White matter hyperintensities and their relationship to cognition: Effects of segmentation algorithm

Meral A. Tubi [a,1], Franklin W. Feingold [a,b,1], Deydeep Kothapalli [a], Evan T. Hare [a], Kevin S. King [c], Paul M. Thompson [a], Meredith N. Braskie [a,*], for the Alzheimer's Disease Neuroimaging Initiative[2]

[a] Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Marina del Rey, CA, 90292, USA
[b] Stanford University, Stanford, CA, 94305, USA
[c] Huntington Medical Research Institute, Imaging Division, Pasadena, CA, 91105, USA

A B S T R A C T

White matter hyperintensities (WMHs) are brain white matter lesions that are hyperintense on fluid attenuated inversion recovery (FLAIR) magnetic resonance imaging (MRI) scans. Larger WMH volumes have been associated with Alzheimer's disease (AD) and with cognitive decline. However, the relationship between WMH volumes and cross-sectional cognitive measures has been inconsistent. We hypothesize that this inconsistency may arise from 1) the presence of AD-specific neuropathology that may obscure any WMH effects on cognition, and 2) varying criteria for creating a WMH segmentation. Manual and automated programs are typically used to determine segmentation boundaries, but criteria for those boundaries can differ. It remains unclear whether WMH volumes are associated with cognitive deficits, and which segmentation criteria influence the relationships between WMH volumes and clinical outcomes.

In a sample of 260 non-demented participants (ages 55–90, 141 males, 119 females) from the Alzheimer's Disease Neuroimaging Initiative (ADNI), we compared the performance of five WMH segmentation methods, by relating the WMH volumes derived using each method to both clinical diagnosis and composite measures of executive function and memory. To separate WMH effects on cognition from effects related to AD-specific processes, we performed analyses separately in people with and without abnormal cerebrospinal fluid amyloid levels.

WMH volume estimates that excluded more diffuse, lower-intensity lesions were more strongly correlated with clinical diagnosis and cognitive performance, and only in those without abnormal amyloid levels. These findings may inform best practices for WMH segmentation, and suggest that AD neuropathology may mask WMH effects on clinical diagnosis and cognition.

1. Introduction

White matter hyperintensities (WMHs) in the brain white matter are lesions having a signal intensity brighter than the surrounding white matter on a magnetic resonance imaging (MRI) fluid attenuation inversion recovery (FLAIR) sequence (Yoshita et al., 2006). WMHs are associated with vascular risk (Prins and Scheltens, 2015; Scott et al., 2015) and may represent increased blood brain barrier permeability, plasma leakage, and degeneration of axons and myelin (Haller et al., 2013). WMH volumes are associated with older age, Alzheimer's disease

(AD), small vessel disease, and cognitive decline, making them a measure of clinical interest (Brickman et al., 2009; Prins and Scheltens, 2015).

AD-specific processes may influence the observed effect of WMHs on clinical diagnosis and cognition. In cross-sectional data, amyloid plaque counts do not correlate as strongly with cognition as neurofibrillary tangle counts (Wilcock and Esiri, 1982). Still the presence of amyloid positivity in cognitively intact older adults is considered to be a sign of preclinical AD (Hane et al., 2017), and is associated with faster longitudinal decline in cognitive function compared to that seen in amyloid-negative older adults (Mortamais et al., 2017). WMHs and amyloid deposition in AD may influence one another (Grimmer et al., 2012; Scott et al., 2015, 2016), and both may contribute to cognitive impairment (Provenzano et al., 2013; Gordon et al., 2015). We used amyloid positivity as a surrogate for AD-specific processes, which may influence cognition independently of, and together with, WMHs. We studied the effect of WMHs on cognition by evaluating the relationship separately in those who were amyloid positive (Aβ+) or negative (Aβ-) (Shaw et al., 2009). We hypothesized that the relationship between WMH volume and cognition would be stronger in those who were Aβ- (and thus had less cognitive variability added by AD-related processes) compared to those who were Aβ+.

Larger WMH volumes have been associated with both decreased global cognitive function (Au et al., 2006; Frisoni et al., 2007; Kloppenborg et al., 2014) and domain specific-cognitive impairment, including executive function (Gunning-Dixon and Raz, 2000; Smith et al., 2011; Lampe et al., 2017; Aljondi et al., 2018) and memory (de Groot et al., 2000; Gunning-Dixon and Raz, 2000; Smith et al., 2011; Lampe et al., 2017). However, results vary among studies that have evaluated the WMHs to cognition relationship. This variability may arise from differences in the classification of lesion boundaries in segmentation methods (Smart et al., 2011; Caligiuri et al., 2015; Wang et al., 2015; Dadar et al., 2017), which may capture physiologically different components (Haller et al., 2013).

Manual segmentation - the gold standard when comparing automated methods - is time consuming, requiring multiple raters, and training to establish intra-rater and inter-rater reliability. Further, expert reviewers in different laboratories may use different visual rating scales or may disagree about what constitutes a clinically-relevant WMH boundary or location. Therefore, acceptable intra-study reliability may not translate into high reliability between methods or studies (Grimaud et al., 1996; Mantyla et al., 1997; Kapeller et al., 2003; Prins et al., 2004; Yoshita et al., 2005). Often, limited information is provided in publications to describe the criteria used for defining manual segmentations - such as whether to include minimally hyperintense lesions, or lighter 'halos' around larger higher-intensity lesions. This makes ground truth and replication across studies difficult (Firbank et al., 2004; Gibson et al., 2010; Smart et al., 2011; Iorio et al., 2013; Griffanti et al., 2018). It is unclear which WMH manual segmentation criteria result in the most clinically-relevant lesion assessments (van Straaten et al., 2006).

We calculated WMH volumes using the default options for five automated WMH segmentation algorithms. Our goal was not to evaluate the software packages themselves, all of which can be optimized, but rather to create a range of typical segmentations that allowed us to identify which features strengthened the sensitivity to detecting a relationship between WMH volumes and cognitive measures in Aβ+ and Aβ- non-demented older adults.

2. Materials and methods

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

2.1. Participants

We evaluated 260 non-demented participants, aged 55–90 years old, from ADNI2 who had all of the following variables available: 1) 3T MRI T1-weighted and fluid attenuated inversion recovery (FLAIR) images, 2) cerebrospinal fluid (CSF) Aβ42 levels (described further in the ADNI methods page http://adni.loni.usc.edu/methods/), and 3) neuropsychological assessment. Both the CSF collection and neuropsychological testing occurred within 18.5 months (average of 3.2 months and 27 days, respectively) of the MRI scan. Four supplemental participants were used for training of our in-house WMH intensity ratio method, and four additional participants were removed after failing FreeSurfer segmentation quality control procedures. Demographic information is tabulated in Table 1. Data analyzed in this study - including MRI scans, CSF amyloid-β_{1-42} (Aβ42) levels, and neuropsychological test scores - were downloaded from the publicly available ADNI Image Data Archive (IDA; https://ida.loni.usc.edu). WMH volumes assessed using one of the five algorithms we evaluated - the intensity histograms algorithm – were also downloaded directly from the ADNI IDA.

2.2. Neuropsychological testing and diagnostic criteria

Participants underwent ADNI baseline neuropsychological testing - including tests of long-term and working memory, language, and executive function - within 3 months of their brain scan. Clinical diagnoses were determined by ADNI as follows: probable AD is assessed according to NINDS/ADRDA criteria (McKhann et al., 1984). However, to minimize the contributions to cognition of neurodegeneration that is specific to AD, our study included only participants with MCI (N = 162) and those who were cognitively intact (N = 98). Participants diagnosed with MCI did not meet the diagnostic criteria for dementia, but did report a

Table 1
Demographic features of the sample analyzed.

Demographic	Aβ-			Aβ+		
	Controls	MCI	Total	Controls	MCI	Total
N	54	89	143	44	73	117
Age (years)	73.07 ± 5.53*	70.00 ± 7.29*	71.16 ± 6.83**	74.68 ± 7.17	73.26 ± 7.64	73.79 ± 7.47**
Sex (M/F)	32/22	46/43	78/65	23/21	40/33	63/54
Education (years) rowhead	16.81 ± 2.60	16.01 ± 2.47	16.31 ± 2.54	16.70 ± 2.47	16.40 ± 2.61	16.51 ± 2.56
ICV (mm^3)	$1.46 \times 10^6 \pm 1.37 \times 10^5$	$1.46 \times 10^6 \pm 1.32 \times 10^5$	$1.46 \times 10^6 \pm 1.33 \times 10^5$	$1.47 \times 10^6 \pm 1.51 \times 10^5$	$1.47 \times 10^6 \pm 1.47 \times 10^5$	$1.47 \times 10^6 \pm 1.48 \times 10^5$

Shown as mean $\pm$ standard deviation. We evaluated group level differences, between amyloid groups (Aβ- vs. Aβ+) and within amyloid groups (control vs. MCI), across age, years of education, and intracranial volume (ICV), using Welch's two-tailed t-tests. We evaluated sex using a χ^2 test.
* Significantly different between controls and MCI within Aβ group, $p < 0.05$.
** Significantly different between Aβ+ and Aβ- participants, $p < 0.05$.

memory complaint. MCI participants had objective memory loss as measured by education-adjusted scores on the Wechsler Memory Scale-Revised - Logical Memory II (WMS-Logical Memory II; Score ≤ 8, 4, or 2 for having completed 16, 8–15, or 0–7 years of education, respectively). They also had a Clinical Dementia Rating (CDR) scores of 0.5 (with a mandatory requirement that CDR memory box score was 0.5 or higher), an absence of significant impairments in other cognitive domains, and preserved daily life activities. Cognitively intact controls did not meet the diagnostic criteria for probable AD or MCI and had no memory complaints. They had a Mini-Mental State Exam (MMSE) score between 24 and 30, a CDR of 0, and scored higher than the education-adjusted MCI thresholds listed above on the Wechsler Memory Scale-Revised - Logical Memory II scores. Participants were excluded if they had a serious neurological condition, neuropsychiatric condition (e.g., major depression, bipolar disorder, schizophrenia), or history of brain injury. We used previously-validated ADNI composite scores for executive function (Gibbons et al., 2012) and memory (Crane et al., 2012). The normalized composite measures of executive function and memory were derived from an iterative process that applied item response theory and confirmatory factory analysis to previously acquired ADNI neuropsychological battery (Crane et al., 2012; Gibbons et al., 2012). The executive function composite score was derived from five clock drawing items (circle, symbol, numbers, hands, and time), Trail Making Test parts A and B, and Category Fluency (animals). The memory composite score was derived from Rey Auditory Verbal Learning Test (RAVLT), AD Assessment Schedule - Cognition (ADAS-Cog), MMSE, and WMS-Logical Memory II.

