In the primary analysis of tau PET-based subtyping (Byun *et al.*, 2015; Risacher *et al.*, 2017; Charil *et al.*, 2019), potential covariates were not considered. Correction for ICV is not necessary unlike in sMRI methods, but age may potentially affect tau PET SUVR (Schöll *et al.*, 2016). Here, we compared subtyping with age-corrected SUVR and uncorrected SUVR.
iv. *The effect of PVC on tau PET-based subtyping methods*
In the primary analysis, we used PVC for reliably quantifying tau PET SUVR, accounting for any off-target binding, especially in the hippocampus (Ikonomovic *et al.*, 2016; Lowe *et al.*, 2016). Here, we compared subtyping between PVC SUVR and non-PVC SUVR.

Statistical analysis

We compared subtyping methods at the group-level in terms of age, sex, mini-mental state exam (MMSE), education and *APOE* ε4 status. Within each subtyping method, hypothesis testing was performed to compare the

Table 1 Overview of the subtyping methods implemented in this study

Method	Type of method	Definition of subtypes	Modality	Measure	Subtypes	Graphical representation
Charil (25)	Hypothesis-driven	Within-sample dependent	tau PET	SUVR	TAD, LP, HS	Fig. 1A
Risacher (6)	Hypothesis-driven	Within-sample dependent	sMRI and tau PET	Automated volumes and SUVR	TAD, LP, HS	Fig. 1B
Byun (5)	Hypothesis-driven	External reference group	sMRI and tau PET	Automated volumes and SUVR	TAD, LP, HS, MA[a]	Fig. 1C
Ferreira (7)	Hypothesis-driven	External reference group	sMRI	Visual ratings	TAD, LP, HS, MA	Fig. 1D
Poulakis (15)	Data-driven	Within-sample dependent	sMRI	Automated volumes	TAD[b], LP, HS, MA	Fig. 1E

HS = hippocampal-sparing; LP = limbic-predominant; MA = minimal atrophy; SUVR = standardized uptake value ratio; TAD = typical AD.
[a]MA corresponds to the subtype identified by the sMRI-based method. For the tau PET-based method, the corresponding subtype would be minimal tau.
[b]The two clusters reflecting typical Alzheimer's disease patterns in the original publication by Poulakis *et al.* (15) were combined into a single typical Alzheimer's disease subtype to allow comparisons across subtyping methods.

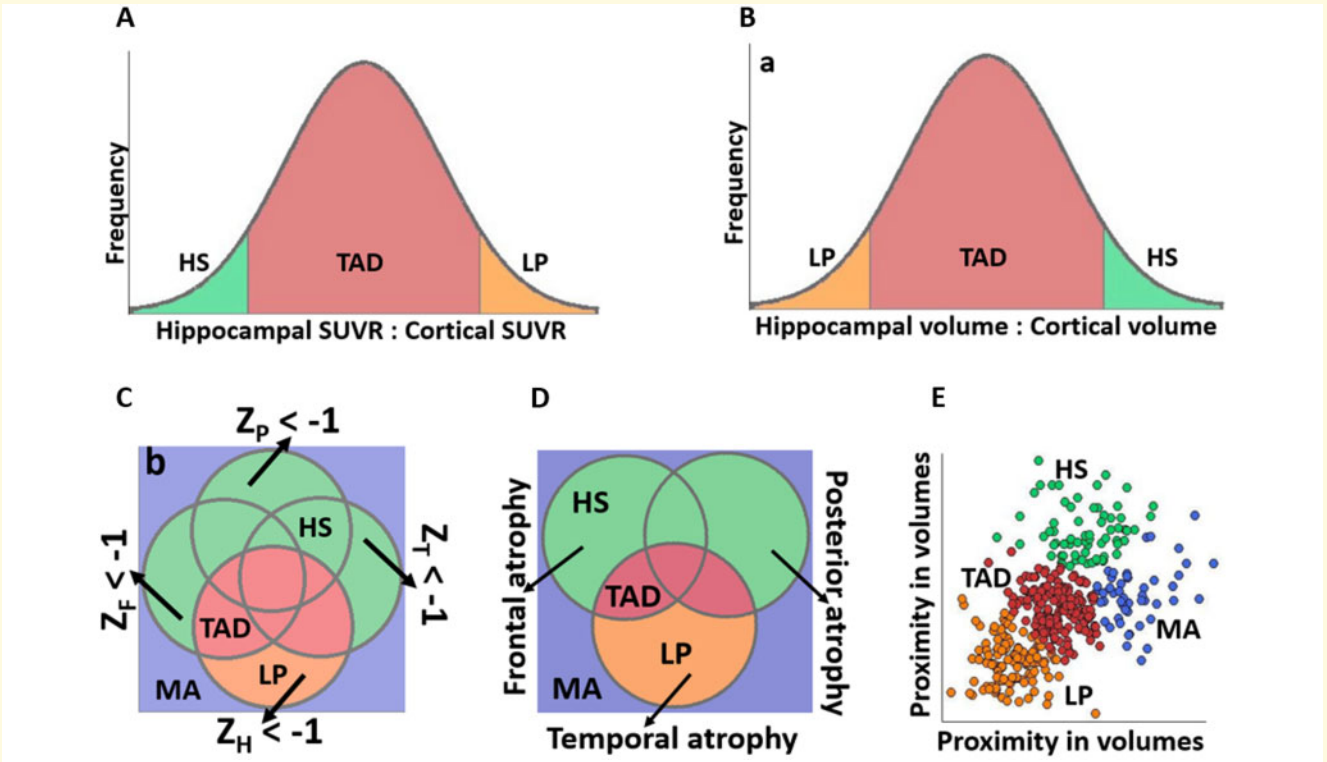


Figure 1 Graphical representation of the subtyping methods implemented in this study. [a]This figure corresponds to the sMRI-based method. For the tau PET-based method, volume measures are replaced with SUVR and classification of LP and HS is reversed; [b]Z_H = z-score for hippocampus; Z_F = z-score for frontal regions; Z_P = z-score for parietal regions; Z_T = z-score for temporal regions. This figure corresponds to the sMRI-based method and z-scores are computed for volumes. For the tau PET-based method, volume measures are replaced with SUVR and abnormal tau levels have z-scores ≥ 1. HS = hippocampal-sparing; LP = limbic-predominant; MA = minimal atrophy; SUVR = standardized uptake value ratio; TAD = typical AD.

distribution of subtypes with the Kruskal–Wallis test. A *P*-value ≤ 0.05 was deemed significant. Group-level cortical thickness and PVC tau PET uptake maps were generated by comparing each subtype with the healthy controls. In each hemisphere, data were smoothed onto the surface using a 10 mm Gaussian kernel with a full width at half maximum. A general linear model was fitted at each vertex. All maps were visualized at $P \leq 0.01$ (uncorrected). Individual-level agreement among subtyping methods was quantified by Cohen's kappa ($\kappa < 0$, no agreement; $\kappa = 0$–0.20, slight agreement; $\kappa = 0.21$–0.40, fair agreement; $\kappa < 0.41$–0.60, moderate agreement;

$\kappa = 0.61–0.80$, substantial agreement; $\kappa = 0.81–1.0$, almost perfect agreement) (Landis and Koch, 1977).

Data availability

Source data are available as a part of the ADNI. All data generated or analysed during this study are included within this article and its supplementary information files.

Results

Table 2(a, b) shows the demographic and clinical characteristics for the sMRI cohort and sMRI-tauPET cohorts, respectively.

Validation of subtyping methods in the sMRI cohort

The frequencies of the subtypes in the sMRI cohort were very similar to the frequencies reported in the original studies (Byun *et al.*, 2015; Ferreira *et al.*, 2017; Risacher *et al.*, 2017; Poulakis *et al.*, 2018; Charil *et al.*, 2019) suggesting we could replicate the subtyping methods (Table 3).

Group-level comparison of subtyping methods in the sMRI and sMRI-tauPET cohorts

Figure 2 shows that, at the group-level, the subtyping methods captured similar demographic and clinical characteristics of the subtypes in both cohorts. Typical Alzheimer's disease was always the most frequent subtype and showed a greater frequency of males and lower MMSE scores relative to the other subtypes. Limbic-predominant Alzheimer's disease showed lower MMSE scores relative to hippocampal-sparing Alzheimer's disease. Hippocampal-sparing Alzheimer's disease was the subtype with the lowest frequency of *APOE* ε4 carriers. Minimal atrophy/minimal tau Alzheimer's disease included younger individuals and showed higher MMSE scores.

