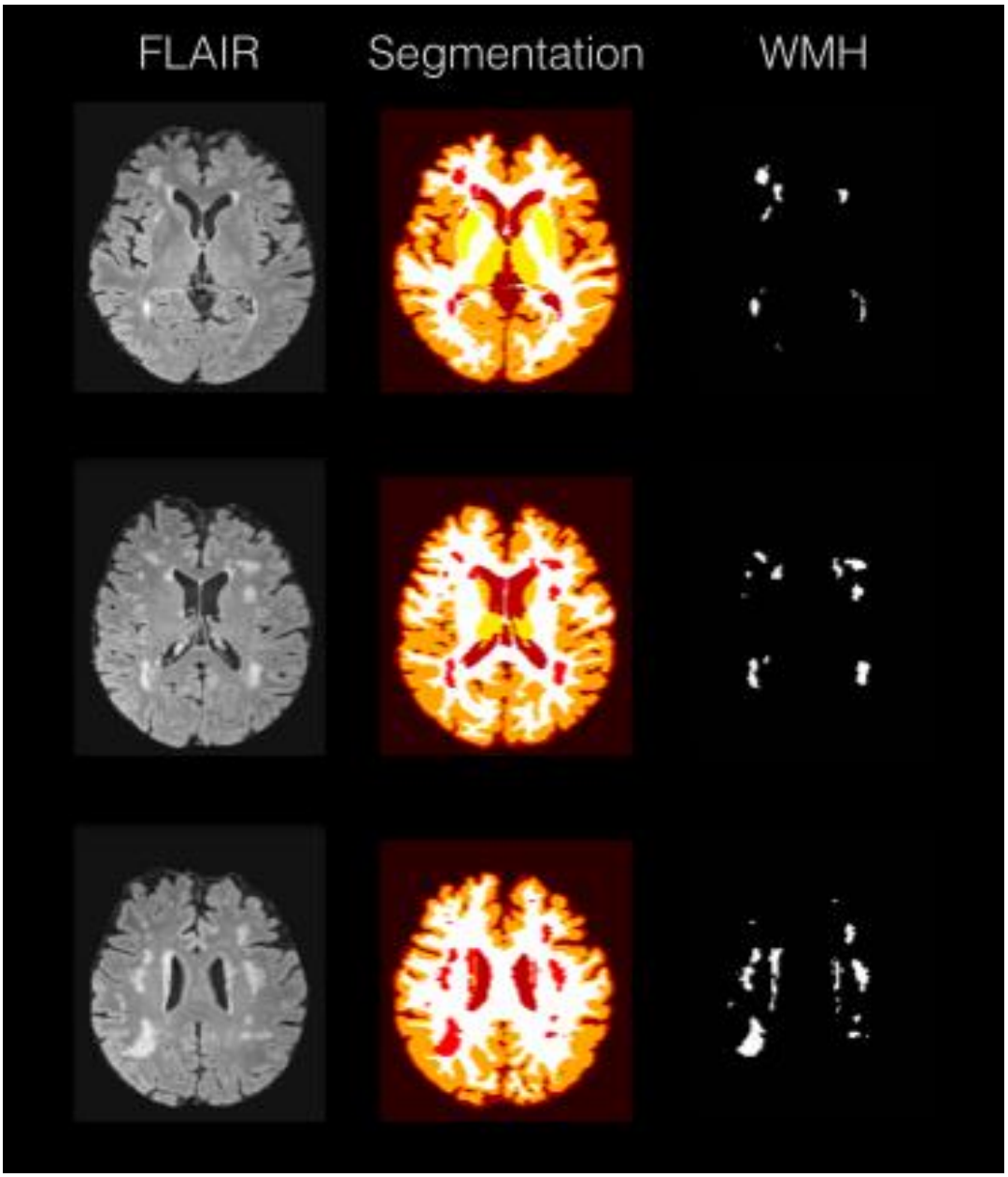
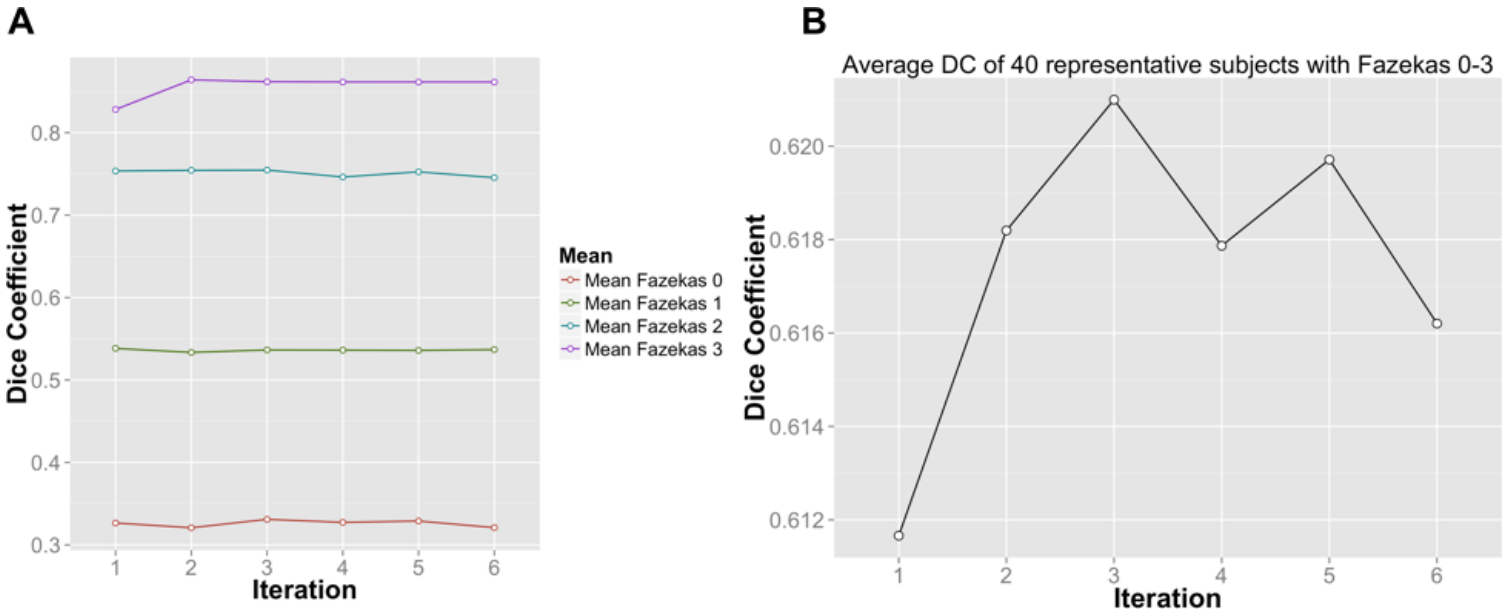
Supplementary Material