2.3. MRI scanning

Participants underwent whole-brain MRI scanning on 3-Tesla scanners across 51 sites in North America. Each participant was scanned using an anatomical T1-weighted sequence (1.2 mm thick sagittal slices; 0.9375 × 0.9375 mm^2 in-plane resolution, 256 × 256 matrix) and a T2-weighted fluid attenuated inversion recovery (FLAIR) sequence (5 mm thick axial slices; 0.86 × 0.86 mm^2 in-plane resolution). All MRI acquisition sites passed rigorous scanner validation tests, and the scan protocols were optimized across sites and manufacturers (GE, Philips, Siemens). A GE scanner was used to acquire MRI data on 61 participants across 14 sites, a Philips scanner was used to acquire MRI data on 46 participants across 10 sites, and a Siemens scanner was used to acquire MRI data on 153 participants across 27 sites. Detailed procedures on scan acquisition and optimization are provided elsewhere (www.adni.loni.usc.edu). All T1-weighted and FLAIR images were visually checked for quality. We did not perform bias correction on the FLAIR scans, because 1) our visual quality control assessment did not find extensive FLAIR field inhomogeneities, and 2) a recent analysis (Hernandez et al., 2016) of bias correction performance on FLAIR white matter hyperintensity progression found that applying a bias field correction was not recommended for FLAIR images. Although there are advantages of correcting the magnetic field inhomogeneities seen in FLAIR, that study found that bias field correction in this modality may result in distortion of

real hyperintensities with a specific expense of subtle intensity differences. No T1 or FLAIR image had artifacts that were severe enough to interfere with structural or WMH segmentations. A board-certified neurologist was part of process of reviewing the FLAIR images and WMH algorithm development.

2.4. CSF collection and analysis

Participants underwent at least one lumbar puncture to obtain CSF for assays of several biomarkers. The sample collection and analysis processes are described in Shaw et al. (2009). Aβ+ participants were defined as those who had CSF Aβ42 levels less than 192 pg/ml, consistent with prior guidelines (Shaw et al., 2009).

2.5. In-house WMH algorithm

We developed a semi-automated method to segment white matter hyperintensities using both T1-weighted and FLAIR images (Fig. 1).

2.5.1. Creating white matter masks

White matter masks were used to exclude hyperintensities other than WMHs from our segmentations. To create these masks, we performed bias field correction on the T1-weighted scans using the Advanced Normalization Tools (ANTs) N4 correction (Tustison et al., 2010). We then submitted these bias-corrected images to FreeSurfer (version 5.3) to obtain tissue-segmentation masks and intracranial volume estimation (Fischl et al., 2002). FreeSurfer estimates intracranial volume (ICV) using the known relationship between the ICV and the linear transform of an individual brain to MNI305 template space (Buckner et al., 2004). Because white matter masks produced by FreeSurfer may omit WMHs, we constructed white matter masks by subtracting the gray matter and CSF masks from the full brain mask. Rarely, WMH were extensive enough that they were contiguous with gray matter on the T1-weighted image. When that happened, the intensity values were similar enough that WMHs were included erroneously in the gray matter mask. We therefore visually inspected and manually edited all WM masks to ensure that the gray matter masks did not include WMHs. Each participant's resulting white matter mask was linearly transformed (6 degrees of freedom) to the participant's own FLAIR image using FMRIB's Linear Image Registration Tool (FLIRT) in FSL (Jenkinson and Smith, 2001; Jenkinson et al., 2002). This white matter mask in FLAIR space was then non-linearly transformed to the FLAIR image using the ANTs symmetric image normalization (SyN) method (Avants et al., 2008). We examined and edited the white matter masks as needed in FLAIR space to ensure all white matter (including WMHs) was included.

Next, for each participant, we constructed a mask of the peripheral white matter alone (which is less likely to contain WMHs) to calculate the mean intensity for the WM that does not contain lesions. To do this, first, we eroded a binary whole brain Montreal Neurological Institute (MNI 152) 1 mm template brain mask by 63% (an arbitrary value chosen to provide a mask that excluded peripheral white matter). We non-linearly transformed this eroded template brain into each individual's T1-
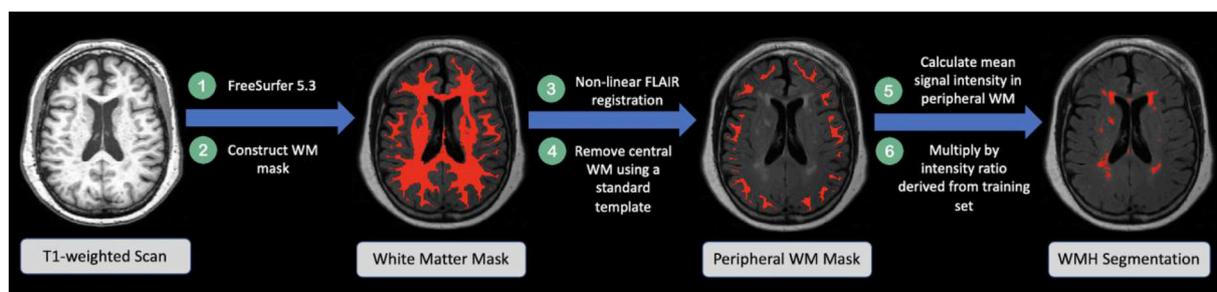

Fig. 1. Flow diagram illustrating the workflow of our method to segment WMH. The intensity ratio is defined as $\frac{\text{Minimum Intensity of WMH}}{\text{Mean WM Intensity without WMH}}$.

weighted image using the ANTs SyN method. The eroded brain masks were then non-linearly transformed into each participant's FLAIR image space using ANTs SyN and were subtracted from the participant's complete white matter mask in FLAIR space to create a mask that contained *only* the brain periphery.

2.5.2. Segmenting WMHs

Our in-house WMH segmentation protocol is illustrated in Fig. 1 and detailed here. First, we created a reference standard segmentation that was visually similar to a manual segmentation in a sub-sample of four training participants who had minimal WMHs on the FLAIR image. We chose participants with minimal WMHs, because for these participants, WMHs were clearly defined and unambiguous, and they contributed minimally to the overall mean white matter intensity for that participant. In our four training participants, we automatically identified WMHs by applying a participant-specific intensity threshold at 99th percentile of the signal intensity in the total white matter for each participant, using the *fslmaths* function in FSL. We arrived at this 99th percentile threshold by visually assessing which threshold adequately segmented these clearly delineated lesions in our test participants. If we had included participants having extensive WMH in this training set, their mean white matter intensity would be low, because the WMHs themselves would reduce the mean signal intensity in the WM mask. Therefore, in participants with extensive WMHs, a 99th percentile intensity threshold would not adequately identify WMHs. Once the WMHs were identified in these four participants, we used their data to calculate a study-specific intensity ratio that could be used to identify high intensity WMHs, even in participants who also have more extensive and diffuse lesions. To calculate a study-specific intensity ratio, across the four training participants, we divided the mean minimum intensity of the WMHs by the mean intensity of the normal-appearing white matter (excluding the WMHs). The voxel, volume, and intensity information derived from the four training participants is tabulated in Table 2. This resulted in a study-specific WMH intensity ratio indicating how much greater the minimum intensity of WMHs was compared with the mean intensity of normal-appearing WM. We then obtained a participant-specific WMH map, by calculating the mean intensity value of each participant's FLAIR image within the peripheral white matter mask (see *2.5.1 Creating white matter masks*) and multiplied it by our study-specific WMH intensity ratio to obtain a threshold, which we then applied to that participant's original FLAIR image.