Figures 3 and 4 and Supplementary Table 2 show that, at the group-level, the subtyping methods captured similar cortical thickness and PVC tau PET uptake maps of the subtypes relative to healthy individuals. Cortical thinning and elevated tau PET uptake included widespread regions in typical Alzheimer's disease; temporal and limbic regions in limbic-predominant Alzheimer's disease; frontal or parietal regions in hippocampal-sparing Alzheimer's disease; and relatively fewer regions in minimal atrophy/minimal tau Alzheimer's disease, across all subtyping methods. Typical and limbic-predominant Alzheimer's disease showed smaller hippocampal volume and greater hippocampal tau PET SUVR relative to hippocampal-sparing and minimal atrophy/minimal tau Alzheimer's disease (boxplots in Figs 3 and 4). Figure 5 shows the group-level tau PET uptake maps for entorhinal-based subtyping instead of hippocampus-based subtyping. Compared to hippocampus-based subtyping, albeit similar maps, hippocampal-sparing Alzheimer's disease in entorhinal-based subtyping showed no tau PET uptake in lateral temporal lobe regions. Greater tau SUVR in the entorhinal cortex was seen in typical and limbic-predominant Alzheimer's disease compared to hippocampal-sparing and minimal tau Alzheimer's disease (boxplots in Fig. 5).

Head-to-head comparison of subtyping methods in the sMRI-tauPET cohort

Figure 6A and C shows the head-to-head comparison of individual-level subtype assignments. Agreement among methods was low, reflected by low values of κ. Agreement among the tau PET-based methods was relatively higher than that of the sMRI-based methods. Since not all methods identify the minimal atrophy/minimal tau Alzheimer's disease, we excluded this subtype in follow-up analyses and observed increased κ values in both cohorts and modalities (Fig. 6B and D). ADNI's participant identifiers are listed in Supplementary Fig. 2 and in Supplementary Data File.

Methodological variations in the sMRI-tauPET cohort

When supplementing our head-to-head comparisons with several methodological variations, we observed the following (Supplementary Data File):

i. *The effect of using three versus seven cortical regions in Risacher's method*
   Results from Risacher's method using three cortical regions were consistent with Risacher's method using seven cortical regions (85% agreement).
ii. *The effect of statistical corrections for ICV and age on sMRI methods*
   Relative to Risacher's method (seven cortical regions and adjusted for ICV and age in a single model), 82% of the individuals were classified consistently when performing the ICV correction only, and 69% when performing the ICV and age correction with separate models. Relative to the variation in Risacher's method using three cortical regions (and adjusted for ICV and age in a single model), 98% of the individuals were classified consistently when performing the ICV correction only, and 74% when performing the ICV and age correction with separate models. Overall, agreements were better in typical Alzheimer's disease (79–88%) compared to the other subtypes (15–83%).
iii. *The effect of statistical corrections for age on tau PET methods*

Table 2 Demographic and clinical characteristics of the cohorts

(a) Validation of subtyping methods in AD dementia patients (sMRI cohort)

	HC (Aβ−)	AD dementia (Aβ+)	*P*-value
N	70	89	
Sex (F,%)	51	39	0.139
Age (years)	75.15 ± 5.22 [62, 89]	74.73 ± 7.72 [57, 88]	0.757
Education (years)	15.66 ± 2.65 [8, 20]	15.16 ± 3.24 [4, 20]	0.402
APOE ε4 carriers (%)	10	74	<0.0001
MMSE	29.04 ± 1.10 [25, 30]	23.48 ± 1.87 [20, 26]	<0.0001
Word recall task	2.86 ± 1.17 [0 5.67]	6.24 ± 1.34 [3.33, 9.33]	<0.0001
Naming objects and fingers	0.08 ± 0.28 [0, 1]	0.43 ± 0.56 [0, 2]	<0.0001
Following commands	0.05 ± 0.23 [0, 1]	0.40 ± 0.63 [0, 3]	<0.0001
Constructional praxis	0.41 ± 0.49 [0, 1]	0.91 ± 0.65 [0, 3]	<0.0001
Ideational praxis	0.05 ± 0.23 [0, 1]	0.33 ± 0.72 [0, 5]	0.0007
Orientation	0.12 ± 0.37 [0, 2]	2.06 ± 1.61 [0, 7]	<0.0001
Word recognition task	2.38 ± 2.24 [0, 12]	6.51 ± 2.88 [1, 2]	<0.0001
Recall of test instructions	0 ± 0 [0, 0]	0.31 ± 0.88 [0, 5]	0.0005
Spoken language	0.02 ± 0.16 [0, 1]	0.30 ± 0.64 [0, 3]	0.0006
Word finding difficulty	0.04 ± 0.20 [0, 1]	0.59 ± 0.95 [0, 4]	<0.0001
ADAS total score	6.06 ± 2.79 [1.67, 14.33]	18.38 ± 6.26 [8.67, 42.67]	<0.0001

(b) Subtyping methods in prodromal AD and AD dementia patients (sMRI-tauPET cohort)

	HC (Aβ−)	Prodromal AD (Aβ+)	AD dementia (Aβ+)	*P*-value
N	200	54	30	
Sex (F, %)	59	48	50	0.285
Age (years)	70.45 ± 5.65 [55.8, 89][a]	74.09 ± 7.34 [59.4, 90.1][a]	77.46 ± 8.27 [55.9, 91.2][a]	<0.0001
Education (years)	16.90 ± 2.31 [11, 20][a]	15.76 ± 2.66 [12, 20][b]	15.77 ± 2.57 [12, 20][b]	0.002
APOE ε4 carriers (%)	22	61	53	<0.0001
MMSE	29.24 ± 1.05 [23, 30][a]	27.48 ± 2.30 [19, 30][a]	22.13 ± 4.23 [9, 30][a]	<0.0001
Word recall task	2.36 ± 1.81 [0 6][a]	4.34 ± 1.49 [1, 7][a]	6.09 ± 1.61 [3, 10][a]	<0.0001
Naming objects and fingers	0.03 ± 0.37 [0, 3]	0.04 ± 0.19 [0, 1][b]	0.56 ± 0.89 [0, 3][b]	<0.0001
Following commands	0.06 ± 0.24 [0, 1]	0.22 ± 0.41 [0, 1][b]	0.40 ± 0.91 [0, 3][b]	0.0011
Constructional praxis	0.33 ± 0.55 [0, 3]	0.48 ± 0.57 [0, 2]	0.72 ± 0.79 [0, 3][b]	0.0084
Ideational praxis	0.05 ± 0.38 [0, 5]	0.08 ± 0.34 [0, 2]	0.28 ± 0.54 [0, 2][b]	<0.0001
Orientation	0.09 ± 0.29 [0, 1][a]	0.50 ± 0.73 [0, 3][a]	2.8 ± 2.1 [0, 7][a]	<0.0001
Word recognition task	4.92 ± 1.91 [0, 10][a]	5.9 ± 2.77 [0, 12][a]	9.28 ± 2.73 [3, 12][a]	<0.0001
Recall of test instructions	0.005 ± 0.07 [0, 1]	0.14 ± 0.53 [0, 3][b]	0.95 ± 1.39 [0, 5][b]	<0.0001
Spoken language	0.005 ± 0.07 [0, 1][a]	0.04 ± 0.19 [0, 1][a]	0.44 ± 1.00 [0, 4][a]	<0.0001
Word finding difficulty	0.04 ± 0.27 [0, 3][a]	0.18 ± 0.43 [0, 2][a]	0.88 ± 1.05 [0, 3][a]	<0.0001
ADAS total score	8.08 ± 2.86 [1, 19.33][a]	11.98 ± 4.66 [3, 24][a]	21.68 ± 7.48 [7, 37][a]	<0.0001

Data are reported as mean ± standard deviation [minimum, maximum]; Hypothesis testing was performed using the Kruskal–Wallis test for the continuous variables and χ^2 test for the nominal variables. Additionally, the Kruskal–Wallis test was performed pairwise between groups in the sMRI-tauPET cohort.
AD = Alzheimer's disease; ADAS = Alzheimer's Disease Assessment Scale-Cognitive Subscale; *APOE* = apolipoprotein; Aβ = amyloid-beta; F = female; HC = healthy control; MMSE = mini-mental state examination; PET = positron emission tomography; sMRI = structural MRI.
[a]Significantly different from each of the other two groups.
[b]Significantly different from HC group only.

