

White matter hyperintensities classified according to intensity and spatial location reveal specific associations with cognitive performance

Luca Melazzini^{a,b}, Clare E Mackay^{b,c,d}, Valentina Bordin^{b,e}, Sana Suri^{c,f}, Enikő Zsoldos^{b,c,f}, Nicola Filippini^c, Abda Mahmood^f, Vaanathi Sundaresan^b, Marina Codari^g, Eugene Duff^{b,h}, Archana Singh-Manoux^{i,j}, Mika Kivimäki^j, Klaus P Ebmeier^f, Mark Jenkinson^b, Francesco Sardanelli^{a,k}, Ludovica Griffanti^{b,c}.

^a Department of Biomedical Sciences for Health, Università degli Studi di Milano, Milan, Italy

^b Wellcome Centre for Integrative Neuroimaging, Oxford Centre for Functional MRI of the Brain, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

^c Wellcome Centre for Integrative Neuroimaging, Oxford Centre for Human Brain Activity, Department of Psychiatry, University of Oxford, Oxford, UK

^d Oxford Health NHS Foundation Trust, Oxford, UK

^e Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

^f Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, UK

^g Department of Radiology, Stanford University School of Medicine, Stanford, USA

^h Department of Paediatrics, University of Oxford, Oxford, UK

ⁱ INSERM U1153, Epidemiology of Ageing and Neurodegenerative diseases, Université de Paris, Paris, France.

^j Department of Epidemiology and Public Health, University College London, London, UK

^k Department of Radiology, IRCCS Policlinico San Donato, Milan, Italy

HIGHLIGHTS

- Heterogeneous measures of WMHs are used in research and clinical practice.
- Location and image intensity should be considered in the assessment of WMHs.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

- T₁-hypointense WMHs were found to be associated with poorer cognitive performance.
- Sub-classes of WMHs provide promising results for translation into the clinic.

ABSTRACT

White matter hyperintensities (WMHs) on T₂-weighted images are radiological signs of cerebral small vessel disease. As their total volume is variably associated with cognition, a new approach that integrates multiple radiological criteria is warranted. Location may matter, as periventricular WMHs have been shown to be associated with cognitive impairments. WMHs that appear as hypointense in T₁-weighted images (T_{1w}) may also indicate the most severe component of WMHs. We developed an automatic method that classifies WMHs into four categories (periventricular/deep and T_{1w}-hypointense/nonT_{1w}-hypointense) using MRI data from 684 community-dwelling older adults from the Whitehall II study. To test if location and intensity information can impact cognition, we derived two general linear models using either overall or subdivided volumes. Results showed that periventricular T_{1w}-hypointense WMHs were significantly associated with poorer performance in several cognitive tests. We found no association between total WMH volume and cognition. These findings suggest that classifying WMHs according to both location and intensity in T_{1w} adds value over and above total WMH volume.

KEYWORDS: Aging, cerebral small vessel diseases, cognition, magnetic resonance imaging, white matter.

INTRODUCTION

White matter hyperintensities (WMHs) on T₂-weighted magnetic resonance images (MRI) are radiological signs of cerebral small vessel disease (SVD) (Wardlaw et al., 2013). They are associated with a higher incidence of stroke and dementia (DeBette and Markus, 2010), mood disorders, motor impairments and urinary incontinence (LADIS Study Group, 2011). Moreover, WMHs are related to cognitive impairments, particularly executive dysfunctions and poorer psychomotor speed (Bolanzadeh et al., 2012). Whilst WMHs have acquired considerable interest in the field of translational and clinical research, the assessment and the reporting of WMH volume are often inconsistent in research studies and medical practice (Frey et al., 2019).

The optimal MRI sequence to assess WMHs is fluid-attenuated inversion recovery (FLAIR). This sequence generates T₂-weighted images where the signal from the cerebrospinal fluid is suppressed and hyperintense regions stand out on a low intensity homogeneous background (Wardlaw et al., 2013). In research, quantification of WMHs is preferred to qualitative assessment due to higher reliability, sensitivity and objectivity of the former (De Guio et al., 2016; van den Heuvel et al. 2006) and the widespread availability of segmentation software. However, the interpretation of quantitative results and comparison between studies remain difficult due to acquisition-related differences (scanner, protocol), discrepancies between processing methods (pre-processing pipelines, method/tool used to extract WMHs measurements) and variations in the definition of what should be considered a WMH (De Guio et al., 2016). Harmonisation methods that reduce or compensate for the variability due to acquisition differences and/or processing discrepancies are being developed to enable comparisons between or pooling of MRI-derived measures from different datasets (Bertani &

Bordin, 2019). Notwithstanding, the lack of a clear definition on what should be segmented as a WMH and whether some WMH sub-classes are more clinically relevant than others warrant further investigation (Frey et al., 2019; Smith et al., 2019; Wardlaw et al., 2013).

Periventricular WMHs are more strongly associated with concurrent cognitive deficits than deep ones (Bolandzadeh et al., 2012). This is in line with longitudinal studies on regional baseline WMH volumes and their association with the risk of transition from intact cognition to mild cognitive impairment and dementia (De Groot et al., 2002, Kim et al., 2015, van Straaten et al., 2008). To explain this finding, the hypothesis of reduced brain reserve in periventricular regions has been put forward (De Groot et al., 2002). Despite this evidence, it is still unclear whether periventricular and deep WMHs would constitute a continuous entity or should be considered and reported separately (DeCarli et al., 2005). If the latter is true, a single method for distinguishing types of WMHs should be adopted among the many that have been proposed (Griffanti et al., 2018).