2.5.3. Regional WMH segmentation

We investigated regional differences in WMH accumulation across three lobes: frontal, temporal, and parietal outlined based on the MNI lobe map atlas from FSL 5.0.7 (maxprob-thr0-1mm). To the extent that the standard lobe map did not cover the entire white matter, we manually extended the lobar gray matter boundaries into the white matter, and visually confirmed that the segmentations were accurate. Fig. 2 depicts before and after we manually extended the lobar gray matter boundaries into the white matter. The lobar masks were registered to each participant's FLAIR space. This allowed us to calculate the WMH volume for the
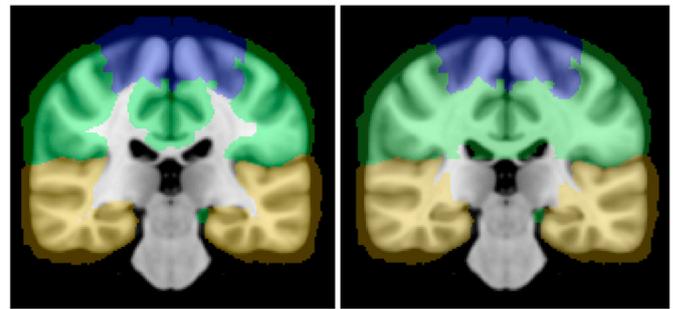
Table 2
Voxel, volume, and intensity information from the participants used to calculate the intensity ratio.

Training Set Participant	WMH Voxels	WMH Volume	WMH minimum intensity	Mean intensity of WM without WMH	Intensity Ratio
1	1425	5262.30	536.65	395.29	1.36
2	1855	6850.22	551.40	404.26	1.36
3	2125	7847.29	579.15	405.49	1.43
4	1966	7260.14	553.34	366.39	1.51
Average	1842.75	6804.99	555.14	392.86	1.42

Mean volume is in mm^3.

Fig. 2. Image on the left depicts the coronal view of the MNI lobe map atlas from FSL 5.0.7 (maxprob-thr0-1mm). The image on the right depicts the lobe map after we manually extended the boundaries of the lobes into the white matter.

frontal, temporal, and parietal lobes.

We also performed an analysis to separate periventricular and deep WMHs. We did this by dilating the ventricle segmentation mask in the participant's FLAIR space by a sphere kernel of 5.16 mm (6 voxels x 0.86 mm resolution). Fig. S1 in the supplemental material illustrates testing performed on the separation of periventricular and deep WMH boundaries by kernel size. We multiplied the dilated ventricle mask by the participant-specific WMH map to construct the periventricular WMH map. The deep WMH volume was calculated by subtracting the participant-specific periventrucular WMH volume from that participant's full WMH volume. We performed quality control on each of the segmentation masks to ensure that the individual mask's boundaries were accurate and the ventricular segmentation did not have artificial enlargement.

2.5.4. Intensity thresholding

We next investigated how varying the inclusiveness of the WMH masks (to include or exclude more diffuse signal surrounding hyperintense lesions) affected the relationship of WMH volume to cognition. To do this we created masks based on different percentages of the constructed intensity ratio. We calculated WMH volumes derived from thresholding at 85%, 90%, 95%, and 105% of the intensity ratio. To determine the new threshold for each percentage we multiplied the percent by the unadjusted intensity ratio and applied the adjusted value to the mean signal intensity in the peripheral mask. Lower threshold percentages provided a more 'lenient' WMH map - that included more diffuse lesions - while higher values included only the highest intensity voxels in the white matter, often associated with more discrete lesions.

2.6. Existing white matter hyperintensity segmentation algorithms

We also evaluated how WMH volumes related to cognition using four WMH algorithms other than our own: 1) an intensity histogram-based algorithm (DeCarli et al., 1995); two algorithms that are part of SPM's lesion segmentation tool (LST): 2) the lesion growth algorithm (LGA) (Schmidt et al., 2012) and 3) the lesion prediction algorithm (LPA) (Schmidt, 2017) (http://www.applied-statistics.de/lst.html); and 4) FSL's brain intensity abnormality classification algorithm (BIANCA) (Griffanti et al., 2016). We used the default settings of each algorithm.

1) The intensity histograms algorithm is the standard method used in ADNI to calculate the WMH volume. This algorithm uses a Bayesian probabilistic method to generate likelihood estimate values for WMH at each voxel in the white matter. These likelihoods are thresholded at three standard deviations above the mean to construct the binary WMH mask. 2) LGA was implemented in the LST toolbox, version 2.0.15 (http://www.statistical-modelling.de/lst.html) for SPM. T1-weighted and FLAIR images were used as inputs. The algorithm selects an initial lesion map and subsequently grows along voxels that are hyperintense relative to surrounding tissue. 3) LPA was implemented in the LST toolbox,

version 2.0.15, for SPM. We used only FLAIR images as input. The algorithm is a binary classifier using a logistic regression model trained on data from 53 participants with severe Multiple Sclerosis (MS). The model covariates include a similar belief map used in the LGA algorithm above and a spatial covariate that accounts for voxel-specific changes in lesion probability. The fitted model parameters are implemented to segment lesions of novel images by estimating the lesion probability across each voxel, outputting a lesion probability map. 4) BIANCA was implemented using FSL. We used a T1-weighted image, FLAIR image, and the same training set as we used for our in-house algorithm. BIANCA classifies each voxel based on intensity and spatial features to output the probability of that voxel being in a WMH. We used BIANCA's default settings and implemented the default probability map threshold of 0.9 (probability of a voxel being a WMH), which historically has optimized the voxel WMH classification false positives and false negative detection rate (Griffanti et al., 2016). We ran BIANCA both with and without including as inputs the same individual WM masks created using our in-house algorithm. Using a WM mask to exclude non-white matter has been shown to reduce false positives (Griffanti et al., 2016).

2.7. Statistics

2.7.1. Amyloid group differences

We stratified the cohort by CSF amyloid level (Aβ-, Aβ+) and evaluated demographic measures both within- and between-amyloid group. In our within-amyloid group analyses, we assessed differences between diagnostic groups (cognitively intact controls or MCI). We used Welch's t-tests to evaluate group differences in age, years of education, and ICV, and a χ^2 test to evaluate group differences in sex.

We covaried for age, sex, years of education, and ICV in all subsequent analyses. Adding scanner manufacturer as a covariate did not modify the relationship between WMH volume and clinical diagnosis and did not significantly contribute to our analyses, so we did not include manufacturer in the statistical models reported throughout the paper (Supplementary Table S1). For all statistical models with a binary dependent variable (such as diagnosis), we performed logistic regression. For all statistical models with a continuous dependent variable (such as cognitive composite scores), we performed multiple linear regression. We used the composite cognitive measures available on the ADNI website. However, when we further evaluated the contribution to our effects of individual neuropsychological subtests within the composite measures, we used Z-score transformed values in our analyses. All statistical analyses were performed in R version 3.5.1 (University of Auckland, Auckland, New Zealand) (R Core Team, 2013).

2.7.2. In-house WMH analysis

We used logistic regression to test our hypothesis that total WMH volume would be more associated with clinical diagnosis in Aβ- participants. To evaluate whether our results were specific to Aβ- participants, we also used logistic regression to relate WMH volume to diagnosis in Aβ+ participants. WMH volume was significantly related to diagnosis in Aβ- participants only. Therefore, all subsequent analyses presented here were performed only in Aβ- participants. Additional analyses performed in Aβ+ participants can be found in the supplementary material (Table S4, Table S5, Table S6, Table S7).

To further investigate our significant results in Aβ- participants we examined whether regional differences in WMH accumulation (in the frontal, temporal, and parietal lobes as well as periventricular (PVWMH) and deep WMH (DWMH; regions across lobes) were associated with clinical diagnosis and executive function and memory. We corrected for multiple comparisons using the false discovery rate (FDR) approach, and report FDR-adjusted p-values (Yekutieli and Benjamini, 1999).

We assessed whether changing the threshold of our in-house WMH algorithm (i.e., including or excluding more diffuse, lower-intensity voxels to WMHs) modified the relationship between WMH volume and clinical diagnosis. To do this, we used logistic regression to test the

association in Aβ- participants between clinical diagnosis and WMH volume calculated using different intensity ratios (at 85%, 90%, 95%, and 105% of the original intensity ratio). We performed additional analyses in segmentation methods that had an available threshold option (LST LGA and BIANCA). Using both the LST LGA method and the BIANCA method, we evaluated the relationship between WMH volume, derived from varying thresholds, and clinical diagnosis in Aβ- participants (Table S8 and Table S9; Fig. S2 and Fig. S3). Because peripheral (deep) WMH may be less intense than periventricular WMH, we further investigated the relationship between deep WMH and diagnosis in Aβ- participants with a more lenient threshold (85% of the intensity ratio) to deep WMH, using our in-house method. We then related deep WMH volume using the 85% intensity threshold to diagnosis in Aβ- participants. This analysis was meant to evaluate possible separate effects of location and intensity of WMHs.

2.7.3. WMH segmentation comparison

To evaluate differences across WMH volumes derived from various algorithms, we related WMH volume (predictor variable) calculated using each of the five segmentation algorithms, to clinical diagnosis (outcome variable), adjusting for age, sex, years of education, and ICV. To test model differences between the WMH segmentation algorithms, we performed a one-way ANOVA with pairwise comparisons, applying FDR to correct for multiple comparisons.

2.7.4. Executive function & memory analysis

To assess whether WMH volume had any cognitive domain-specific effects, we performed multiple linear regression to relate WMH volume, derived from the segmentation algorithm that detected strongest associations, to composite scores of executive function and memory (Crane et al., 2012; Gibbons et al., 2012). To further investigate any significant findings between WMH volume and executive function and memory, we evaluated the relationship between WMH volume and the composite score subtests. In this neuropsychological subtest analysis, we corrected for multiple comparisons by applying false discovery rate (FDR) and reported FDR adjusted p-values (Yekutieli and Benjamini, 1999).