Table 3 Frequencies of the subtypes compared with previously published studies in the sMRI cohort

Subtype	Risacher		Byun		Ferreira		Poulakis	
	Pub.	This study	Pub.	This study	Pub.	This study	Pub.[a]	This study
Typical AD	69	69	59	55	51	52	69	66
Hippocampal-sparing	17	19	12	17	17	19	7	10
Limbic-predominant	14	12	19	21	17	18	4	1
Minimal atrophy			10	7	15	11	19	23

Data are reported as % and rounded to the nearest integer for readability.
AD = Alzheimer's Disease; Pub. = published study.
[a]Frequencies of subtypes based on the ADNI cohort only, since the original study by Poulakis *et al.* (15) also includes another cohort.

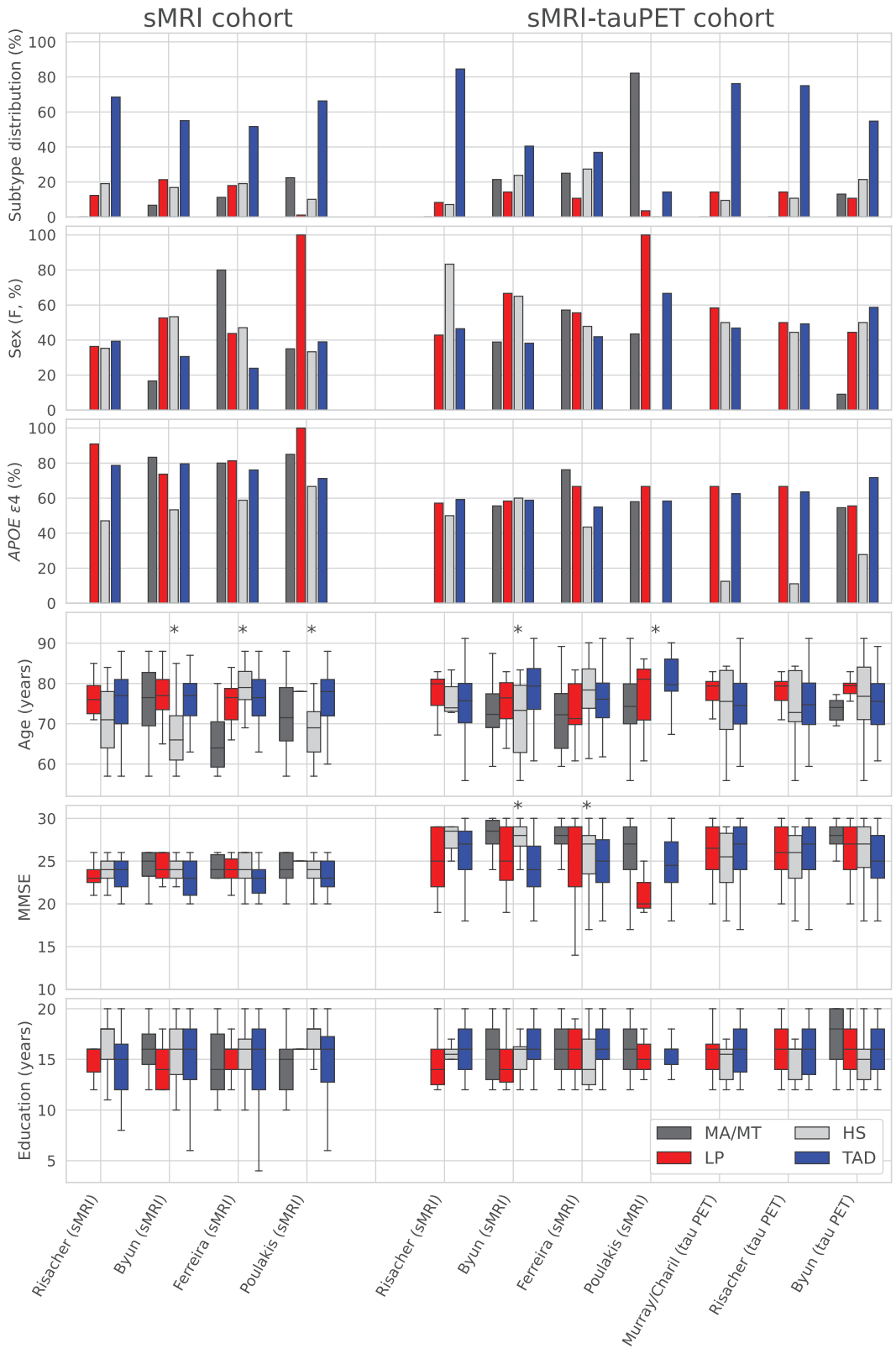


Figure 2 Demographic and clinical characteristics captured by the different subtyping methods. The bar plots for sex and APOE ε4 show what percentage of patients in each subtype were females and APOE ε4 carriers, respectively. Kruskal–Wallis hypothesis testing was conducted comparing the subtypes within each method; *$P < 0.05$ within the subtyping method. APOE = apolipoprotein; F = female; HS = hippocampal-sparing; LP = limbic-predominant; MA = minimal atrophy; MMSE = mini mental state exam; MT = minimal tau; PET = positron emission tomography; sMRI = structural magnetic resonance imaging; TAD = typical AD.