White matter hyperintensities can either be iso- or hypointense in T_1 -weighted images (Spilt et al., 2006; Wardlaw et al., 2013). To the best of our knowledge, there are no studies performed on brain tissue samples that investigate whether there is any difference in the underlying pathology between non T_1 -hypointense *versus* T_1 -hypointense WMHs. However, clinical research in multiple sclerosis has showed that demyelinating T_1 -hypointense lesions (“black holes”) represent permanent damage to the white matter and are associated with cognitive impairments (Nowaczyk et al., 2019). In WMHs of presumed vascular origin, the co-located hypointensity in T_1 -weighted images (here on referred to as T_1 -hypointense WMHs) may indicate more severe damage to the white matter than WMHs that are visible in T_2 -

weighted images without any corresponding hypointensities in the T_1 -weighted images (here on referred to as *non* T_1 -hypointense WMHs). Quantitative measures of the motion of water molecules *in vivo* using diffusion tensor imaging (DTI) metrics have shown microstructural changes in the white matter in WMH areas (Wardlaw et al., 2015). Altered DTI metrics were reported in multiple sclerosis lesions that turn into permanent “black holes” (Naismith et al., 2010). Similarly, we could expect lower fractional anisotropy and higher radial diffusivity to reflect more severe axonal and myelin damage in T_1 -hypointense WMHs than in *non* T_1 -hypointense WMHs. T_1 -hypointense WMHs could thereby represent the portion of WMHs that carries the highest clinical impact.

For the reasons outlined above, we tested the hypothesis that periventricular (rather than deep) and T_1 -hypointense (rather than non T_1 -hypointense) WMHs may indicate the most severe forms of lesion in terms of impact on cognitive function. To this purpose, we classified WMHs according to both their anatomical location and intensity in T_1 -weighted images in a large cohort of community-dwelling older adults, and studied whether this classification could provide added value on the association between WMHs and cognitive function. We implemented a method for categorising WMHs according to spatial location and intensity in T_1 -weighted images that was automatic and objective and used multiple linear regression analysis to see if these WMH sub-classes show specific associations with validated scores of cognitive functions.

Given the current limitations and discrepancies in WMHs definition, our ultimate goal was to identify which sub-class(es) are specifically linked to cognitive function. This would inform future guidelines to focus the assessment on clinically-relevant radiological criteria of WMHs

beyond their total extent, with a clear and objective definition of WMHs radiological appearance and location.

METHODS

The study sample was drawn from 800 participants of the Whitehall II imaging sub-study (Filippini et al., 2014), which is part of a larger prospective occupational cohort study of British civil servants established in 1985 (Marmot and Brunner, 2005). Ethical approval was obtained from the University of Oxford Central University Research Ethics Committee, and the UCL Medical School Committee on the Ethics of Human Research. Informed written consent was obtained from all participants. Socio-demographic, health and lifestyle variables and current cognitive function were assessed at the time of the MRI.

MRI data were acquired at the Oxford Centre for Functional MRI of the Brain (FMRIB), Wellcome Centre for Integrative Neuroimaging (University of Oxford), using a 3-T, Siemens Magnetom Verio (Erlangen, Germany) scanner with a 32-channel receive head coil from April 2012 to December 2014 (N = 550 participants) and a 3-T Siemens Prisma (Erlangen, Germany) with a 64-channel receive head-neck coil from July 2015 to December 2016 (N = 250 participants) due to a scanner upgrade. Details of acquisition protocols are shown in Zsoldos et al. (2020) and Filippini et al. (2014) and are reported in Supplementary Table S1. For the purpose of this study we used high-resolution T₁-weighted images, FLAIR images and diffusion weighted images (DWI).

All images were processed and analysed using FMRIB Software Library (FSL) v.6.0 tools (Jenkinson et al., 2012). Participants' T₁-weighted and FLAIR images were skull-stripped with

FSL-BET (Smith et al., 2002) and bias field corrected with FSL-FAST (Zhang et al., 2001). DWI scans were pre-processed as described in (Filippini et al., 2014) and a diffusion tensor model was fit at each voxel to obtain maps of fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD) and radial diffusivity (RD). T_1 -weighted and FA images were then linearly registered to the corresponding FLAIR with FSL-FLIRT (Jenkinson et al., 2001). WMHs segmentation was performed with FSL-BIANCA (Griffanti et al., 2016), using intensity features (T_1 -weighted, FLAIR and FA), local average intensities (3 voxels kernel), and spatial features (MNI coordinates obtained from the transformation between FLAIR and MNI for each subject, weighting factor of 2). To avoid scanner-specific biases in the estimates, BIANCA was trained with WMH masks manually delineated in a sub-sample of individuals scanned on the Prisma ($n = 24$) and Verio ($n = 24$) scanners and an independent sample from the UK Biobank study ($n = 12$). The total WMH mask included voxels exceeding a probability of 0.9 of being a WMH and located within a white matter mask as described in Griffanti et al. (2016). The total WMH volume was adjusted for the total brain volume and log transformed for statistical analysis.

We separated WMHs voxels into T_1 -hypointense and non T_1 -hypointense. To achieve this, we used FSL-FAST (Zhang et al., 2001) on T_1 -weighted images to perform tissue type segmentation and calculate maps of partial volume estimates (PVE) for the three classes (grey matter, white matter and cerebrospinal fluid). Due to their low-intensity values, T_1 -hypointense WMHs are classified by FAST as either grey matter or cerebrospinal fluid. We therefore classified voxels as non T_1 -hypointense WMHs the voxels within the total WMH mask where the corresponding white matter PVE was greater than 0.5. We then obtained T_1 -hypointense WMHs by subtraction.

We used a cluster-based approach to separate between periventricular and deep WMHs, similar to the “continuity to ventricle” criterion described by (Griffanti et al., 2018). To do so, we created an extended ventricle mask (i.e. a ventricle mask that extended beyond the ventricular boundaries) in the Montreal Neurological Institute (MNI) space. The extended ventricle mask consisted of the probability maps -set with a very low threshold- of the lateral ventricles, thalami and fornix bilaterally. We transformed the mask to the single-subject FLAIR space via the corresponding T_1 using linear (Jenkinson et al., 2002) and non-linear registration (Andersson et al., 2007), and classified as periventricular WMHs the clusters that overlapped with any part of the mask. Deep WMHs were then defined by subtraction.