3. Results

3.1. Between and within-amyloid group comparison

In a within-amyloid group analysis, we found that in Aβ- participants, cognitively intact controls were significantly older than those with MCI ($t = 2.850$; $p = 0.005$). In Aβ+ participants, no statistically significant differences were found. When diagnosis was not considered, Aβ+ participants were significantly older than the Aβ- participants ($t = 2.941$; $p = 0.004$; Table 1). We controlled for age in all further analyses along with sex, years of education, and estimated ICV.

3.2. Regional relationship to diagnosis

In Aβ- participants only, higher total WMH volume derived from our in-house algorithm was significantly associated with worse clinical diagnosis ($z = 2.373$, WMH volume partial $p = 0.018$). In this model, higher age ($z = -3.417$, partial $p < 0.001$) and lower educational level ($z = -1.842$, partial $p = 0.065$) were also significantly associated with poorer clinical diagnosis. Additionally, larger frontal, parietal, and periventricular WMH volume were associated with worse clinical diagnosis in Aβ- participants (Table 3). Results from the total WMH and regional analysis can be found in Table 3. In a follow-up analysis in Aβ- participants, we related regional WMH volume to executive function and memory scores. None of the individual regions were related to executive function (Supplementary Table S2) or memory (Supplementary Table S3).

We performed a follow-up analysis to further identify regional

Table 3
Associations between in-house derived WMH volume by region and clinical diagnosis in Aβ- participants.

Region	Controls		MCI		z-value	Partial p-value	FDR adjusted p-value
	Mean volume ± SD	Median (IQR)	Mean volume ± SD	Median (IQR)			
Total	2839 ± 2684	2084 (943–3763)	4379 ± 6609	2145 (1118–4118)	2.373	0.018*	–
Frontal	584 ± 893	262 (119–622)	1389 ± 2639	418 (139–1512)	3.057	0.002*	0.010*
Parietal	739 ± 866	456 (152–944)	1489 ± 3032	356 (149–1146)	2.303	0.021*	0.035*
Temporal	117 ± 135	61 (18–188)	144 ± 200	65 (26–183)	1.448	0.148	0.148
Periventricular	1940 ± 1706	137 (807–2720)	2823 ± 3353	1667 (797–3285)	2.727	0.006*	0.015*
Deep	899 ± 1445	352 (92–972)	1555 ± 3570	383 (124–945)	1.757	0.079•	0.099•

Mean volume is in mm^3. Each relationship was evaluated using a logistic regression, adjusted for age, sex, years of education, and ICV. Multiple comparison correction was applied to frontal, parietal, temporal, periventricular, and deep WMH volume analyses, using FDR adjusted values. Clinical diagnoses: MCI = 1; control = 0. SD = Standard Deviation; IQR = Interquartile Range.
*$p < 0.05$.
•$p < 0.10$, indicating a trend level association.

specificity of periventricular and deep WM effects, and found that larger WMH volumes in the frontal periventricular and parietal periventricular regions were significantly associated with worse clinical diagnosis in Aβ- participants (Table 4).

No significant associations were found between total or regional WMH and clinical diagnosis in Aβ+ participants. All analyses in Aβ+ participants can be found in the supplemental material (Table S4, Table S5, Table S6, Table S7).

3.3. Intensity threshold modification

We next evaluated whether including less intense/more diffuse WMH voxels in the WMH volume measure affected the relationship we saw between WMH volume and clinical diagnosis in Aβ- participants. We did this by adjusting the intensity ratio thresholds used to define WMHs. In Aβ- participants, higher total WMH volume derived from both the unadjusted WMH threshold (i.e., 100%) and the 105% intensity threshold (which further excluded lower-intensity voxels) were significantly associated with poorer clinical diagnosis (Table 5, Fig. 3). Thresholds of 85%, 90%, and 95% of the original WMH threshold included lower intensity voxels characteristic of diffuse lesions; WMH volumes calculated using these more inclusive thresholds were not significantly associated with clinical diagnosis (Table 5). For each intensity threshold we tested, the covariate of older age was significantly associated with poorer clinical diagnosis and lower educational level attained had a trend level association with poorer clinical diagnosis. Modification of thresholds using the LGA and BIANCA methods can be found in the supplementary material (Table S8, Fig. S2, and Table S9, Fig. S3). In a follow-up analysis we investigated whether applying a more lenient threshold to deep WMH resulted in a larger relationship between deep WMH volume and clinical diagnosis. We found that, -when thresholded at 85% of the intensity ratio to allow the inclusion of lower-intensity lesions, deep WMH volume still was not significantly related to diagnosis ($z = -0.307$, $p = 0.759$).

3.4. Clinical associations detected by other WMH algorithms

Within our Aβ- group, we assessed the association between diagnosis and total WMH volume. We did this using six logistic regression analyses, one for each WMH segmentation method: 1) our in-house intensity-based algorithm; 2) a previously-published WMH segmentation based on mathematical modeling of MR pixel intensity histograms (DeCarli et al., 1995): and three methods freely available online - 3) LST - LGA (Schmidt et al., 2012), 4) LST - LPA, (Schmidt, 2017), 5) FSL - BIANCA using an optional WM mask as input, and 6) FSL – BIANCA without using an optional WM mask as input (Griffanti et al., 2016) (Fig. 4).

In non-demented Aβ- participants, greater total WMH volume was significantly associated with MCI diagnosis, calculated using four algorithms: 1) our in-house algorithm ($p = 0.018$), 2) an algorithm based on MR pixel intensity histograms ($p = 0.001$) (DeCarli et al., 1995), 3) LGA ($p < 0.001$) (Schmidt et al., 2012), and 4) BIANCA using the optional WM mask as input ($p = 0.032$) (Griffanti et al., 2016). Total WMH volume was not significantly associated with diagnosis using LPA ($p = 0.086$) or BIANCA without using the optional WM mask as input ($p = 0.088$), using the default options (Table 6).

We performed a one-way ANOVA with pairwise comparisons (Table 7) to determine whether WMH volume was significantly different across algorithms. We found that WMH volumes calculated using LPA were significantly different from WMH volumes using all other methods. Our in-house method, LGA, BIANCA (masked), and the intensity histogram method did not provide WMH volumes that were significantly different from one another.

3.5. Executive function and memory

For simplicity of presentation, in subsequent analyses, we used the WMH segmentation method that produced the strongest association to clinical diagnosis (LGA) to further evaluate the relationship between WMH volume and cognition – specifically, neuropsychological composite

Table 4
Associations between in-house derived WMH volume by sub region and diagnosis in Aβ- participants.

Region	Controls		MCI		z-value	Partial p-value	FDR adjusted p-value
	Mean volume ± SD	Median (IQR)	Mean volume ± SD	Median (IQR)			
Frontal Periventricular	470 ± 725	193 (78–538)	954 ± 1488	351 (69–128)	3.252	0.001*	0.007*
Parietal Periventricular	453 ± 423	373 (144–618)	774 ± 1217	298 (129–787)	2.440	0.015*	0.044*
Temporal Periventricular	93 ± 112	48 (10–141)	105 ± 152	44 (15–138)	1.177	0.239	0.239
Frontal Deep	114 ± 252	34 (8–109)	435 ± 1297	54 (17–211)	1.907	0.057•	0.085•
Parietal Deep	286 ± 593	40 (2–250)	715 ± 1989	39 (2–242)	1.931	0.053•	0.085•
Temporal Deep	25 ± 51	6 (0–27)	39 ± 77	8 (0–44)	1.327	0.184	0.221

Mean volume is in mm^3. Each relationship was evaluated using a logistic regression, adjusted for age, sex, years of education, and ICV. Multiple comparison correction was applied to frontal, parietal, temporal, periventricular, and deep WMH volume analyses, using FDR-adjusted values. Clinical diagnoses: MCI = 1; control = 0. SD = Standard Deviation; IQR = Interquartile Range.
*$p < 0.05$.
•$p < 0.10$, indicating a trend level association.