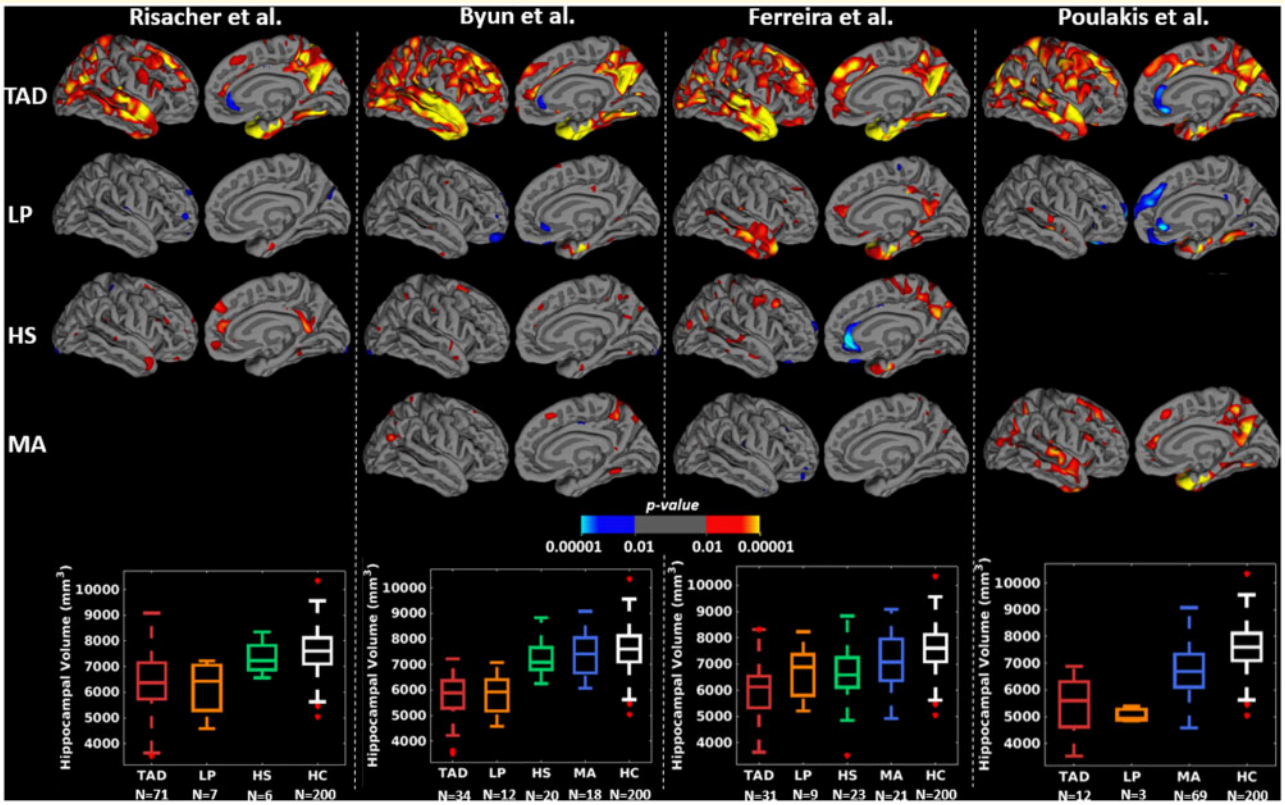


Figure 3 Group-level cortical thickness maps across subtyping methods in the sMRI-tauPET cohort. For simplicity, only left lateral and medial views are presented since very similar results were obtained for the right lateral and medial views. Differences in cortical thickness maps are shown in each subtype relative to HC, generated by fitting general linear model at each vertex. Yellow-red regions reflect thinner cortex in Alzheimer's disease subtypes relative to HC. All brain maps are uncorrected for multiple comparisons at $P < 0.01$. Risacher *et al.* identified three subtypes only and hence, there are no cortical maps corresponding to MA subtype. Poulakis *et al.* identified all four subtypes. However, the HS subtype (one individual) had to be excluded from the study due to invalid tau PET data. HC=healthy control; HS = hippocampal-sparing; LP = limbic-predominant; MA= minimal atrophy; TAD = typical AD.

Over 80% of the individuals were consistently classified with and without age-adjusted tau SUVR (agreement for: Charil's method = 89%; Risacher's method = 100%; Byun's method = 80%).

iv. *The effect of PVC on tau PET-based subtyping methods* Over 80% of the individuals were consistently classified with PVC and non-PVC SUVR (agreement for: Charil's method = 87%; Risacher's method = 89%; Byun's method = 80%). Overall, agreements were better in typical Alzheimer's disease (83–94%) compared to the other subtypes (56–78%).

Discussion

The field of biological subtypes of Alzheimer's disease has expanded rapidly in the last decade, with numerous recent publications on neuropathological, MRI and PET data. However, the great methodological variability is complicating reaching a definitive understanding of the heterogeneity within Alzheimer's disease. The current study is the first head-to-head comparison of several subtyping methods in the same cohort. We found that different methods identify subtypes that are largely comparable at the group level (similar frequencies, demographic, clinical characteristics, cortical thinning and tau PET uptake). However, strikingly, the individual-level agreement among subtyping methods is very low when compared head-to-head. This result may have important implications for advancing the implementation of precision medicine. Below, we discuss several factors that may explain this finding and ways to minimize this problem in future studies.

Comparability across studies at the group level suggests a convergence of results and initial consensus on the existence of three to four major subtypes: typical, limbic-predominant and hippocampal-sparing Alzheimer's disease in all the studies, and minimal atrophy Alzheimer's disease in several studies. Minimal atrophy Alzheimer's disease is only identified when considering disease severity, while the other subtypes are identified when considering typicality (Ferreira *et al.*, 2020). The dimensions of

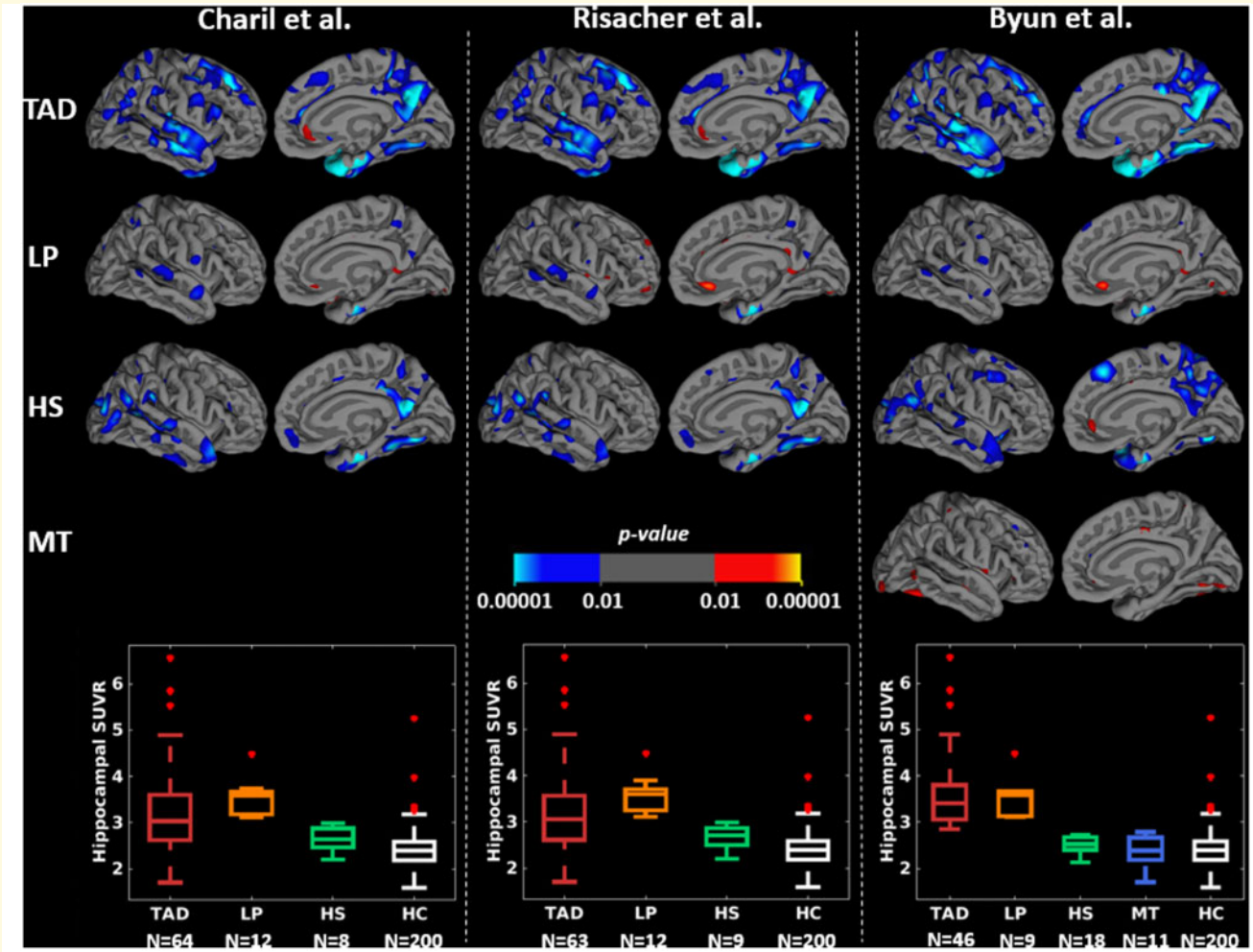


Figure 4 Group-level PVC tau PET uptake maps across subtyping methods using the hippocampus in the sMRI-tauPET cohort. For simplicity, only left lateral and medial views are presented since very similar results were obtained for the right lateral and medial views. Differences in tau PET uptake maps are shown in each subtype relative to HC, generated by fitting general linear model at each vertex. Cyan regions reflect greater PVC tau PET uptake in Alzheimer's disease subtypes relative to HC. All brain maps are uncorrected for multiple comparisons at $P < 0.01$. HC = healthy control; HS = hippocampal-sparing; LP = limbic-predominant; MT = minimal tau; TAD = typical AD.

severity and typicality were defined in a recent conceptual framework for biological subtypes of Alzheimer's disease (Ferreira *et al.*, 2020). Typicality spans from limbic-predominant to hippocampal-sparing, with typical Alzheimer's disease in-between. Severity differentiates minimal atrophy from typical Alzheimer's disease, accounting for neurodegeneration.

The seminal study by Murray *et al.* (2011) based subtyping on tau NFT in the hippocampus and three cortical regions. Importantly, all the patients had a pathological diagnosis of Alzheimer's disease with Braak stage of V or VI (Braak and Braak, 1995). This means that all patients had NFT in the hippocampus by definition, and the method focused on separating the subset of patients with NFT predominantly in the hippocampus (limbic-predominant Alzheimer's disease) from those with NFT predominantly in the cortical regions (hippocampal-sparing Alzheimer's disease). Remainder of the patients had a rather balanced NFT count in the hippocampus and cortical regions and were classified as typical Alzheimer's disease.