LesionTOADS in the healthy elderly

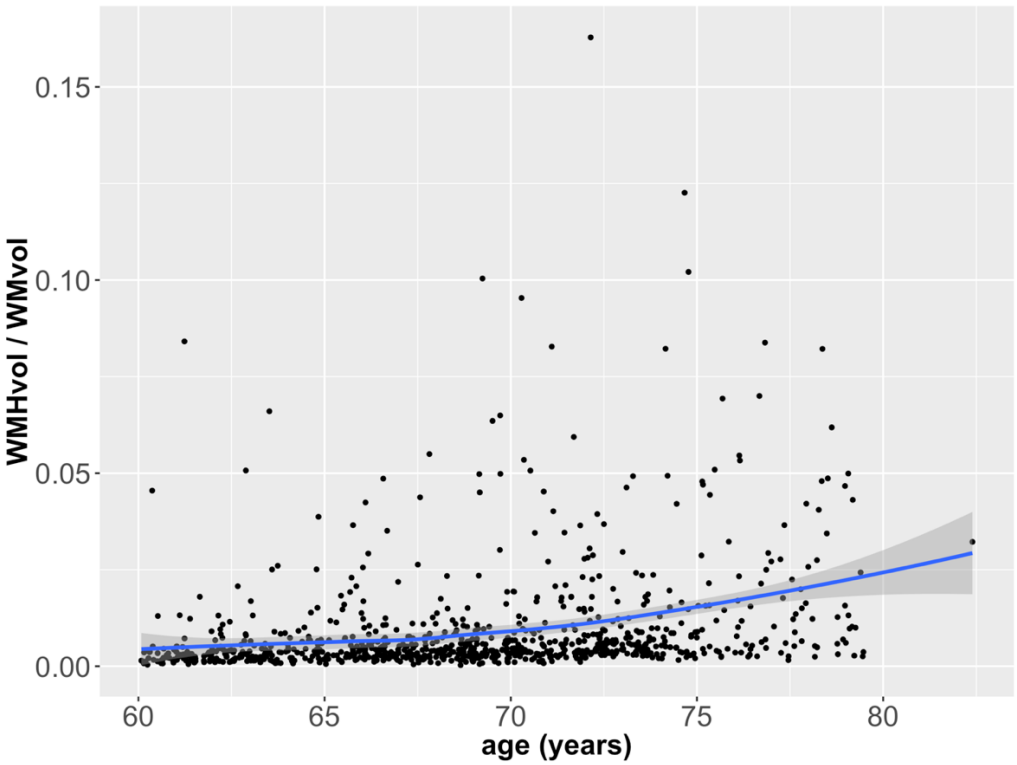
Lesion TOADS—a computer-based WMH segmentation algorithm that was previously validated for patients with multiple sclerosis—was used to automatically determine WMH volume on T1-weigthed MPRAGE and FLAIR images.[14] As WMH differ in their pattern, intensity, and extent to those of multiple sclerosis, we adapted the algorithm to the needs of age-related WMH in our cohort in a two-step iterative process and revalidated it the algorithm on our cohort. Given the large variety of WMH, their appearance on FLAIR images had a wider range of intensities across participants. Therefore, from the first segmentation given by LesionTOADS (step 1), we renormalized the contrast of input FLAIR images to better separate WMH from healthy tissue, modeling the FLAIR intensity inside the brain as a mixture of Gaussian and outlier distributions (step 2). The intensity boundary between tissue intensity and WMH is estimated from setting the ratio of segmented WMH to brain volume as an outlier ratio, and performing the LesionTOADS step again on the re-normalized intensities. This method ensures that the relative intensities of the FLAIR in different images is similar across a large cohort of subjects. We then performed a revalidation of LesionTOADS with groups of different White matter disease severity. Groups were defined by experienced neuroradiologists according to the widely used Fazekas scale,[2] and each Fazekas subcategory had ten participants (forty participants in total). Manually delineated WMH maps served as the “gold standard”. Three iterations were necessary to achieve an improved stable result (Supplementary Figure 2). The algorithm also performed a full brain segmentation into cerebral and cerebellar cortex, white matter, ventricles, brainstem, and subcortex (Supplementary Figure 1).