Table 5

Associations between in-house derived total WMH volume and clinical diagnosis by intensity threshold in Aβ- participants.

| Intensity Threshold Percentage (Ratio Number)	Controls		MCI		z-value	Partial p-value
	Mean volume ± SD	Median (IQR)	Mean volume ± SD	Median (IQR)		
85% (1.207)	18206 ± 13717	14906 (7098–25585)	18524 ± 14935	14959 (8663–21507)	0.904	0.366
90% (1.278)	8830 ± 7714	6990 (3015–11760)	10145 ± 10898	6655 (3931–10841)	1.431	0.152
95% (1.349)	4750 ± 4324	3576 (1598–6191)	6426 ± 8417	3651 (1961–6329)	2.109	0.035
100% (1.42)	2839 ± 2684	2084 (943–3763)	4379 ± 6609	2145 (1118–4118)	2.373	0.018*
105% (1.491)	1799 ± 1831	1245 (563–2395)	3066 ± 5208	1338 (645–2781)	2.388	0.017*

Mean volume is in mm^3. Each relationship was evaluated using a logistic regression, adjusted for age, sex, years of education, and ICV. Clinical diagnoses were coded as MCI = 1; control = 0. SD = Standard Deviation; IQR = Interquartile Range.
*$p < 0.05$.
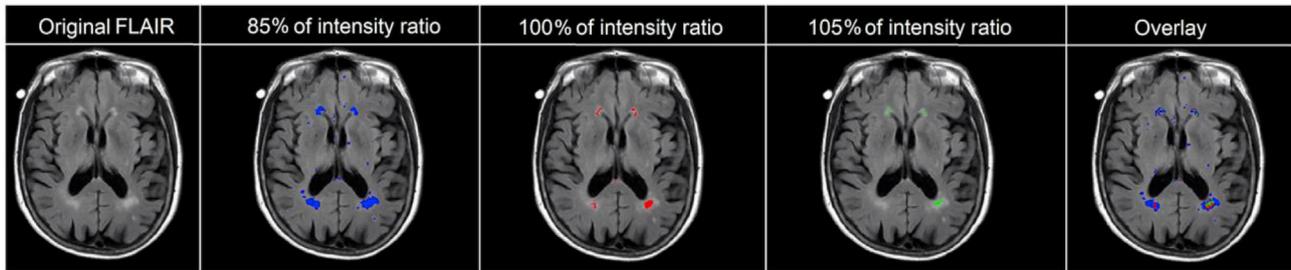
Fig. 3. WMH boundary segmentation based on varying intensity thresholds of the study-specific intensity ratio. The far-right image illustrates the 85%, 100%, and 105% threshold masks all overlaid on the base FLAIR image for comparison purposes.
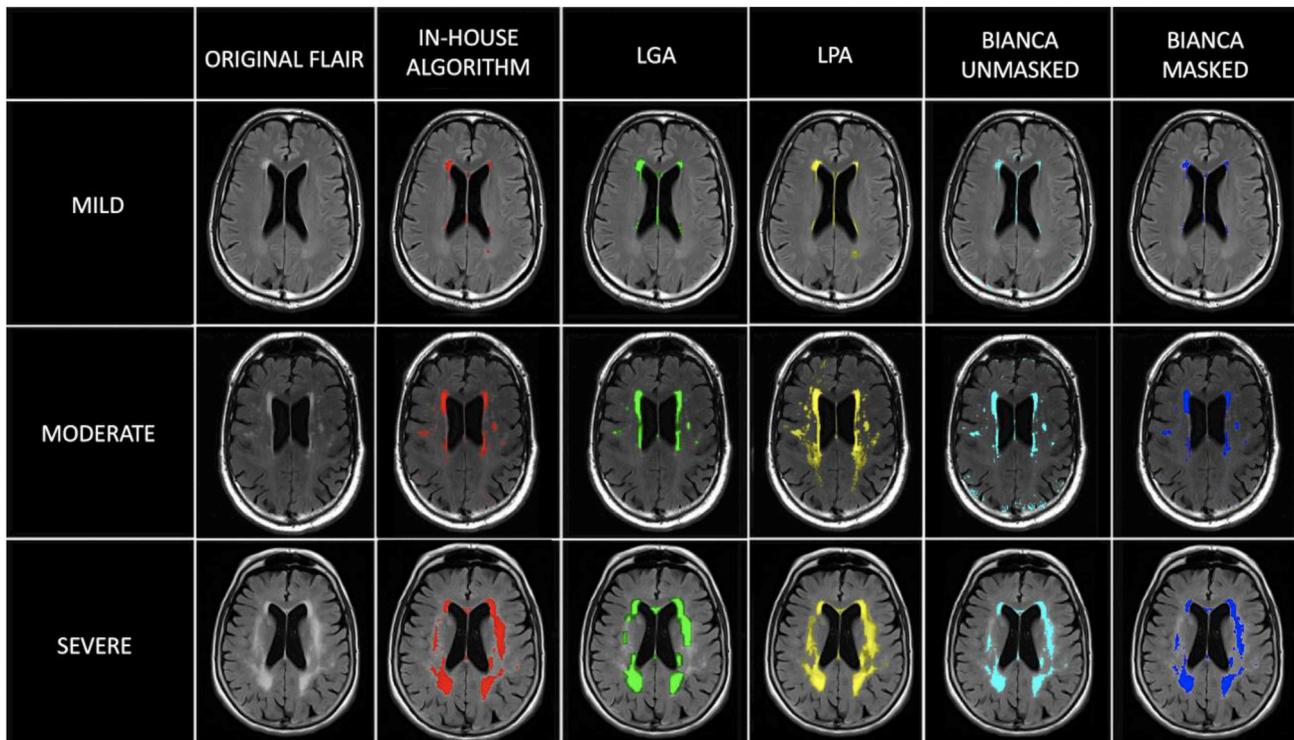
Fig. 4. Range of WMH severity and variation in white matter segmentation methods. The severity was evaluated as WMH volume corrected for ICV. We defined mild WMH volume in a participant, when the individual's total WMH volume was less than the mean total WMH volume across participants. Moderate WMH volume was defined as the individual's total WMH volume being between the mean and two standard deviations above the mean across participants, and severe WMH volume when the individual's total WMH volume was greater than two standard deviations above the mean across participants. For BIANCA, "masked" indicates that the same WM mask generated for our in-house algorithm was used as input for the analysis.

measures of memory or executive function, although it is important to note that LGA WMH volumes were not significantly different from those calculated using the in-house, BIANCA (masked), or intensity histogram algorithms, all of which were significantly associated with diagnosis (Table 7). Using multiple linear regression, we found that greater LGA-derived WMH volumes were significantly associated with lower executive function composite scores (omnibus $p < 0.001$; WMH volume partial $t = 2.33$; $p = 0.021$). LGA-derived WMH volumes were not significantly

Table 6
Associations between WMH volume by segmentation method and clinical diagnosis in Aβ- participants.

WMH Segmentation Method	Controls		MCI		z-value	Partial p-value
	Mean volume ± SD	Median (IQR)	Mean volume ± SD	Median (IQR)		
In-house	2839 ± 2684	2084 (943, 3763)	4379 ± 6609	2145 (1118–4118)	2.373	0.018*
Intensity histograms	3133 ± 2788	2340 (1359–4076)	5955 ± 8686	2713 (1243–6664)	3.231	0.001*
LGA	2658 ± 3478	1525 (575–3076)	5457 ± 8473	1864 (315–6903)	3.533	< 0.001*
LPA	20163 ± 19443	12338 (6713–28781)	20617 ± 21128	11772 (5057–30651)	1.715	0.086•
BIANCA (masked)	4272 ± 4063	2404 (1494–6386)	5442 ± 7020	2705 (1257–6564)	2.145	0.032*
BIANCA (unmasked)	14165 ± 5568	13610 (10209-17196)	16147 ± 8050	15285 (9845–20916)	1.705	0.088•

Mean volume is in mm3. Each relationship was evaluated using a logistic regression, adjusted for age, sex, years of education, and ICV. For BIANCA, "masked" indicates that a WM mask was used as input for the analysis. Clinical diagnoses: MCI = 1; control = 0. SD = Standard Deviation; IQR = Interquartile Range.
*p < 0.05.
• p < 0.10, indicating a trend level association.

Table 7
One-way ANOVA with pairwise comparisons.

	In-House	Intensity Histogram	LGA	LPA
Intensity Histogram	p = 0.66	–	–	–
LGA	p = 0.78	p = 0.78	–	–
LPA	p < 0.001*	p < 0.001*	p < 0.001*	–
BIANCA (masked)	p = 0.66	p = 0.93	p = 0.78	p < 0.001*

p-values displayed are corrected for multiple comparisons using the false discovery rate (FDR). For BIANCA, "masked" indicates that a WM mask was used as input for the analysis.
*p < 0.05.

correlated with composite memory scores (omnibus $p < 0.001$; WMH volume partial $t = 1.629$; $p = 0.106$).

We further investigated our significant result to determine whether certain neuropsychological subtests may be driving the association between LGA-derived WMH volumes and executive function. We found that greater WMH volume was significantly associated with lower Category Fluency score (omnibus $p < 0.001$; WMH volume partial $t = 3.12$; FDR adjusted $p = 0.013$). LGA-derived WMH volumes were not associated with Trail Making Test Part A (omnibus $p < 0.001$; WMH volume partial $t = 1.96$; FDR corrected $p = 0.157$) or Part B (omnibus $p < 0.001$; WMH volume partial $t = 0.79$; FDR adjusted $p = 0.516$), or any of the three clock drawing subscores: symbol (omnibus $p < 0.001$; WMH volume partial $t = 1.389$; FDR adjusted $p = 0.516$); numbers (WMH volume partial $t = 1.389$; FDR adjusted $p = 0.330$); or time (omnibus $p < 0.001$; WMH volume partial $t = 0.151$; FDR adjusted $p = 0.880$). We did not examine the relationship between WMH volume and the clock drawing circle or hand scores test, as there was ceiling effect on these tests – on the clock drawing circle subtest, all 143 of the Aβ- participants received a perfect score and on the hand subtest, 142 of the 143 participants received a perfect score.

4. Discussion

We investigated, in a sample of non-demented Aβ- older adults, the most clinically relevant features of WMH boundary selection, by relating WMH volume (using 5 different algorithms) to clinical diagnosis and cognitive function. We found that 1) larger total, frontal, parietal, and periventricular WMH volumes, derived from our in-house algorithm, were significantly associated with a worse clinical diagnosis in Aβ- older adults, 2) limiting WMH boundaries to voxels having the highest-intensity thresholds strengthened the relationship between WMH volume and clinical diagnosis, and 3) the most clinically relevant WMH segmentation algorithms (LGA, intensity histogram, our in-house method, and BIANCA with the WM mask option) were methods that limited boundary selection to the most high-intensity areas of the WMHs.