Murray's method (Murray *et al.*, 2011) motivated many subsequent sMRI studies (Ferreira *et al.*, 2020; Habes *et al.*, 2020). However, these studies rely on sMRI, a marker of unspecific neurodegeneration. This raises several problems. Firstly, while sMRI can reliably track neuropathologically defined subtypes (Whitwell *et al.*, 2012), the actual distribution of NFT in sMRI subtypes remains largely unknown. Recent studies have provided interesting preliminary data on tau PET uptake in sMRI-based subtypes (Jeon *et al.*, 2019; Ossenkoppele *et al.*, 2020). Secondly, the published sMRI subtype studies quite likely included patients in Braak NFT Stage IV or lower. Thirdly, most sMRI studies investigated cohorts including both amyloid-beta positive and negative

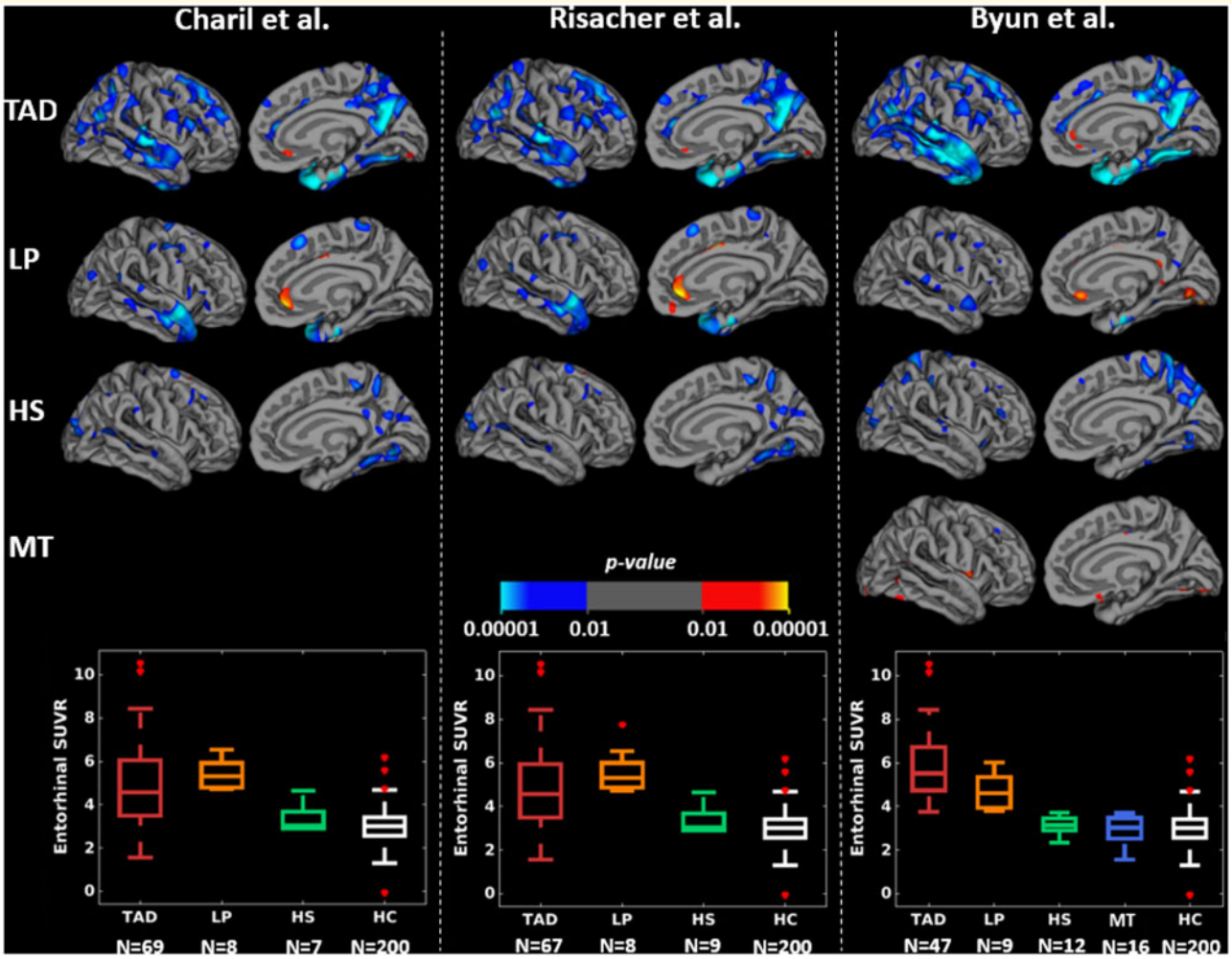


Figure 5 Group-level PVC tau PET uptake maps across subtyping methods using the entorhinal cortex in the sMRI-tauPET cohort. For simplicity, only left lateral and medial views are presented since very similar results were obtained for the right lateral and medial views. Differences in tau PET uptake maps are shown in each subtype relative to HC, generated by fitting general linear model at each vertex. Blue-cyan regions reflect PVC greater tau PET uptake in Alzheimer's disease subtypes relative to HC. All brain maps are uncorrected for multiple comparisons at $P < 0.01$. HC = healthy control; HS = hippocampal-sparing; LP = limbic-predominant; MT = minimal tau; TAD = typical AD.

Alzheimer's disease dementia patients except a few (Risacher *et al.*, 2017; ten Kate *et al.*, 2018), while all the patients in Murray *et al.* (2011) had a pathological diagnosis of Alzheimer's disease. Fourthly, neurodegeneration is downstream to NFT pathology (Dubois *et al.*, 2014), and there is a time gap until overt brain atrophy can be visually observed or captured by automatic methods for data analysis. Nonetheless, some data-driven methods may capture subtle differences in regional covariance in the absence of overt brain atrophy, mitigating this problem. Altogether, we still need a better understanding of the correspondence between neuropathologically, sMRI- and tau PET-defined subtypes. A major contribution of our current study is that subtypes identified with sMRI and tau PET are not interchangeable at the individual level.

At the group level, findings for the demographic and clinical measures were in agreement with previously reported studies and a recent meta-analysis (Ferreira *et al.*, 2020). Broadly, typical Alzheimer's disease was the most frequent subtype; typical and limbic-predominant Alzheimer's disease were older in comparison to the hippocampal-sparing and minimal atrophy Alzheimer's disease; MMSE scores were mostly comparable across subtypes with minimal atrophy Alzheimer's disease showing the highest scores; a lower proportion of *APOE* ε4 carriers belonged to hippocampal-sparing relative to typical and limbic-predominant Alzheimer's disease and hippocampal-sparing Alzheimer's disease had the highest levels of education.

Overall, head-to-head comparisons revealed greater agreement of tau PET-based methods than the sMRI-based methods. This could be potentially attributed to: (i) lower resolution of tau PET and smaller proximity of the key regions involved in subtyping, (ii) comparison of