We finally combined the two criteria and obtained four WMH masks for the following subclasses for each participant: periventricular T_1 -hypointense WMHs; periventricular non T_1 -hypointense WMHs; deep T_1 -hypointense WMHs; deep non T_1 -hypointense WMHs. Then, each mask was used to derive the corresponding WMH volume, which was adjusted for the total brain volume and log transformed for statistical analysis.

We performed univariate multiple linear regression using the univariate general linear model function type III sum of squares on SPSS version 25.0 (IBM Corp. Armonk, NY). The following measures of participants’ cognition were selected as indices of global functioning, executive function, processing speed, and memory, in line with (Bolanzadeh et al. 2012): Montreal cognitive assessment (MoCA), trail-making test (TMT, A and B), digit span forward, backwards and sequence, digit symbol, digit coding, Boston naming-60 test (BNT), phonemic (letter) fluency (FLU-L) and semantic (category) fluency (FLU-C) tests. For details on the cognitive tests please refer to (Filippini et al., 2014). Demographic variables (age at the examination, sex,

total years of education, systolic blood pressure, diastolic blood pressure) were used as covariates of no interest. For each cognitive test (dependent variable in the univariate multiple linear regression), two models were investigated, in which participants' demographic variables were kept unchanged while WMHs were included as either WMHs total volume (corresponding to the BIANCA output) or subdivided WMH volumes (corresponding to the four sub-classes reported above).

To better investigate the meaning of T₁-hypointensity in our sample, we performed two *post-hoc* analyses. First, we studied WMH microstructure using DTI-derived metrics, given the lack of histopathological data for this dataset. Accordingly, we compared T₁-hypointense and non T₁-hypointense WMHs in terms of average fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD) and radial diffusivity (RD) using paired *t*-tests to evaluate potential differences in the underlying microstructure. Second, since we noticed that a WMH often includes both T₁-hypointense and non T₁-hypointense voxels, we adopted an alternative T₁-weighted intensity-based classification of WMHs by dividing the total WMH map into WMH clusters with and without T₁-hypointense voxels. This sub-classification was then used to look at the prevalence of WMH clusters with a T₁-hypointense component and to better interpret the results of the main analysis.

We also used Pearson correlations to further investigate the associations between WMH sub-classes, total WMH volume and age.

Statistical significance was set at $\alpha=0.05$ (Di Leo & Sardanelli, 2020).

RESULTS

Participants characteristics (demographics, cognitive scores and MRI measures) are reported in Table 1. One hundred and sixteen participants were excluded from our analysis due to incomplete or poor-quality images (N = 44), neurological disorders (N = 30) and inaccurate WMH segmentation masks (N = 42) to leave a total of 684 community-dwelling older adults. An example of the WMH segmentation output resulting from the method we developed is provided in Figure 1.

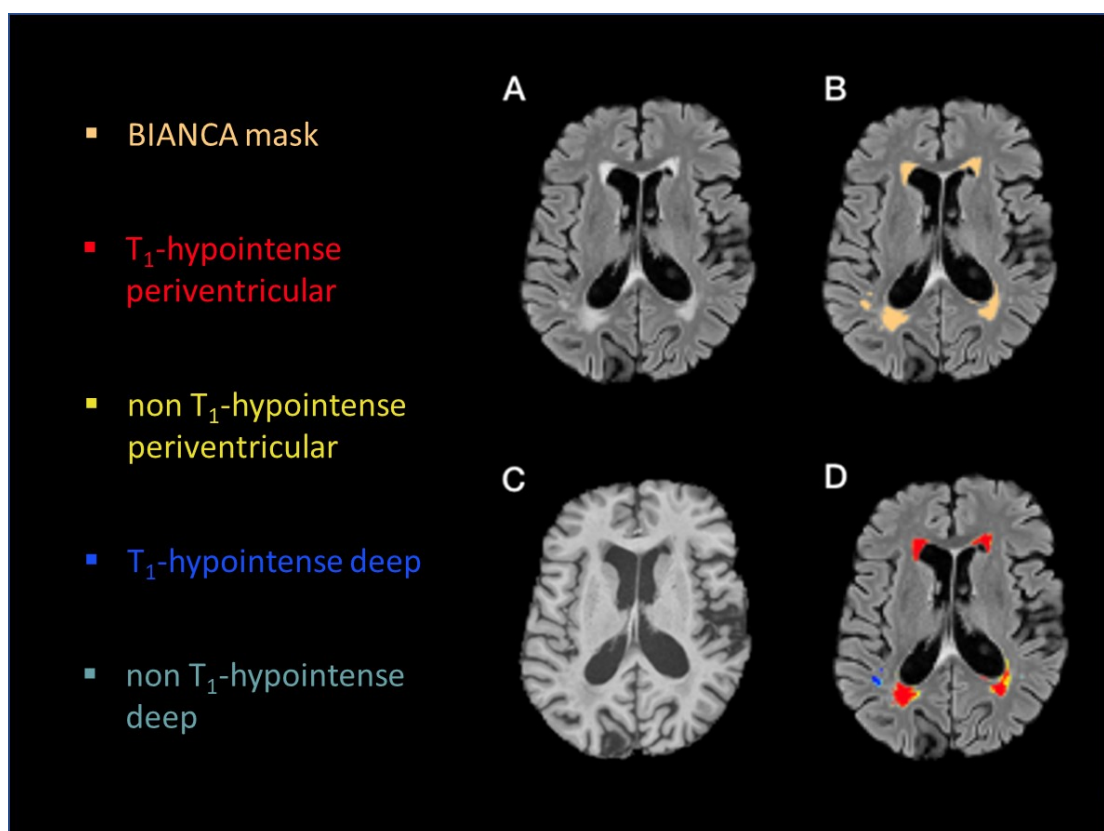


Figure 1. Native images and image processing output. (A) FLAIR image, (B) WMH mask (BIANCA mask - mustard colour) on FLAIR, (C) T_1 -weighted image, (D) WMH sub-classes on FLAIR: red = T_1 -hypointense periventricular WMHs; yellow = non T_1 -hypointense periventricular WMHs; blue = T_1 -hypointense deep WMHs; light blue = non T_1 -hypointense deep WMHs.