Our study is the first to compare multiple WMH segmentation methods using clinical diagnosis and cognitive composite measures to assess clinical relevance.

We found a significant effect between WMH volume and diagnosis only in Aβ- participants. Here, Aβ positivity may reflect AD-related neuropathological changes more broadly, which may be associated with cognitive effects, even in cognitively intact older adults (Braskie et al., 2010; Amariglio et al., 2012; Ho and Nation, 2018). Variability from such AD-related effects on cognition could add statistical noise to WMH-related cognitive effects in Aβ+ participants, making those effects harder to detect. The relationship between cerebral amyloidosis and WMHs is currently debated (Roseborough et al., 2017), although emerging evidence indicates that Aβ and WMHs may have both independent and interactive effects (Scott et al., 2016; Schreiner et al., 2018). WMH accumulation in Aβ- participants may represent an increased vulnerability to developing abnormal levels of Aβ later, although future longitudinal studies are needed to clarify this possibility. Additionally, the interaction between amyloid and WMHs may also make it more difficult to detect an effect on cognition that is specifically attributable to WMHs in Aβ+ participants. It is possible that Aβ may interact with the effect of WMH accumulation on cognition differently when evaluating cohorts with a broader range of diagnoses, such as those with symptomatic AD (Provenzano et al., 2013). However, within our cohort of non-demented older adults, the effect of subtle increases in WMH volume on cognition was only detectable in individuals without abnormal amyloid levels, suggesting that AD-specific processes may mask the effect of WMH accumulation before clinical onset of AD. These findings are consistent with our hypothesis that the relationship between WMH volume and clinical diagnosis is dependent upon both WMH boundary selection and amyloid-positivity status.

We found that larger total, frontal, and parietal WMH volumes were significantly associated with worse clinical diagnosis, while WMH in the temporal lobe were not. We also found that periventricular, but not deep WMH were significantly associated with clinical diagnosis, which is consistent with past findings relating periventricular WMH to global cognition (Kim et al., 2008; Bolandzadeh et al., 2012; Griffanti et al., 2018). However, our study found trend level significance when relating deep WMH to clinical measures. The ability to detect a robust effect could be a result of small discrete lesions having a different intensity distribution from larger lesions, with deep WMHs appearing lighter than periventricular WMHs. Therefore, deep WMHs could be under-segmented by various segmentation methods, resulting in periventricular WMHs appearing to have a stronger relationship to clinical variables than deep WMHs. To test this, we performed an additional test in which we used a more lenient threshold to segment less hyperintense deep WMHs. We found that when we segmented the lighter regions of deep WMHs, the relationship to clinical diagnosis became weaker, suggesting that both the location and intensity of lesions are important to clinical relevance. Periventricular and deep WMHs are both associated with severe myelin loss and increased microglia activity (Simpson et al., 2007), but may also have different etiological origins. Periventricular

WMHs may be related to arterial pressure, plasma leakage, blood brain barrier permeability, and decline in total cerebral blood flow, while deep WMHs may be associated with axonal loss, arteriolosclerosis, and body mass index (ten Dam et al., 2007; Haller et al., 2013; Wharton et al., 2015; Griffanti et al., 2018). Although future work is needed to illuminate the mechanisms of these findings, our work suggests that disruption of global cognitive processes may be related to region-specific changes.

We found that WMH boundary selection was an important algorithm feature that modified the degree to which logistic regression could capture the relationship between WMH volume and clinical diagnosis. The WMH volumes calculated such that only the most hyperintense voxels were included, were best associated with clinical diagnosis. We further validated our intensity threshold findings by determining that the WMH segmentation methods most associated with diagnosis (using default settings) were the four algorithms that most limited WMHs to highest intensity voxels (LGA, in-house method, the intensity histogram method, and BIANCA with the WM mask included as an input). The segmentations that were significantly associated with diagnosis provided smaller WMH volumes for the same scans. A visual review suggests that these segmentations captured the most discrete and highly intense regions (Fig. 4). Additionally, when we tested different thresholds for selecting voxels using the three algorithms that allowed such adjustments, the thresholds that resulted in smaller WMH volumes composed of the most intense voxels were most closely related to clinical diagnosis. Our results suggest that optimization of algorithm parameters to capture the most intense WMH voxels will yield more robust classification results relevant to clinical diagnosis.

Using only default options, the WMH volumes that resulted in the largest volumes were derived from LPA and BIANCA without the WM mask and were not significantly associated with clinical diagnosis. LPA included lighter and more diffuse hyperintense regions in addition to brighter, more discrete lesions. Using additional optional parameters for LPA may have produced significant associations with diagnosis. BIANCA without the optional WM mask provided WMH estimates that appeared visually similar to the less inclusive in-house, histogram, and LGA methods, but also included some non-white matter hyperintense regions, such as in the cortex, cerebellum, and brainstem regions. Applying the BIANCA option to input a white matter mask prevented the inclusion of erroneous non-WM voxels in the WMH map, and WMH volume estimated using BIANCA with a WM mask as input was significantly associated with clinical diagnosis. We implemented each algorithm using the default parameters to provide varying segmentation results among segmentation methods allowing us to better investigate which WMH characteristics were most clinically relevant. Our purpose was not to recommend any one software package over another. Rather, our findings highlight the importance of limiting the WMH search to white matter regions and segmenting only the most hyperintense voxels, regardless of the algorithm used.

In follow-up analyses of Aβ- older adults, greater WMH volumes were associated with lower executive function composite scores. Although previous literature relating WMH volume to cognitive function is variable (Prins and Scheltens, 2015), mounting evidence demonstrates that WMH volume has both broad and specific effects on cognitive function (Hedden et al., 2012; Kloppenborg et al., 2014). Globally, WMHs have been associated with future cognitive decline (Boyle et al., 2016), impacting multiple neuropsychological domains (Gunning-Dixon and Raz, 2000; Au et al., 2006). However, WMHs most consistently have been associated deficits in processing speed and executive function (Debette et al., 2010; Murray et al., 2010; Kloppenborg et al., 2014; Lampe et al., 2017), consistent with our current findings.

The association we found between WMH volume and executive function was driven primarily by deficits in category fluency – a type of verbal fluency test that here involves freely generating as many animal names as possible within a set time period. Category fluency, a sensitive marker for cognitive impairment, is impacted by frontal lobe WMH accumulation (Gootjes et al., 2004), as it recruits both frontal and

temporal lobe brain regions (Mummery et al., 1996; Gootjes et al., 2004; Baldo et al., 2006; Peter et al., 2016). Although temporal lobe WMH volume was not associated with category fluency measures, the effect of WMH volume on category fluency may be attributed to disruption of more global structural cortical connections, such as between the frontal and temporal lobes (Wiseman et al., 2018).

Our study included data only from non-demented older participants, and therefore may not be generalizable to participants with Alzheimer's disease and other diseases affecting the white matter, such as multiple sclerosis. We related total and regional WMH volume to cognition in our study, but did not evaluate how the number of total or regional WMHs in each brain related to cognition. Such an analysis would be interesting and may yield different results. Use of intensity threshold to calculate WMHs, as in our method, may differently capture small and large lesions, as small discrete lesions may have a different intensity distribution from larger more diffuse lesions. We used a visual quality control assessment of each WM mask, and manual editing as needed for accuracy. Because our WM mask was created by subtracting the gray matter and CSF masks from the whole brain on the T1-weighted images, editing was only required when the WMH and gray matter, which have similar intensities on the T1-weighted images, were contiguous, in which case, the automatic segmentation may include WMH erroneously in the gray matter mask. This was not a common occurrence in the ADNI cohort, whose participants do not tend to have extensive WM pathology in the periphery. However, in a cohort that includes many participants with very extensive WMH pathology, this manual editing step may be more time consuming. Additionally, we used a limited set of open-source WMH toolboxes which may not capture all the possible variability of WMH boundary segmentations. These automated WMH segmentation methods have optional parameters that use different variations of location and intensity as inputs into either a linear or nonlinear classifier. We used the default settings on the various packages in order to arrive at variable segmentations, but optimization of these parameters may have resulted in significant associations between the WMH segmentation volumes and clinical diagnosis. Our intent here was not to evaluate the software packages *per se,* but to determine what type of segmentation would be most clinically relevant. Our converging results suggest that multiple algorithms may generate useful segmentations.

Overall, our study sought to systematically assess automatic WMH segmentations to identify the most clinically meaningful results. Our findings suggest that WMH segmentations that exclude the lightest and most diffuse hyperintensities have the strongest clinical relevance and that this relationship is most evident only in Aβ- older adults. This suggests that AD-specific processes, such as amyloid accumulation, may mask the cognitive consequences of WMHs. However, evaluation of higher intensity WMH volumes is a useful metric to classify global cognitive function and assess domain-specific changes in executive function in older adults. Our work is an initial step toward harmonizing WMH segmentation protocols, allowing for more robust and reliable investigations on how WMHs mechanistically relate to cognition and sub-optimal brain aging.

Declaration of competing interest

The authors do not have financial interests or conflicts of interest related to the topic of the paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.116327.

References

Aljondi, R., Szoeke, C., Steward, C., Gorelik, A., Desmond, P., 2018. The effect of midlife cardiovascular risk factors on white matter hyperintensity volume and cognition two decades later in normal ageing women. Brain Imag. Behav.