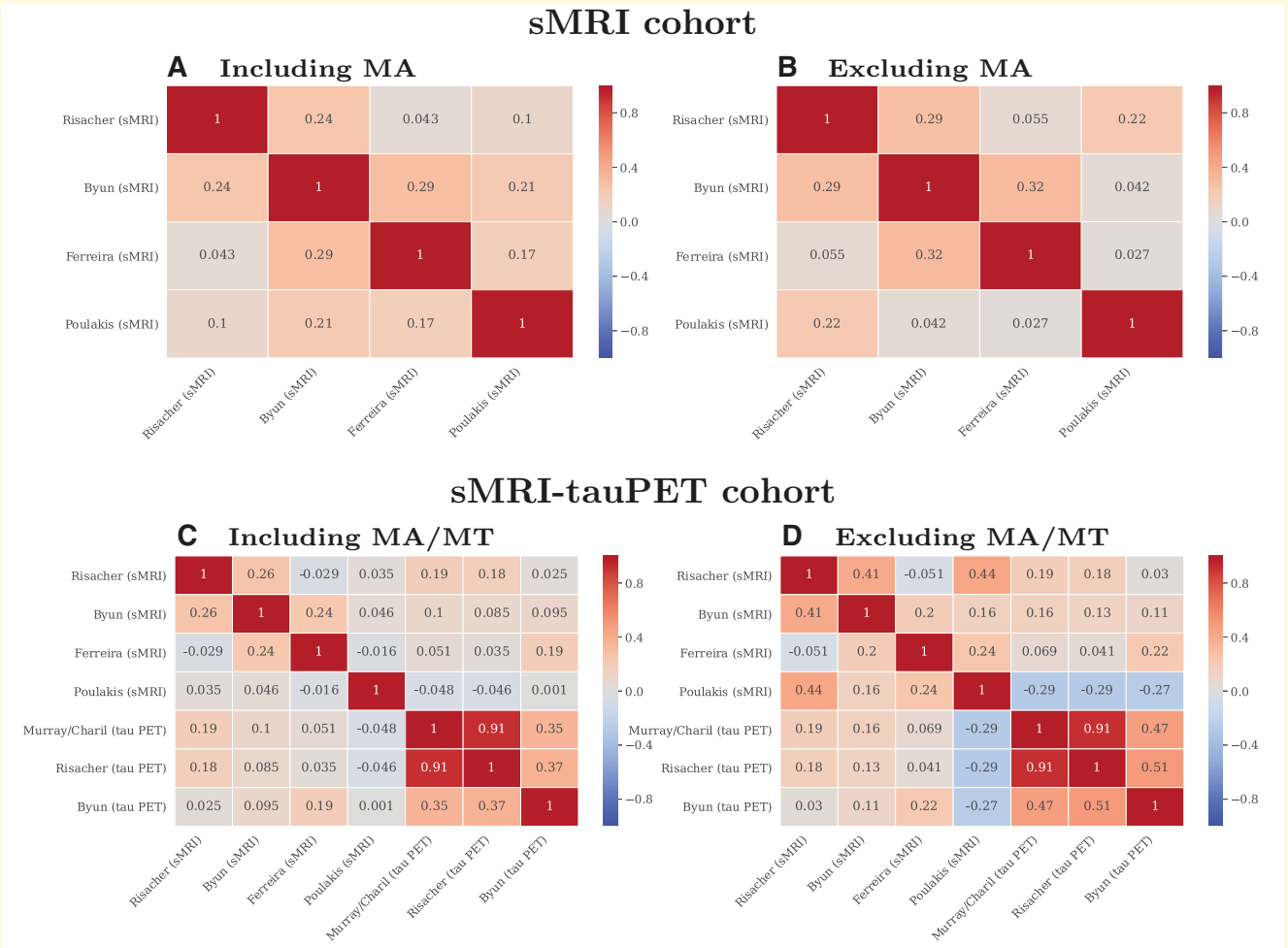


Figure 6 Individual-level agreement among subtyping methods as illustrated by Cohen's kappa values. MA = minimal atrophy AD; MT = minimal tau; PET = positron emission tomography; sMRI = structural magnetic resonance imaging.

merely three tau PET-based subtyping methods with relatively less methodological variability or (iii) a more consistent and direct emulation of postmortem NFT captured in Murray *et al.* (2011) by tau PET compared to sMRI which can capture variance unrelated to subtyping. Future tau PET-based subtyping methods could shine light on this finding. Low levels of agreement across subtyping methods at the individual-level could be ascribed to a combination of one or more of the following factors: (i) variation in cut points of atrophy or tau uptake used to define the subtypes which may differ from dataset to dataset; (ii) accounting for (or lack thereof) the two dimensions of subtypes, namely typicality and severity, to different degrees by different methods; (iii) forced allocation of each individual into a subtype without an associated measure of (un)certainty; (iv) accounting for (or lack thereof) within-subtype variability (i.e. strong/weak resemblance) in the biological profiles of individuals assigned to the same subtype. We call for the investigation of these factors as a promising avenue to increase the agreement between subtyping methods in the future.

Biologically, the head-to-head agreements are best understood by considering individual exemplars. A consistent scenario is participant identifier 2239: across the sMRI methods, this individual was classified as hippocampal-sparing or minimal atrophy Alzheimer's disease whereas, across the tau PET methods, the individual was classified as typical Alzheimer's disease. The difference in sMRI-based subtyping could be attributed to differences in cut points for abnormality across methods. The fact that the corresponding tau PET-based subtype was typical Alzheimer's disease (higher severity) could suggest greater tau pathology relative to structural atrophy. A more challenging case is participant identifier 6377: across the sMRI methods, this individual was classified as typical, hippocampal-sparing, limbic-predominant, or minimal atrophy Alzheimer's disease, whereas across the tau PET-based method, the individual was classified as limbic-predominant Alzheimer's disease. Some differences in sMRI-based subtyping are relatively more plausible than others, considering the above-mentioned typicality and severity dimensions (Ferreira *et al.*, 2020). To instantiate, it

may be plausible that this individual demonstrated typical Alzheimer's disease (with one method; Risacher *et al.*, 2017) and limbic-predominant Alzheimer's disease (with another method; Ferreira *et al.*, 2017), as these two subtypes are close to each other along the typicality dimension (Ferreira *et al.*, 2020). However, classification as limbic-predominant Alzheimer's disease (with one method; Ferreira *et al.*, 2017) and hippocampal-sparing Alzheimer's disease (with another method; Byun *et al.*, 2015) seem incompatible, since these two subtypes correspond to the extremities of the typicality dimension. Therefore, a classification with all four subtypes for the same individual leaves the case biologically uninterpretable, calling for consensus across subtyping in the field as we aim for precision medicine.

Despite having several caveats, previous neuroimaging-based subtyping studies have made important contributions. Byun *et al.* (2015) and Risacher *et al.* (2017) translated the NFT-based method by Murray *et al.* (2011) to sMRI data, and Charil *et al.* (2019) translated the method to tau PET. Our analyses of methodological variations showed that the age correction made a stronger impact on agreements among methods than the number of cortical regions or the PVC. This impact was more prominent for sMRI-based methods than for tau PET-based methods; and for limbic-predominant and hippocampal-sparing subtypes than for typical Alzheimer's disease. Contribution of aging to hippocampal atrophy may be at the basis of this finding. Lower disagreement in typical Alzheimer's disease relative to the other subtypes is akin to the diagnostic challenge in the clinical setting. An interesting result of our study is that the method of adjustment (single model for all covariates versus separate models for each covariate) increased the disagreement. Future studies should take this finding into account when deciding on how to correct for potential confounders.

Ongoing research is moving the field forward by characterization of subtypes not only in Alzheimer's disease dementia but also at earlier stages such as prodromal Alzheimer's disease (Zhang *et al.*, 2016; ten Kate *et al.*, 2018; Young *et al.*, 2018; Machulda *et al.*, 2019). Preliminary data show that such characterization could be extended and evaluated at even the earliest stages of preclinical Alzheimer's disease or individuals with subjective cognitive decline (Jung *et al.*, 2016). In speculation, relative to full-blown dementia, atrophy levels are likely modest even if there exists overt tau pathology at pre-dementia stages. This could result in a greater dissociation between atrophy and tau pathology, further leading to lower agreement across subtyping methods. In this scenario, group-level comparisons alone are insufficient. Individual-level agreement is thus warranted, and lack thereof will prevent or delay the use of subtyping in clinical routine, clinical trials, and research. Therefore, there is an urgent need for harmonization of the different subtyping methods.

To this end, we advocate for establishing a framework for benchmarking for future studies. A possibility could be selection of a well-characterized cohort (preferably with multimodal antemortem and postmortem data in a longitudinal setting). This could include preparing a dataset comprised of cognitively normal individuals (amyloid-beta negative) and individuals on the Alzheimer's disease continuum (preclinical Alzheimer's disease, prodromal Alzheimer's disease and Alzheimer's disease dementia). Multiple longitudinal biomarkers during life, such as neuroimaging such as MRI (structural, diffusion, functional, etc.) and PET (fluorodeoxyglucose, amyloid, tau), cerebrospinal fluid, plasma and neuropsychological measures, could enable characterization of the subtypes *in vivo* while neuropathological assessments can provide a ground truth for subtyping. Unimodal (based on a single image modality) as well as multimodal (based on combination of image modalities) subtypes should be differentiated and demonstrated within the same cohort. Additionally, the establishment of clear evaluation metrics would allow for comparison of the performance of the subtyping methods and could include group-level characteristics, individual-level results, cut points for each measure used for subtyping, variability in cut points after accounting for potential covariates, the certainty of assignment of subtype, variability in biomarker profiles within the same subtype, etc. Greater similarity across multiple evaluation metrics across methods would thus, ensure harmonization across subtyping methods. The dataset should be standard so that it can be utilized by future subtyping methods to ensure individual-level consistency across methods. The dataset should also be open and accessible to all researchers in the field. Once validated, the subtyping method could obviously be extended to independent populations and data. As a preliminary step, we provide all the data used for subtyping in this study along with ADNI participant identifiers (Supplementary Data File).