Table 1. Sample characteristics.

Descriptive statistics	N	Minimum	Maximum	Mean	Standard deviation
<i>Demographics</i>					
Age (years)	684	60.34	83.03	69.65	5.06
Sex (M:F)	551:133				
Total education (years)	684	13	28	19.10	2.85
Systolic blood pressure (mmHg)	681	93	226	141.09	17.54
Diastolic blood pressure (mmHg)	680	47	114	77.21	10.77
<i>Cognitive scores</i>					
MoCA	684	17	30	27.25	2.25
Trail making A	682	13	125	30.98	11.73
Trail making B	680	24	321	67.08	33.84
Trail making B-A	680	-8	253	36.25	28.73
Digit span forward	683	5	16	11.08	2.28
Digit span backwards	683	4	16	9.65	2.45
Digit span sequence	683	0	16	10.05	2.48
Digit symbol	683	13	46	30.78	5.75
Digit coding	683	13	114	63.02	13.22
Boston naming test-60	684	15	60	57.38	4.66
Letter fluency	684	3	31	15.78	4.64
Category fluency	684	3	40	22.36	5.34

MRI measures

Total brain volume (dm ³)	684	1.023	1.928	1.456	0.133
Total WMH volume (cm ³)	684	0.339	35.627	5.855	3.715
Periventricular T ₁ -hypointense WMHs (cm ³)	684	0.240	31.448	2.679	3.025
Periventricular non T ₁ -hypointense WMHs (cm ³)	684	0.050	6.379	2.138	0.831
Deep T ₁ -hypointense WMHs (cm ³)	684	0.001	5.274	0.434	0.577
Deep non T ₁ -hypointense WMHs (cm ³)	684	0.003	3.976	0.604	0.565

Legend: MoCA, Montreal cognitive assessment; WMHs, white matter hyperintensities.

White matter hyperintensities expressed as total volume were not significantly associated with any of the cognitive tests. However, when the four sub-classes of WMHs were instead used in the model, we found statistically significant associations between periventricular T₁-hypointense WMHs and poorer performance on the TMT-A ($p = 0.011$), digit symbol ($p = 0.028$) and digit coding ($p = 0.009$) tests. Conversely, non T₁-hypointense periventricular WMHs were associated with higher scores on the digit backwards test ($p = 0.023$). Also, deep non T₁-hypointense WMHs were found to be positively associated with the letter ($p = 0.004$) and category ($p = 0.036$) fluency tests. Regression analysis results for the models where WMH sub-classes were significantly associated with cognitive performances are shown in Table 2. Comprehensive results for all the models, including all covariates, are showed in Supplementary Table S2.

Table 2. Multiple linear regression analysis results. Model results and significant WMH predictors of cognitive scores adjusted for age at the examination, sex, total years of education, systolic blood pressure and diastolic blood pressure.

Model	Parameter estimates
-------	---------------------

Cognitive test	F	p-value	Adjusted R ²	Significant WMH covariate	β (SE)	p-value [#]
Trail making A	10.082	<0.001	0.108	T ₁ -hypointense periventricular WMHs	4.792 (1.870)	0.011
Digit span backwards	5.917	<0.001	0.061	Non T ₁ -hypointense periventricular WMHs	1.300 (0.571)	0.023
Digit span sequence	4.988	<0.001	0.050	T ₁ -hypointense periventricular WMHs	-0.787 (0.408)	0.054 [^]
Digit symbol	8.249	<0.001	0.088	T ₁ -hypointense periventricular WMHs	-2.045 (0.927)	0.028
Digit coding	11.062	<0.001	0.118	T ₁ -hypointense periventricular WMHs	-5.524 (2.093)	0.009
Letter fluency	3.737	<0.001	0.035	Non T ₁ -hypointense deep WMHs	2.097 (0.723)	0.004
Category fluency	8.071	<0.001	0.086	Non T ₁ -hypointense deep WMHs	1.829 (0.871)	0.036

Legend: SE, standard error; WMHs, white matter hyperintensities.

[#]P-values did not remain significant after adjustment for multiple comparisons across cognitive tests.

[^]Predictor that shows a trend for significance (0.05<p<0.06).

When looking at microstructural differences in WMHs classified according to the intensity in T₁-weighted images criterion, T₁-hypointense WMHs showed significantly lower FA, and higher MD, AD and RD than non T₁-hypointense WMHs (p < 0.001 for all measures). Results of the within-subject paired *t*-tests for the averages of DTI-derived metrics between T₁- and non T₁-hypointense WMHs are reported in Table 3.

Table 3. Comparison between the T₁- and non T₁- hypointense WMHs in terms of DTI-derived metrics. Paired *t*-tests results. Diffusivity values are expressed in (mm²/s) x 10⁻³.

Metric	T ₁ -hypointense WMHs		Non T ₁ -hypointense WMHs		<i>t</i>	p-value [#]
	Mean	SD	Mean	SD		
FA	0.26	0.03	0.37	0.04	-86.21	<0.001
MD	1.12	0.09	0.92	0.06	71.24	<0.001

AD	1.42	0.11	1.30	0.08	33.03	<0.001
RD	0.97	0.09	0.73	0.06	88.14	<0.001

Legend: FA, fractional anisotropy; MD, mean diffusivity; AD, axial diffusivity; RD, radial diffusivity; SD, standard deviation; WMHs, white matter hyperintensities.