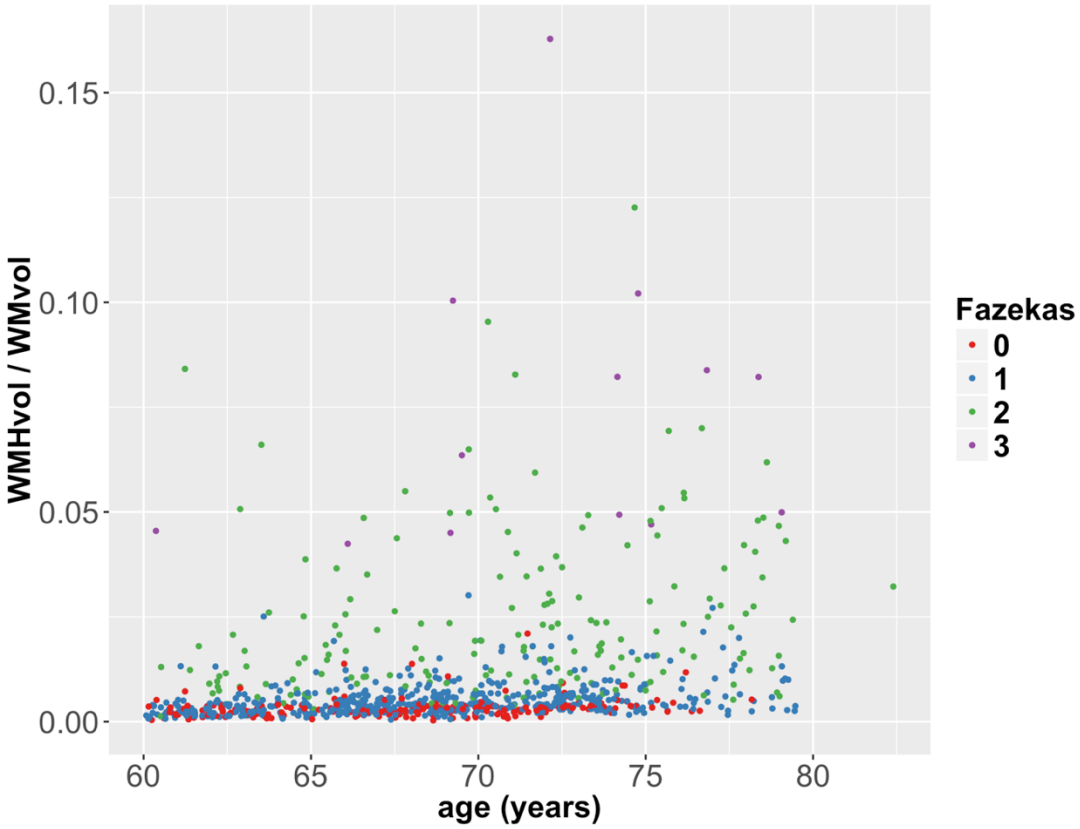
Supplementary Figure 1 (from left to right): Output of LesionTOADS; FLAIR image with WMH, full brain segmentation of T1weighted image, WMH mask