This study has some limitations. The cohort was part of the ADNI, which has strict selection criteria and excludes individuals with non-amnestic presentations or cerebrovascular pathology. It is likely that agreement among subtyping methods is different in clinically oriented or more heterogeneous cohorts. The number of Alzheimer's disease dementia patients was limited, and prodromal Alzheimer's disease patients, in which the degree of atrophy may be smaller than in Alzheimer's disease dementia, were overrepresented within the sMRI-tau PET cohort used to demonstrate the lack of consensus in subtyping. However, the additional and relatively large sMRI cohort of Alzheimer's disease dementia patients strengthens and illustrates the case in point. Hypothesis-driven methods are well covered in our study (Murray *et al.*, 2011; Byun *et al.*, 2015; Ferreira *et al.*, 2017; Risacher *et al.*, 2017; Charil *et al.*, 2019). However, previous subtyping studies have applied many different data-driven methods. Methods, especially involving clustering,

can differ on if and/or how they account for critical aspects of certainty of subtype allocation and variability within each identified subtype. We selected Poulakis' method (Poulakis *et al.*, 2018), which resulted in notably distinct subtyping potentially due to being most methodologically different from the rest, and our current study cannot provide direct insight on methods used by other groups (Ferreira *et al.*, 2020; Habes *et al.*, 2020). However, the selection of subtyping methods illustrates the case made in the current study. We based our analyses on cross-sectional tau PET and sMRI data. The next step should be to include longitudinal data. However, the availability of such a dataset is limited at present, particularly for tau PET. Longitudinal data will be relevant to investigate disease progression in the subtypes, disentangling the disagreement due to the temporal lag between NFT accumulation (tau PET) and brain atrophy (sMRI) from pure methodological noise. Finally, the tau PET tracer used in our study, [^{18}F]AV-1451, is a first generation tracer with known off-target binding (Leuzy *et al.*, 2019). Better agreement among the tau PET-based subtyping methods than their sMRI counterparts could indicate that they need to be further pursued. Second generation tau PET tracers (e.g. 18F-RO-948, 18F-MK-6240 and 18F-PI-2620) would be relatively more sensitive and specific, especially at the preclinical and prodromal stages of Alzheimer's disease, although their longitudinal trajectories remain to be fully investigated and validated (Bischof *et al.*, 2020).

The field of biological subtypes is expanding rapidly with the investigation of multiple modalities/biomarkers and extending to pre-dementia stages and other neurodegenerative diseases (Habes *et al.*, 2020). We conclude that subtyping methods may appear comparable across studies, at the group-level. However, a major finding of the present study is the large disagreement among subtyping methods based on tau PET and especially sMRI at the individual level. Hence, there is an urgent need for consensus and harmonization across subtyping methods. To achieve this, we suggest establishment of an accessible and standard framework for benchmarking. A comprehensive dataset along with clear evaluation metrics will facilitate a fair comparison, ultimately ensuring better agreement among future subtyping methods.

Supplementary material

Supplementary material is available at *Brain Communications* online.

Competing interests

The authors have no conflicts of interest.

Funding

This study was supported by the Swedish Foundation for Strategic Research (SSF); the Strategic Research Programme in Neuroscience at Karolinska Institutet (StratNeuro); the Swedish Research Council (VR, 2016-02282); the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet; Center for Innovative Medicine (CIMED); the Swedish Alzheimer Foundation; the Swedish Brain Foundation; the Åke Wiberg Foundation; Demensfonden; Stiftelsen Olle Engkvist Byggmästare; Birgitta och Sten Westerberg; Demensförbundet; Stiftelsen för Ålderssjukdomar vid Karolinska Institutet. Michel Grothe is supported by the 'Miguel Servet' program [CP19/00031] of the Spanish Instituto de Salud Carlos III (ISCIIIFEDER). The funding sources did not have any involvement in the study design; collection, analysis, and interpretation of data; writing of the report; and the decision to submit the article for publication.

Acknowledgements

Data collection and sharing for this study 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: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. 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 Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles.

References

Bischof GN, Dodich A, Ashton NJ, Boccardi M, Barthel H, Carrillo MC, et al. Alzheimer's disease biomarker roadmap 2020: second-generation tau PET tracers. In: 2020 Alzheimer's Association International Conference. ALZ; 2020.

Braak H, Braak EVA. Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol Aging 1995; 16: 271–8.

Byun MS, Kim SE, Park J, Yi D, Choe YM, Sohn BK, Alzheimer's Disease Neuroimaging Initiative, et al. Heterogeneity of regional brain atrophy patterns associated with distinct progression rates in Alzheimer's disease. PLoS One 2015; 10: e0142756.

Charil A, Shcherbinin S, Southekal S, Devous MD, Mintun M, Murray ME, et al. Tau subtypes of Alzheimer's disease determined in vivo using flortaucipir PET imaging. J Alzheimer's Dis 2019; 1–12.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006; 31: 968–80.

Destrieux C, Fischl B, Dale A, Halgren E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 2010; 53: 1–15.

Dong A, Honnorat N, Gaonkar B, Davatzikos C. CHIMERA: clustering of heterogeneous disease effects via distribution matching of imaging patterns. IEEE Trans Med Imaging 2016; 35: 612–21.

Dong A, Toledo JB, Honnorat N, Doshi J, Varol E, Sotiras A, et al. Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. Brain 2017; 140: 735–47.

Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol 2014; 13: 614–29.

Ekman U, Ferreira D, Westman E. The A/T/N biomarker scheme and patterns of brain atrophy assessed in mild cognitive impairment. Sci Rep 2018; 8: 1–10.

Ferreira D, Cavallin L, Larsson E-M, Muehlboeck J-S, Mecocci P, Vellas, B. et al.; the AddNeuroMed consortium and the Alzheimer's Disease Neuroimaging Initiative. Practical cut-offs for visual rating scales of medial temporal, frontal and posterior atrophy in Alzheimer's disease and mild cognitive impairment. J Intern Med 2015; 278: 277–90.

Ferreira D, Nordberg A, Westman E. Biological subtypes of Alzheimer disease: a systematic review and meta-analysis. Neurology 2020; 94: 436–48.

Ferreira D, Pereira JB, Volpe G, Westman E. Subtypes of Alzheimer's disease display distinct network abnormalities extending beyond their pattern of brain atrophy. Front Neurol 2019; 10: 524.

Ferreira D, Shams S, Cavallin L, Viitanen M, Martola J, Granberg T, et al. The contribution of small vessel disease to subtypes of Alzheimer's disease: a study on cerebrospinal fluid and imaging biomarkers. Neurobiol Aging 2018; 70: 18–29.

Ferreira D, Verhagen C, Hernández-Cabrera JA, Cavallin L, Guo C-J, Ekman U, et al. Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: longitudinal trajectories and clinical applications. Sci Rep 2017; 7: 46263.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002; 33: 341–55.

Greve DN, Salat DH, Bowen David Izquierdo-Garcia SL, Schultz AP, Catana C, Becker JA. et al. Different partial volume correction methods lead to different conclusions: an 18F-FDG-PET study of aging. Neuroimage 2016; 132: 334–43.

Habes M, Grothe MJ, Tunc B, McMillan C, Wolk DA, Davatzikos C. Disentangling heterogeneity in Alzheimer's disease and related dementias using data-driven methods. Biol Psychiatry 2020; 88: 70–82.