#P-values remained significant after adjustment for multiple comparisons across metrics.

Results for the separation of the total WMH volume into WMH clusters with and without corresponding T₁-hypointense voxels are reported in Table 4.

Table 4. Number and size of WMH clusters with and without T₁-hypointense voxels.

	Number	Average volume <i>per subject</i> (mm ³)	Average volume <i>per cluster</i> (mm ³)
WMH clusters	223.91 ± 103.77	5855.01 ± 3715	35.50 ± 41.70
Clusters without T ₁ -hypointense voxels	112.43 ± 59.43	295.72 ± 188.87	2.58 ± 0.72
Clusters with T ₁ -hypointense voxels	111.48 ± 57.55	5559.29 ± 3754.63	66.16 ± 63.13
<i>Rims</i>		2446.69 ± 1021.26	27.96 ± 18.99
<i>Cores</i>		3112.61 ± 3227.86	49.72 ± 18.99

We found that 50% of the WMH clusters had T₁-hypointense voxels. However, despite constituting half of all clusters, these WMHs composed 94% of the whole WMH volume. The remaining 6 % of the WMH volume consisted of small non T₁-hypointense WMH clusters, i.e. WMH clusters without co-located hypointensity in T₁-weighted images.

The volume of non T₁-hypointense WMHs is the sum of non T₁-hypointense clusters and the hyperintense rims of T₁-hypointense clusters (a schematic representation is depicted in Figure 2).

Given the unexpected significant positive associations of non T₁-hypointense WMHs with the digit span backwards, letter and category fluency, we investigated which of these two components (either non T₁-hypointense clusters or rims of T₁-hypointense clusters) drove the observed associations. When fitting multiple linear regression models with either of these two components as the only WMH predictor for the fluency and digit span backwards scores, we found that non T₁-hypointense clusters were significantly associated with the fluency performances (FLU-L: $\beta = 1.976$, $p = 0.001$; FLU-C: $\beta = 2.023$, $p = 0.004$). Conversely, rims were significant predictors of the digit span backwards scores ($\beta = 0.916$, $p = 0.035$).

Additional multiple regression models and Pearson correlations among WMH sub-classes, total WMH volume and age supporting these findings are reported in Supplementary Table S3 and S4.

DISCUSSION

In this study we sought to provide a clinically-oriented insight into WMHs by developing an automated method for classifying WMHs according to spatial location (periventricular *versus* deep WMHs) and lesion intensity in the corresponding T₁-weighted image (T₁-hypointense *versus* non T₁-hypointense WMHs). We fitted univariate multiple linear regression models using the volumes of WMH sub-classes as predictors for the participants' performance in several cognitive tests and then further explored the microstructural properties of T₁-hypointense WMHs to understand the meaning of this radiological appearance. Our classification proved to be clinically meaningful, as periventricular T₁-hypointense WMHs were found to be linked to poorer performance on multiple cognitive tests, including the trail

making A, digit symbol and digit coding tests. Our results suggest that periventricular T₁-hypointense WMHs would represent the best WMH biomarker for cognitive impairment.

Our finding of periventricular WMHs being associated with poorer cognition is in line with previous large longitudinal population-based studies on non-demented elderly (de Groot et al., 2002; Griffanti et al., 2018), where periventricular WMHs were found to be more strongly associated with participants' cognition than deep WMHs. Increased volume of periventricular WMHs in the Alzheimer Disease Neuroimaging Initiative dataset was also associated with evidence of beta-amyloid deposition in the brain (Marnane et al., 2016), suggesting a synergistic damage driven by concurrent SVD and Alzheimer's pathology in these areas. Periventricular regions are also characterised by high density of long associating fibres which link the cortex to the deep grey matter and other distant brain territories (Filley 1998). For this reason, they are potentially susceptible to pathologies that damage cortical arteries and eventually provoke distal hypoperfusion (Moody et al., 1990). Disrupted cholinergic activity is related to periventricular (and not deep) WMHs and may be involved in the physiopathological pathway that underpins the observed cognitive scores (Bohnen et al., 2009). Moreover, high periventricular WMHs were found to be associated with frontal cortical thinning, where both imaging findings were independently linked to executive dysfunctions (Seo et al., 2012). Altogether, these findings endorse our results of periventricular WMHs being more strongly associated with cognitive impairment than deep WMHs.

Although the location criterion is well established, less is known about the meaning of T₁-intensity in WMHs. This aspect has been well-investigated in demyelinating disease where hypointense lesions in T₁-weighted images are more likely to represent low axonal density

and irreversible tissue damage (Bitsch et al., 2001; van Walderveen et al., 1998). Our DTI analysis showed decreased FA and increased MD, AD and RD in WMHs T₁-hypointense WMHs. Not surprisingly, these findings mirror similar microstructural results found in multiple sclerosis (Vavasour et al., 2018) and suggest more severe damage to the white matter in T₁-hypointense areas than their T₁-isointense counterpart.

T₁-hypointense and non T₁-hypointense WMHs may represent two distinct entities with different meanings. Alternatively, WMHs could start as small punctate FLAIR hyperintensities and later develop a T₁-hypointense “core”. Despite the lack of longitudinal data in our cohort, we attempted to investigate further the meaning of the intensity in T₁-weighted images and the theoretical evolution of WMHs and their intensity in T₁-weighted images from a cross-sectional basis. Within the total WMH mask, we separated WMH clusters that contained some T₁-hypointense voxels (T₁-hypointense clusters) from those that did not (non T₁-hypointense clusters) (Table 4). Since most of the WMH volume comprised (relatively big) clusters with a T₁-hypointense component, our results suggest that the evolution of a WMH may be a cascade of events starting from small punctate lesions leading to bigger lesions with a T₁-hypointense core and surrounding non T₁-hypointense rim (Figure 2).