Supplementary Figure 2: Validation of LesionTOADS; Overlap between manual and automated segmentations during the renormalization step: A) Mean dice coefficient in different lesion groups; B) Average dice coefficient of all 40 participants



Supplementary Figure 3A: Normalized WMH volume (WMHvol/WMvol) across age in 890 healthy participants (60 - 82 years)



Supplementary Figure 3B: Normalized WMH volume (WMHvol/WMvol) across age in 890 healthy participants (60 - 82 years), color coding: different Fazekas groups

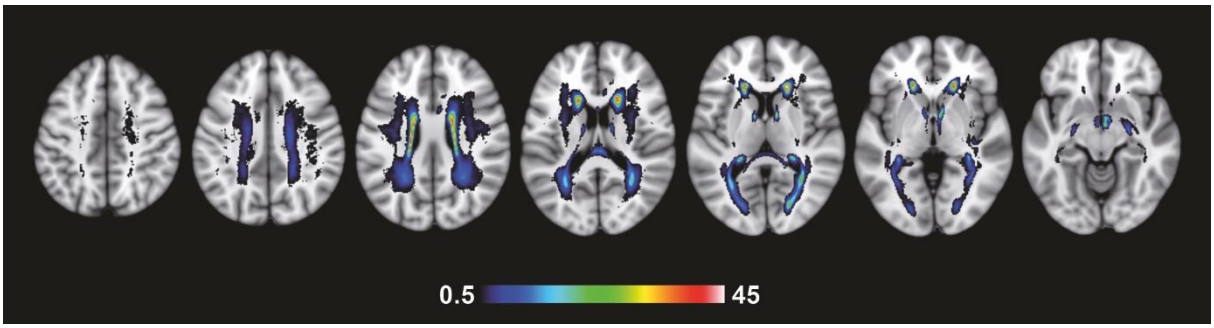
LesionTOADS in our sample

WMH volume in our cohort was assessed first visually according to the widely used Fazekas scale.[3] Forty subjects (ten of each Fazekas category) were subsequently manually delineated and served as the validation sample for LesionTOADS, a computer-based lesion segmentation algorithm[1] originally designed for multiple sclerosis patients. As WMH of vascular and inflammatory origin are similar in appearance, the algorithm rendered satisfactory results. By segmenting jointly brain structures and lesions from the MPRAGE and the FLAIR, false positive WMH detected in other tissue were minimized. High spatial resolution FLAIR images (1 mm isovoxel) increased the accuracy of the WMH volume assessment. WMH typically differ in intensity, pattern, and distribution. While some WMH are subtle in intensity differences and blurry because of diffuse configuration—also referred to as dirty-appearing

white matter[4]—other WMH are circumscribed focal lesions of high intensity signal. The latter is easily depicted while the former is easily missed by lesion segmentation algorithms in general as well as for LesionTOADS. A frequent challenge for detecting WMH in our cohort were motion artifacts because of discomfort in the scanner related to a lengthy scanning protocol, which lead to false positive WMH classification. Through careful inspection, we removed images with motion-induced, artificial WMH volume. We further noted occasional misclassification of tissue, for example, T2-hyperintense appearing plexus choroideus often falsely segmented as WMH. As a result of these precautions, the number of false positives is relatively small and consistent in all participants, independent of WMH volume, and can be neglected as systematic error. The high dice coefficient in participants with high lesion load and the lower dice coefficient in participants with few lesions in the revalidation of LesionTOADS likewise reflect false positives carrying a larger weight in participants with low or no WMH (Supplementary Figure 2). The overall dice coefficient of LesionTOADS is comparable with competing state-of-the-art lesion segmentation algorithms (MICCAI segmentation challenge[5]), and was consistent with multiple sclerosis lesion segmentation scores in our revalidation. It is also worth mentioning that the "gold standard" of a manual delineation cannot be considered an objective truth. We conclude that LesionTOADS is a competitive algorithm and well suited for segmenting age-related WMH in the general population, and that automated WMH segmentation is the more precise, rapid, and objective method by which to quantify WMH volume in comparison to visual rating scales.