Amariglio, R.E., Becker, J.A., Carmasin, J., Wadsworth, L.P., Lorius, N., Sullivan, C., Maye, J.E., Gidicsin, C., Pepin, L.C., Sperling, R.A., Johnson, K.A., Rentz, D.M., 2012. Participantive cognitive complaints and amyloid burden in cognitively normal older individuals. Neuropsychologia 50, 2880–2886.

Au, R., Massaro, J.M., Wolf, P.A., Young, M.E., Beiser, A., Seshadri, S., D'Agostino, R.B., DeCarli, C., 2006. Association of white matter hyperintensity volume with decreased cognitive functioning: the Framingham Heart Study. Arch. Neurol. 63, 246–250.

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12, 26–41.

Baldo, J.V., Schwartz, S., Wilkins, D., Dronkers, N.F., 2006. Role of frontal versus temporal cortex in verbal fluency as revealed by voxel-based lesion symptom mapping. J. Int. Neuropsychol. Soc. 12, 896–900.

Bolandzadeh, N., Davis, J.C., Tam, R., Handy, T.C., Liu-Ambrose, T., 2012. The association between cognitive function and white matter lesion location in older adults: a systematic review. BMC Neurol. 12, 126.

Boyle, P.A., Yu, L., Fleischman, D.A., Leurgans, S., Yang, J., Wilson, R.S., Schneider, J.A., Arvanitakis, Z., Arfanakis, K., Bennett, D.A., 2016. White matter hyperintensities, incident mild cognitive impairment, and cognitive decline in old age. Ann. Clin. Transl. Neurol. 3, 791–800.

Braskie, M.N., Klunder, A.D., Hayashi, K.M., Protas, H., Kepe, V., Miller, K.J., Huang, S.C., Barrio, J.R., Ercoli, L.M., Siddarth, P., Satyamurthy, N., Liu, J., Toga, A.W., Bookheimer, S.Y., Small, G.W., Thompson, P.M., 2010. Plaque and tangle imaging and cognition in normal aging and Alzheimer's disease. Neurobiol. Aging 31, 1669–1678.

Brickman, A.M., Muraskin, J., Zimmerman, M.E., 2009. Structural neuroimaging in Alzheimer's disease: do white matter hyperintensities matter? Dialogues Clin. Neurosci. 11, 181–190.

Buckner, R.L., Head, D., Parker, J., Fotenos, A.F., Marcus, D., Morris, J.C., Snyder, A.Z., 2004. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage 23, 724–738.

Caligiuri, M.E., Perrotta, P., Augimeri, A., Rocca, F., Quattrone, A., Cherubini, A., 2015. Automatic detection of white matter hyperintensities in healthy aging and pathology using magnetic resonance imaging: a review. Neuroinformatics 13, 261–276.

Crane, P.K., Carle, A., Gibbons, L.E., Insel, P., Mackin, R.S., Gross, A., Jones, R.N., Mukherjee, S., Curtis, S.M., Harvey, D., Weiner, M., Mungas, D., Alzheimer's Disease Neuroimaging Initiative, 2012. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). Brain Imag. Behav. 6, 502–516.

Dadar, M., Maranzano, J., Misquitta, K., Anor, C.J., Fonov, V.S., Tartaglia, M.C., Carmichael, O.T., Decarli, C., Collins, D.L., 2017. Performance comparison of 10 different classification techniques in segmenting white matter hyperintensities in aging. Neuroimage 157, 233–249.

de Groot, J.C., de Leeuw, F.E., Oudkerk, M., van Gijn, J., Hofman, A., Jolles, J., Breteler, M.M., 2000. Cerebral white matter lesions and cognitive function: the Rotterdam Scan Study. Ann. Neurol. 47, 145–151.

Debette, S., Beiser, A., Hoffmann, U., Decarli, C., O'Donnell, C.J., Massaro, J.M., Au, R., Himali, J.J., Wolf, P.A., Fox, C.S., Seshadri, S., 2010. Visceral fat is associated with lower brain volume in healthy middle-aged adults. Ann. Neurol. 68, 136–144.

DeCarli, C., Murphy, D.G., Tranh, M., Grady, C.L., Haxby, J.V., Gillette, J.A., Salerno, J.A., Gonzales-Aviles, A., Horwitz, B., Rapoport, S.I., et al., 1995. The effect of white matter hyperintensity volume on brain structure, cognitive performance, and cerebral metabolism of glucose in 51 healthy adults. Neurology 45, 2077–2084.

Firbank, M.J., Lloyd, A.J., Ferrier, N., O'Brien, J.T., 2004. A volumetric study of MRI signal hyperintensities in late-life depression. Am. J. Geriatr. Psychiatry 12, 606–612.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.

Frisoni, G.B., Galluzzi, S., Pantoni, L., Filippi, M., 2007. The effect of white matter lesions on cognition in the elderly–small but detectable. Nat. Clin. Pract. Neurol. 3, 620–627.

Gibbons, L.E., Carle, A.C., Mackin, R.S., Harvey, D., Mukherjee, S., Insel, P., Curtis, S.M., Mungas, D., Crane, P.K., Alzheimer's Disease Neuroimaging Iniative, 2012. A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. Brain Imag. Behav. 6, 517–527.

Gibson, E., Gao, F., Black, S.E., Lobaugh, N.J., 2010. Automatic segmentation of white matter hyperintensities in the elderly using FLAIR images at 3T. J. Magn. Reson. Imaging 31, 1311–1322.

Gootjes, L., Teipel, S.J., Zebuhr, Y., Schwarz, R., Leinsinger, G., Scheltens, P., Moller, H.J., Hampel, H., 2004. Regional distribution of white matter hyperintensities in vascular dementia, Alzheimer's disease and healthy aging. Dement. Geriatr. Cognit. Disord. 18, 180–188.

Gordon, B.A., Najmi, S., Hsu, P., Roe, C.M., Morris, J.C., Benzinger, T.L., 2015. The effects of white matter hyperintensities and amyloid deposition on Alzheimer dementia. Neuroimage Clin. 8, 246–252.

Griffanti, L., Zamboni, G., Khan, A., Li, L., Bonifacio, G., Sundaresan, V., Schulz, U.G., Kuker, W., Battaglini, M., Rothwell, P.M., Jenkinson, M., 2016. BIANCA (Brain Intensity AbNormality Classification Algorithm): a new tool for automated segmentation of white matter hyperintensities. Neuroimage 141, 191–205.

Griffanti, L., Jenkinson, M., Suri, S., Zsoldos, E., Mahmood, A., Filippini, N., Sexton, C.E., Topiwala, A., Allan, C., Kivimaki, M., Singh-Manoux, A., Ebmeier, K.P., Mackay, C.E., Zamboni, G., 2018. Classification and characterization of periventricular and deep white matter hyperintensities on MRI: a study in older adults. Neuroimage 170, 174–181.

Grimaud, J., Lai, M., Thorpe, J., Adeleine, P., Wang, L., Barker, G.J., Plummer, D.L., Tofts, P.S., McDonald, W.I., Miller, D.H., 1996. Quantification of MRI lesion load in multiple sclerosis: a comparison of three computer-assisted techniques. Magn. Reson. Imaging 14, 495–505.

Grimmer, T., Faust, M., Auer, F., Alexopoulos, P., Forstl, H., Henriksen, G., Perneczky, R., Sorg, C., Yousefi, B.H., Drzezga, A., Kurz, A., 2012. White matter hyperintensities predict amyloid increase in Alzheimer's disease. Neurobiol. Aging 33, 2766–2773.

Gunning-Dixon, F.M., Raz, N., 2000. The cognitive correlates of white matter abnormalities in normal aging: a quantitative review. Neuropsychology 14, 224–232.

Haller, S., Kovari, E., Herrmann, F.R., Cuvinciuc, V., Tomm, A.M., Zulian, G.B., Lovblad, K.O., Giannakopoulos, P., Bouras, C., 2013. Do brain T2/FLAIR white matter hyperintensities correspond to myelin loss in normal aging? A radiologic-neuropathologic correlation study. Acta Neuropathol. Commun. 1, 14.

Hane, F.T., Robinson, M., Lee, B.Y., Bai, O., Leonenko, Z., Albert, M.S., 2017. Recent progress in Alzheimer's disease Research, Part 3: diagnosis and treatment. J. Alzheimer's Dis. 57, 645–665.

Hedden, T., Mormino, E.C., Amariglio, R.E., Younger, A.P., Schultz, A.P., Becker, J.A., Buckner, R.L., Johnson, K.A., Sperling, R.A., Rentz, D.M., 2012. Cognitive profile of amyloid burden and white matter hyperintensities in cognitively normal older adults. J. Neurosci. 32, 16233–16242.

Hernandez, V., Gozalez-Castro, V., Ghandour, D.T., Wang, X., Doubal, F., Munoz Maniega, S., Armitage, P.A., Wardlaw, J.M., 2016. On the computational assessment of white matter hyperintensity progression: difficulties in method selection and bias field correction performance on images with significant white matter pathology. Neuroradiology 58, 475–485.

Ho, J.K., Nation, D.A., 2018. Neuropsychological profiles and trajectories in preclinical Alzheimer's disease. J. Int. Neuropsychol. Soc. 24, 693–702.