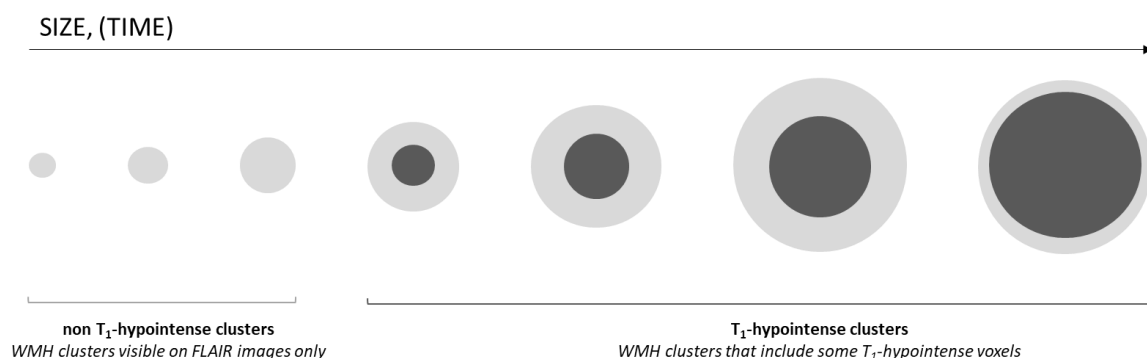


Figure 2. Diagram representing the potential evolution of WMHs over time. WMHs, white matter hyperintensity.

This mechanism could explain the significant association we observed between higher volume of deep non T₁-hypointense WMHs and higher scores at the fluency tests. Since the volume of non T₁-hypointense WMHs includes both the volume of non T₁-hypointense clusters and the outer rim of the clusters with a T₁-hypointense core, we hypothesized that the positive association of deep non T₁-hypointense WMHs with higher cognitive scores at the fluency tests could be driven by non-T₁ hypointense clusters (i.e. smaller, less severe WMHs, that have not yet progressed into a WMH with a T₁-hypointense component). We tested this hypothesis by fitting multiple linear regression models with either the volume of all non T₁-hypointense clusters or only the volume of deep non T₁-hypointense WMHs voxels as the only imaging predictors for the fluency tests (Supplementary Table S3). Notably, non T₁-hypointense clusters were more strongly associated with fluency scores than deep non T₁-hypointense WMHs. Thus, it is likely that the volume of clusters of non T₁-hypointense WMHs drove the observed significance of deep non T₁-hypointense WMHs as predictors of better cognitive scores at the fluency tests.

The positive associations of non T₁-hypointense clusters with fluency scores could be due to the fact that these small lesions are more frequent in younger individuals with a low WMH volume. The hypothesized evolution into T₁-hypointense clusters would then explain the apparent decrease of this type of lesions in individuals with lower fluency scores.

Since non T₁-hypointense clusters are small, we cannot exclude the possibility that some of these are false positives from BIANCA segmentation. However, WMH masks have been

visually checked so that we exclude results that could be driven by errors in WMHs segmentation.

Our finding of higher non T₁-hypointense periventricular WMHs linked to higher scores at the digit span backwards test could instead be explained by the other component of non T₁-hypointense WMHs, the non T₁-hypointense voxels belonging to T₁-hypointense clusters (i.e. the hyperintense “rims”). In fact, when fitting the multiple linear regression model with the volumes of rims of all T₁-hypointense clusters as the only imaging predictor, we found a positive association with the digit span backwards that was very similar to the one given by non T₁-hypointense periventricular WMHs alone (Supplementary Table S3). Although rim and core belong to the same physical entity (i.e. the WMH cluster) they are likely to have opposite meanings. On the one hand, T₁-hypointense WMHs voxels predict bad cognitive scores, as seen for the Trail Making Test A, digit symbol and digit coding tests. On the other hand, in spite of the positive association between rims and cores in terms of overall volume, the former are predictors of higher cognitive scores from the digit span backwards test. Thus, within the context of our hypothesized evolution of WMHs, rims would represent those WMH areas belonging to T₁-hypointense clusters that have not turned T₁-hypointense yet and theoretically “withstand” further tissue damage. When this occurs, the number of non T₁-hypointense voxels would decrease because they become T₁-hypointense. This would in turn explain the positive relationship with cognition observed in our results. Further investigation would be necessary to explore the potential cascade of events leading to change in T₁ intensity in a longitudinal setting, since the hypothesis of rims as WMHs associated with healthy cognitive aging is very speculative.

Our study has some limitations in terms of the data and the methodology.

The Whitehall II imaging sub-study dataset shows a strong gender imbalance, skewed towards men, since it reflects the demographic of British civil servants at the time of recruitment in the main study. Moreover, the age range of the sample is quite narrow (60-84 years). Future work is therefore needed to test the generalisability of our results. Our approach for the sub-classification of WMHs relies on automated segmentation of WMHs with BIANCA, tissue type segmentation with FAST, and registration of images with different spatial resolutions. As already mentioned, we cannot exclude inaccuracies in the WMH masks, despite visual inspection of the results. We used FAST segmentation as a proxy for defining T_1 -hypointensity, therefore inaccuracies in the segmentation would translate to inaccuracies in the sub-classification. Moreover, we performed linear and non-linear registrations between images with different resolutions (FLAIR, T_1 and MNI space) and the interpolation process could have slightly affected the segmented volumes.