Hwang J, Kim CM, Jeon S, Lee JM, Hong YJ, Hoon Roh J, Alzheimer's Disease Neuroimaging Initiative, et al. Prediction of Alzheimer's disease pathophysiology based on cortical thickness patterns. Alzheimer's Dementia 2016; 2: 58–67.

Ikonomovic MD, Abrahamson EE, Price JC, Mathis CA, Klunk WE. [F-18] AV-1451 positron emission tomography retention in choroid plexus: more than 'off-target' binding. Ann Neurol 2016; 80: 307–8.

Jeon S, Kang JM, Seo S, Jeong HJ, Funck T, Lee S-Y, et al. Topographical heterogeneity of Alzheimer's disease based on MR imaging, tau PET, and amyloid PET. Front Aging Neurosci 2019; 11: 211.,

Joshi AD, Pontecorvo MJ, Clark CM, Carpenter AP, Jennings DL, Sadowsky CH, the Florbetapir F 18 Study Investigators, et al. Performance characteristics of amyloid PET with florbetapir F 18 in patients with Alzheimer's disease and cognitively normal subjects. J Nuclear Med 2012; 53: 378–84.

Jung N-Y, Seo SW, Yoo H, Yang J-J, Park S, Kim YJ, et al. Classifying anatomical subtypes of subjective memory impairment. Neurobiology of Aging 2016; 48: 53–60.

Kate M, ten E, Dicks PJ, Visser WM, van der Flier CE, Teunissen F, Barkhof Alzheimer's Disease Neuroimaging Initiative, et al. Atrophy Subtypes in Prodromal Alzheimer's Disease Are Associated with Cognitive Decline. Brain 2018; 141: 3443–56., and

Landis Richard JKoch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977; 33: 159–74.

Lee CM, Jacobs HIL, Marquié M, Becker JA, Andrea NV, Jin DS, et al. 18F-Flortaucipir Binding in Choroid Plexus: Related to Race and Hippocampus Signal. JAD 2018; 62: 1691–702.

Lemoine L, Leuzy A, Chiotis K, Rodriguez-Vieitez E, Nordberg A. Tau positron emission tomography imaging in tauopathies: the added hurdle of off-target binding. Alzheimer's Dement 2018; 10: 232–6.

Leuzy A, Chiotis K, Lemoine L, Gillberg P-G, Almkvist O, Rodriguez-Vieitez E, et al. Tau PET imaging in neurodegenerative tauopathies—still a challenge. Mol Psychiatry 2019; 24: 1112–34.

Lowe VJ, Curran G, Fang P, Liesinger AM, Josephs KA, Parisi JE, et al. An autoradiographic evaluation of AV-1451 tau PET in dementia. Acta Neuropathol Commun 2016; 4: 58.

Machado A, Ferreira D, Grothe MJ, Eyjolfsdottir H, Almqvist PM, Cavallin L, et al. The Cholinergic System and Treatment Response in Subtypes of Alzheimer's disease. Alzheimer's Res Ther 2020; 12: 51.

Machulda MM, Lundt ES, Albertson SM, Kremers WK, Mielke MM, Knopman DS, et al. Neuropsychological subtypes of incident mild cognitive impairment in the Mayo Clinic study of aging. Alzheimer's Dement 2019; 15: 878–87.

Marinescu RV, Eshaghi A, Lorenzi M, Young AL, Oxtoby NP, Garbarino S, et al.; and Alzheimer's Disease Neuroimaging Initiative. DIVE: a spatiotemporal progression model of brain pathology in neurodegenerative disorders. NeuroImage 2019; 192: 166–77.

Mårtensson G, Ferreira D, Cavallin L, Muehlboeck J-S, Wahlund L-O, Wang C, et al. AVRA: automatic visual ratings of atrophy from MRI images using recurrent convolutional neural networks. NeuroImage: Clinical 2019; 23: 101872.

Mårtensson G, Ferreira D, Granberg T, Cavallin L, Oppedal K, Padovani A, et al.. The reliability of a deep learning model in clinical out-of-distribution MRI data: a multicohort study. Med Image Anal 2020; 66: 101714.

Muehlboeck J, Westman E, Simmons A. TheHiveDB image data management and analysis framework. Front Neuroinform 2014; 7: 49.

Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack C, Jagust W, et al. The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin 2005; 15: 869–77.

Murray ME, Graff-Radford NR, Ross OA, Petersen RC, Duara R, Dickson DW. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. Lancet Neurol 2011; 10: 785–96.

Na HK, Kang DR, Kim S, Seo SW, Heilman KM, Noh Y, et al. Malignant progression in parietal-dominant atrophy subtype of Alzheimer's disease occurs independent of onset age. Neurobiol Aging 2016; 47: 149–56.

Noh Y, Jeon S, Lee JM, Seo SW, Kim GH, Cho H, et al. Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. Neurology 2014; 83: 1936–44.,

Oppedal K, Ferreira D, Cavallin L, Lemstra AW, Ten Kate M, Padovani A, for the Alzheimer's Disease Neuroimaging Initiative, et al. A signature pattern of cortical atrophy in dementia with lewy bodies: a study on 333 patients from the European DLB Consortium. Alzheimer's Dement 2019; 15: 400–9.

Ossenkoppele R, Lyoo C H, Sudre C H, Westen D, Cho H, Ryu Y H, et al.. Distinct tau PET patterns in atrophy-defined subtypes of Alzheimer's disease. Alzheimer's Dement 2020; 16: 335–44.

Park J-Y, Na HK, Kim S, Kim H, Kim HJ, Seo SW, Alzheimer's Disease Neuroimaging Initiative, et al. Robust identification of Alzheimer's disease subtypes based on cortical atrophy patterns. Sci Rep 2017; 7: 1–14.

Persson K, Sakshaug Eldholm R, Barca ML, Cavallin L, Ferreira D, Brita Knapskog A, et al. MRI-assessed atrophy subtypes in Alzheimer's disease and the cognitive reserve hypothesis. PLoS One 2017; 12: e0186595.

Poulakis K, Ferreira D, Pereira JB, Smedby Ö, Vemuri P, Westman E. Fully bayesian longitudinal unsupervised learning for the assessment and visualization of AD heterogeneity and progression. Aging 2020; 12: 12622–47.

Poulakis K, Pereira JB, Mecocci P, Vellas B, Tsolaki M, Kłoszewska I, et al. Heterogeneous patterns of brain atrophy in Alzheimer's disease. Neurobiol Aging 2018; 65: 98–108.

Risacher SL, Anderson WH, Charil A, Castelluccio PF, Shcherbinin S, Saykin AJ, Alzheimer's Disease Neuroimaging Initiative, et al. Alzheimer disease brain atrophy subtypes are associated with cognition and rate of decline. Neurology 2017; 89: 2176–86. and

Rousset OG. Correction for partial volume effects in PET: principle and validation. J Nucl Med 1998; 39: 904–11.

Schöll M, Samuel N, Lockhart Daniel R, Schonhaut James P, O'Neil M, Janabi R, Ossenkoppele, et al. PET imaging of tau deposition in the aging human brain. Neuron 2016; 89: 971–82.

Shaw LM, Vanderstichele H, Knapik M, -Czajka CM, Clark PS, Aisen RC, Alzheimer's Disease Neuroimaging Initiative, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol 2009; 65: 403–13.

Varol E, Sotiras A, Davatzikos C, Alzheimer's Disease Neuroimaging Initiative. HYDRA: revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. Neuroimage 2017; 145: 346–64.

Whitwell JL, Dickson DW, Murray ME, Weigand SD, Tosakulwong N, Senjem ML, et al. Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: a case-control study. Lancet Neurol 2012; 11: 868–77.

Whitwell JL, Graff-Radford J, Tosakulwong N, Weigand SD, Machulda M, Senjem ML, et al. [18 F]AV-1451 clustering of entorhinal and cortical uptake in Alzheimer's disease. Ann Neurol 2018; 83: 248–57.

Young AL, Marinescu RV, Oxtoby NP, Bocchetta M, Yong K, Firth NC, The Genetic FTD Initiative (GENFI), et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. Nat Commun 2018; 9: 1–16.

Zhang X, Mormino EC, Sun N, Sperling RA, Sabuncu MR, Yeo BTT; Alzheimer's Disease Neuroimaging Initiative. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. Proc Natl Acad Sci USA 2016; 113: E6535–E6544.