Iorio, M., Spalletta, G., Chiapponi, C., Luccichenti, G., Cacciari, C., Orfei, M.D., Caltagirone, C., Piras, F., 2013. White matter hyperintensities segmentation: a new semi-automated method. Front. Aging Neurosci. 5, 76.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5, 143–156.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841.

Kapeller, P., Barber, R., Vermeulen, R.J., Ader, H., Scheltens, P., Freidl, W., Almkvist, O., Moretti, M., del Ser, T., Vaghfeldt, P., Enzinger, C., Barkhof, F., Inzitari, D., Erkinjunti, T., Schmidt, R., Fazekas, F., 2003. Visual rating of age-related white matter changes on magnetic resonance imaging: scale comparison, interrater agreement, and correlations with quantitative measurements. Stroke 34, 441–445.

Kim, K.W., MacFall, J.R., Payne, M.E., 2008. Classification of white matter lesions on magnetic resonance imaging in elderly persons. Biol. Psychiatry 64, 273–280.

Kloppenborg, R.P., Nederkoorn, P.J., Geerlings, M.I., van den Berg, E., 2014. Presence and progression of white matter hyperintensities and cognition: a meta-analysis. Neurology 82, 2127–2138.

Lampe, L., Kharabian-Masouleh, S., Kynast, J., Arelin, K., Steele, C.J., Loffler, M., Witte, A.V., Schroeter, M.L., Villringer, A., Bazin, P.L., 2017. Lesion location matters: the relationships between white matter hyperintensities on cognition in the healthy elderly. J. Cereb. Blood Flow Metab. 39, 36–43.

Mantyla, R., Erkinjuntti, T., Salonen, O., Aronen, H.J., Peltonen, T., Pohjasvaara, T., Standertskjold-Nordenstam, C.G., 1997. Variable agreement between visual rating scales for white matter hyperintensities on MRI. Comparison of 13 rating scales in a poststroke cohort. Stroke 28, 1614–1623.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E., 1984. Clinical diagnosis of Alzheimer's disease. Neurology 34, 939–944.

Mortamais, M., Ash, J.A., Harrison, J., Kaye, J., Kramer, J., Randolph, C., Pose, C., Albala, B., Ropacki, M., Ritchie, C.W., Ritchie, K., 2017. Detecting cognitive changes in preclinical Alzheimer's disease: a review of its feasibility. Alzheimers Dement 13, 468–492.

Mummery, C.J., Patterson, K., Hodges, J.R., Wise, R.J., 1996. Generating 'tiger' as an animal name or a word beginning with T: differences in brain activation. Proc. Biol. Sci. 263, 989–995.

Murray, M.E., Senjem, M.L., Petersen, R.C., Hollman, J.H., Preboske, G.M., Weigand, S.D., Knopman, D.S., Ferman, T.J., Dickson, D.W., Jack Jr., C.R., 2010. Functional impact of white matter hyperintensities in cognitively normal elderly participants. Arch. Neurol. 67, 1379–1385.

Peter, J., Kaiser, J., Landerer, V., Kostering, L., Kaller, C.P., Heimbach, B., Hull, M., Bormann, T., Kloppel, S., 2016. Category and design fluency in mild cognitive impairment: performance, strategy use, and neural correlates. Neuropsychologia 93, 21–29.

Prins, N.D., Scheltens, P., 2015. White matter hyperintensities, cognitive impairment and dementia: an update. Nat. Rev. Neurol. 11, 157–165.

Prins, N.D., van Straaten, E.C., van Dijk, E.J., Simoni, M., van Schijndel, R.A., Vrooman, H.A., Koudstaal, P.J., Scheltens, P., Breteler, M.M., Barkhof, F., 2004. Measuring progression of cerebral white matter lesions on MRI: visual rating and volumetrics. Neurology 62, 1533–1539.

Provenzano, F.A., Muraskin, J., Tosto, G., Narkhede, A., Wasserman, B.T., Griffith, E.Y., Guzman, V.A., Meier, I.B., Zimmerman, M.E., Brickman, A.M., 2013. White matter hyperintensities and cerebral amyloidosis: necessary and sufficient for clinical expression of Alzheimer disease? JAMA Neurol. 70, 455–461.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Roseborough, A., Ramirez, J., Black, S.E., Edwards, J.D., 2017. Associations between amyloid beta and white matter hyperintensities: a systematic review. Alzheimers Dement 13, 1154–1167.

Schmidt, P., 2017. Bayesian Inference for Structured Additive Regression Models for Large-Scale Problems with Applications to Medical Imaging. PhD thesis. LudwigMaximilians-Universität München.

Schmidt, P., Gaser, C., Arsic, M., Buck, D., Forschler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V.J., Zimmer, C., Hemmer, B., Muhlau, M., 2012. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. Neuroimage 59, 3774–3783.

Schreiner, S.J., Kirchner, T., Narkhede, A., Wyss, M., Van Bergen, J.M.G., Steininger, S.C., Gietl, A., Leh, S.E., Treyer, V., Buck, A., Pruessmann, K.P., Nitsch, R.M., Hock, C., Henning, A., Brickman, A.M., Unschuld, P.G., 2018. Brain amyloid burden and cerebrovascular disease are synergistically associated with neurometabolism in cognitively unimpaired older adults. Neurobiol. Aging 63, 152–161.

Scott, J.A., Braskie, M.N., Tosun, D., Thompson, P.M., Weiner, M., DeCarli, C., Carmichael, O.T., Alzheimer's Disease Neuroimaging I, 2015. Cerebral amyloid and hypertension are independently associated with white matter lesions in elderly. Front. Aging Neurosci. 7, 221.

Scott, J.A., Braskie, M.N., Tosun, D., Maillard, P., Thompson, P.M., Weiner, M., DeCarli, C., Carmichael, O.T., 2016. Cerebral amyloid is associated with greater white-matter hyperintensity accrual in cognitively normal older adults. Neurobiol. Aging 48, 48–52.

Shaw, L.M., Vanderstichele, H., Knapik-Czajka, M., Clark, C.M., Aisen, P.S., Petersen, R.C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W., Lee, V.M., Trojanowski, J.Q., 2009. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative participants. Ann. Neurol. 65, 403–413.

Simpson, J.E., Fernando, M.S., Clark, L., Ince, P.G., Matthews, F., Forster, G., O'Brien, J.T., Barber, R., Kalaria, R.N., Brayne, C., Shaw, P.J., Lewis, C.E., Wharton, S.B., 2007. White matter lesions in an unselected cohort of the elderly: astrocytic, microglial and oligodendrocyte precursor cell responses. Neuropathol. Appl. Neurobiol. 33, 410–419.

Smart, S.D., Firbank, M.J., O'Brien, J.T., 2011. Validation of automated white matter hyperintensity segmentation. J. Aging Res. 391783, 2011.

Smith, E.E., Salat, D.H., Jeng, J., McCreary, C.R., Fischl, B., Schmahmann, J.D., Dickerson, B.C., Viswanathan, A., Albert, M.S., Blacker, D., Greenberg, S.M., 2011. Correlations between MRI white matter lesion location and executive function and episodic memory. Neurology 76, 1492–1499.

ten Dam, V.H., van den Heuvel, D.M., de Craen, A.J., Bollen, E.L., Murray, H.M., Westendorp, R.G., Blauw, G.J., van Buchem, M.A., 2007. Decline in total cerebral blood flow is linked with increase in periventricular but not deep white matter hyperintensities. Radiology 243, 198–203.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29, 1310–1320.

van Straaten, E.C., Fazekas, F., Rostrup, E., Scheltens, P., Schmidt, R., Pantoni, L., Inzitari, D., Waldemar, G., Erkinjuntti, T., Mantyla, R., Wahlund, L.O., Barkhof, F., 2006. Impact of white matter hyperintensities scoring method on correlations with clinical data: the LADIS study. Stroke 37, 836–840.

Wang, R., Li, C., Wang, J., Wei, X., Li, Y., Zhu, Y., Zhang, S., 2015. Automatic segmentation and volumetric quantification of white matter hyperintensities on fluid-attenuated inversion recovery images using the extreme value distribution. Neuroradiology 57, 307–320.

Wharton, S.B., Simpson, J.E., Brayne, C., Ince, P.G., 2015. Age-associated white matter lesions: the MRC cognitive function and ageing study. Brain Pathol. 25, 35–43.

Wilcock, G.K., Esiri, M.M., 1982. Plaques, tangles and dementia. A quantitative study. J. Neurol. Sci. 56, 343–356.

Wiseman, S.J., Booth, T., Ritchie, S.J., Cox, S.R., Muñoz Maniega, S., Valdés Hernández, M.D., Dickie, D.A., Royle, N.A., Starr, J.M., Deary, I.J., Wardlaw, J.M., Bastin, M.E., 2018. Cognitive abilities, brain white matter hyperintensity volume, and structural network connectivity in older age. Hum. Brain Mapp. 39, 622–632.

Yekutieli, D., Benjamini, Y., 1999. Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. J. Stat. Plan. Inference 82, 171–196.

Yoshita, M., Fletcher, E., DeCarli, C., 2005. Current concepts of analysis of cerebral white matter hyperintensities on magnetic resonance imaging. Top. Magn. Reson. Imaging 16, 399–407.

Yoshita, M., Fletcher, E., Harvey, D., Ortega, M., Martinez, O., Mungas, D.M., Reed, B.R., DeCarli, C.S., 2006. Extent and distribution of white matter hyperintensities in normal aging, MCI, and AD. Neurology 67, 2192–2198.