Finally, our hypothesis on the evolution of WMHs should be interpreted cautiously and prompt further longitudinal studies. For example, it would be very valuable to follow up participants and study how WMH sub-classes evolve over time to validate the proposed theory. Another interesting future development would be looking at how the different WMH sub-classes are related to incidence of disease using risk models in well-balance longitudinal datasets.

Despite these limitations, this study presents some novel theoretical and methodological insights that can contribute to better understanding of the role of WMHs in cognitive aging. The methods developed herein can be easily adopted in other research settings. The extended ventricle mask and the scripts created for images post-processing are publicly

available¹. These scripts can also be equally applied to any manually- or automatically-derived WMH masks, other than those from BIANCA, to obtain the four WMH sub-classes presented in this study.

In conclusion, we showed that information from spatial location and intensity in T₁-weighted images provides potentially clinically useful insights into the meaning of WMHs with regards to participants' cognitive function. Notably, the combination of these two criteria revealed an association with cognitive scores related to global functioning, executive function, processing speed, and memory that the WMH total volume alone could not provide.

ACKNOWLEDGMENTS

We thank all Whitehall II participants for their time, the Whitehall II staff at the University College London, Mandy Pipkin and Barbora Krausova for assisting with recruitment and data collection, the FMRI Radiographers team for data acquisition, IT and support teams at the Wellcome Centre for Integrative Neuroimaging for their helpful collaboration. The study follows MRC data sharing policies (<https://www.mrc.ac.uk/research/policies-and-guidance-for-researchers/data-sharing/>). Data will be accessible via the Dementias Platform UK (<https://portal.dementiasplatform.uk/>) after 2020.

FUNDING

¹ Git repository: <https://git.fmrib.ox.ac.uk/ludovica/wmh-sub-classes>

The study was supported by the UK Medical Research Council (MRC) grants “Dementias Platform UK” (MR/L023784/2) and “Predicting MRI abnormalities with longitudinal data of the Whitehall II Substudy” (UK Medical Research Council: G1001354, PI: KPE), and by the HDH Wills 1965 Charitable Trust (Nr: 1117747, PI: KPE). This study was also supported by the Wellcome Centre for Integrative Neuroimaging, which has core funding from the Wellcome Trust (203139/Z/16/Z).

C.E.M., N.F. and L.G. were supported by the National Institute for Health Research (NIHR) Oxford Health Biomedical Research Centres (BRC), a partnership between Oxford Health NHS Foundation Trust and the University of Oxford. L.G. was also supported by the Oxford Parkinson’s Disease Centre (Parkinson’s UK Monument Discovery Award) and the MRC Dementias Platform UK. E.Zs, K.P.E. and S.S. were supported by the European Union’s Horizon 2020 programme ‘Lifebrain’ (732592). S.S. was also supported by an Alzheimer’s Society Junior Research Fellowship (Grant ref: 441). V.S. and M.J. were supported by the Wellcome Centre for Integrative Neuroimaging. M.J. was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC), and this research was funded by the Wellcome Trust (215573/Z/19/Z). A.S.-M. receives research support from the US National Institutes of Health (R01AG056477). M.K. was supported by NordForsk, the UK Medical Research Council (MRC S011676), the Academy of Finland (311492), and the US National Institutes on Aging (NIA R01AG056477, RF1AG062553).

REFERENCES

Andersson, J. L., Jenkinson, M., & Smith, S. (2007). Non-linear registration aka Spatial normalisation FMRIB Technical Report TR07JA2. FMRIB Analysis Group of the University of Oxford.

Bertani, I. & Bordin, V. (2019). Harmonizing large-scale imaging databases to provide integrated assessments of the role of white matter hyperintensities in cognitive aging. Full PDF file accessible at: <https://www.politesi.polimi.it/handle/10589/150951>. Last consulted on 2 July 2020.

Bitsch, A., Kuhlmann, T., Stadelmann, C., Lassmann, H., Lucchinetti, C., & Brück, W. (2001). A longitudinal MRI study of histopathologically defined hypointense multiple sclerosis lesions. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 49(6), 793-796.

Bohnen, N. I., Mueller, M. L., Kuwabara, H., Constantine, G. M., & Studenski, S. A. (2009). Age-associated leukoaraiosis and cortical cholinergic deafferentation. *Neurology*, 72(16), 1411-1416.

Bolanzadeh, 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 neurology*, 12(1), 126.

De Groot, J. C., De Leeuw, F. E., Oudkerk, M., Van Gijn, J., Hofman, A., Jolles, J., & Breteler, M. M. (2002). Periventricular cerebral white matter lesions predict rate of cognitive decline. *Annals of neurology*, 52(3), 335-341.

De Guio, F., Jouvent, E., Biessels, G. J., Black, S. E., Brayne, C., Chen, C., ... & Duering, M. (2016). Reproducibility and variability of quantitative magnetic resonance imaging markers in cerebral small vessel disease. *Journal of Cerebral Blood Flow & Metabolism*, 36(8), 1319-1337.

DeBette, S., & Markus, H. S. (2010). The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *Bmj*, 341, c3666.

DeCarli, C., Fletcher, E., Ramey, V., Harvey, D., & Jagust, W. J. (2005). Anatomical mapping of white matter hyperintensities (wmh) exploring the relationships between periventricular WMH, deep WMH, and total WMH burden. *Stroke*, 36(1), 50-55.

Di Leo, G., & Sardanelli, F. (2020). Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach. *European Radiology Experimental*, 4(1), 1-8.

Filippini, N., Zsoldos, E., Haapakoski, R., Sexton, C. E., Mahmood, A., Allan, C. L., ... & Auerbach, E. (2014). Study protocol: the Whitehall II imaging sub-study. *BMC psychiatry*, 14(1), 159.

Filley, C. M. (1998). The behavioral neurology of cerebral white matter. *Neurology*, 50(6), 1535-1540.

Frey, B. M., Petersen, M., Mayer, C., Cheng, B., & Thomalla, G. (2019). Characterisation Of White Matter Hyperintensities In Large-scale MRI-Studies: A Systematic Review. *Frontiers in Neurology*, 10, 238.

Griffanti, L., Jenkinson, M., Suri, S., Zsoldos, E., Mahmood, A., Filippini, N., ... & Singh-Manoux, A. (2018). Classification and characterization of periventricular and deep white matter hyperintensities on MRI: a study in older adults. *Neuroimage*, 170, 174-181.

Griffanti, L., Zamboni, G., Khan, A., Li, L., Bonifacio, G., Sundaresan, V., ... & Jenkinson, M. (2016). BIANCA (Brain Intensity AbNormality Classification Algorithm): a new tool for automated segmentation of white matter hyperintensities. *Neuroimage*, 141, 191-205.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *Neuroimage*, 62(2), 782-790.

Kim, S., Choi, S. H., Lee, Y. M., Kim, M. J., Kim, Y. D., Kim, J. Y., ... & Shim, Y. S. (2015). Periventricular white matter hyperintensities and the risk of dementia: a CREDOS study. *International psychogeriatrics*, 27(12), 2069-2077.

LADIS Study Group. (2011). 2001–2011: a decade of the LADIS (Leukoaraiosis And DISability) Study: what have we learned about white matter changes and small-vessel disease? *Cerebrovascular diseases*, 32(6), 577-588.

Marmot, M., & Brunner, E. (2005). Cohort profile: the Whitehall II study. *International journal of epidemiology*, 34(2), 251-256.

Marnane, M., Al-Jawadi, O. O., Mortazavi, S., Pogorzelec, K. J., Wang, B. W., Feldman, H. H., ... & Alzheimer's Disease Neuroimaging Initiative. (2016). Periventricular hyperintensities are associated with elevated cerebral amyloid. *Neurology*, 86(6), 535-543.

Moody, D. M., Bell, M. A., & Challa, V. R. (1990). Features of the cerebral vascular pattern that predict vulnerability to perfusion or oxygenation deficiency: an anatomic study. *American Journal of Neuroradiology*, 11(3), 431-439.

Naismith, R. T., Xu, J., Tutlam, N. T., Scully, P. T., Trinkaus, K., Snyder, A. Z., ... & Cross, A. H. (2010). Increased diffusivity in acute multiple sclerosis lesions predicts risk of black hole. *Neurology*, 74(21), 1694-1701.

Nowaczyk, N., Kalinowska-Łyszczarz, A., Paprzycki, W., Michalak, S., Kaźmierski, R., & Pawlak, M. A. (2019). Spatial distribution of white matter degenerative lesions and cognitive dysfunction in relapsing-remitting multiple sclerosis patients. *Neurologia i neurochirurgia polska*, 53(1), 18-25.

Patenaude, B., Smith, S. M., Kennedy, D. N., & Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage*, 56(3), 907-922.

Seo, Sang Won, et al. "Cortical thinning related to periventricular and deep white matter hyperintensities." *Neurobiology of aging* 33.7 (2012): 1156-1167.

Smith, E. E., Biessels, G. J., De Guio, F., de Leeuw, F. E., Duchesne, S., Düring, M., ... & Thrippleton, M. J. (2019). Harmonizing brain magnetic resonance imaging methods for vascular contributions to neurodegeneration. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 11, 191-204.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping*, 17(3), 143-155.

Spilt, A., Goekoop, R., Westendorp, R. G. J., Blauw, G. J., de Craen, A. J. M., & Van Buchem, M. A. (2006). Not all age-related white matter hyperintensities are the same: a magnetization transfer imaging study. *American journal of neuroradiology*, 27(9), 1964-1968.

Van den Heuvel, D. M. J., Ten Dam, V. H., De Craen, A. J. M., Admiraal-Behloul, F., Van Es, A. C. G. M., Palm, W. M., ... & Westendorp, R. G. J. (2006). Measuring longitudinal white matter changes: comparison of a visual rating scale with a volumetric measurement. *American Journal of Neuroradiology*, 27(4), 875-878.

van Straaten, E. C., Harvey, D., Scheltens, P., Barkhof, F., Petersen, R. C., Thal, L. J., ... & DeCarli, C. (2008). Periventricular white matter hyperintensities increase the likelihood of progression from amnesic mild cognitive impairment to dementia. *Journal of neurology*, 255(9), 1302.

Van Walderveen, M. A. A., Kamphorst, W., Scheltens, P. H., Van Waesberghe, J. H. T. M., Ravid, R., Valk, J., ... & Barkhof, F. (1998). Histopathologic correlate of hypointense lesions on T1-weighted spin-echo MRI in multiple sclerosis. *Neurology*, 50(5), 1282-1288.

Vavasour, I. M., Meyers, S. M., Mädler, B., Harris, T., Fu, E., Li, D. K., ... & Laule, C. (2019). Multicenter measurements of T1 relaxation and diffusion tensor imaging: intra and intersite reproducibility. *Journal of Neuroimaging*, 29(1), 42-51.

Wardlaw, J. M., Smith, E. E., Biessels, G. J., Cordonnier, C., Fazekas, F., Frayne, R., ... & Black, S. E. (2013). Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *The Lancet Neurology*, 12(8), 822-838.

Wardlaw, J. M., Valdés Hernández, M. C., & Muñoz-Maniega, S. (2015). What are white matter hyperintensities made of? Relevance to vascular cognitive impairment. *Journal of the American Heart Association*, 4(6), e001140.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, 20(1), 45-57.

Zsoldos, E., Mahmood, A., Filippini, N., Suri, S., Heise, V., Griffanti, L., ... & Ebmeier, K. P. (2020). Association of mid-life stroke risk with structural brain integrity and memory performance at older ages: A longitudinal cohort study. *Brain Communications*.