Our processing pipelines with the detailed parameter settings is publicly available online:
https://figshare.com/s/34efe4617cb2e65c03f0
https://figshare.com/s/3c1ea16ef5ab67aec6cc

It can be easily visualized or applied to new data in cbstools.
https://github.com/piloubazin/cbstools-public

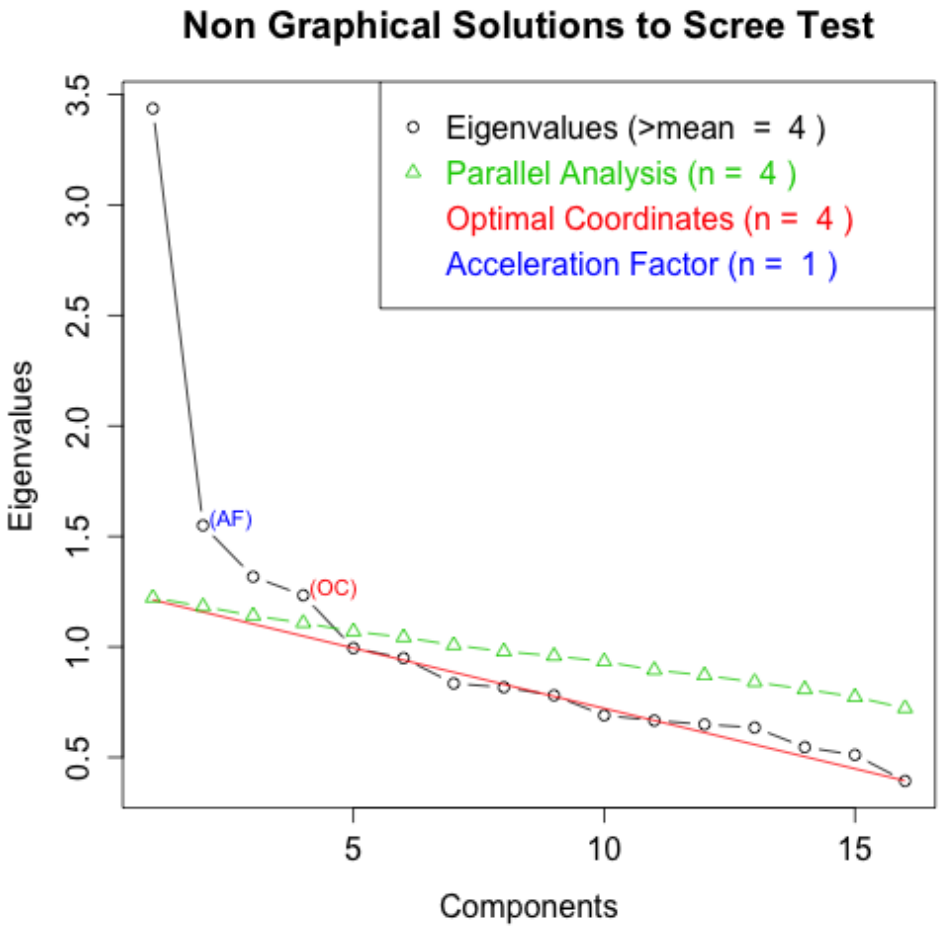


Supplementary Figure 4: WMH frequency maps; overlaid onto a standard MNI template; color bar: WMH frequency in percent

Supplementary Table 1: Kaiser-Meyer-Olkin Factor Adequacy

MSA total	**0.78**
MSA for each item	
Wortschatztest	0.84
Phonemic Verbal Fluency (S-Words)	0.79
Semantic Verbal Fluency (Animals)	0.83
SIDAM Working Memory	0.84
SIDAM Language	0.8
SIDAM Long Term Memory	0.82
Boston Naming Test	0.79
Delayed Recall Words	0.71
Recognition Words	0.76
Immediate Recall Words	0.79
SIDAM Short Term Memory	0.86
CERAD Figure Drawing	0.71
SIDAM Figure Drawing	0.74
Delayed Figure Drawing	0.83
TMT A	0.60
TMT B/A	0.69

Abbreviations: CERAD = Consortium to Establish a Registry for Alzheimer's Disease; MSA = Measure of Sampling Adequacy; SIDAM = the Structured Interview for Diagnosis of Dementia of Alzheimer type; TMT = Trail Making Test



Supplementary Figure 5: Scree plot of 16 items

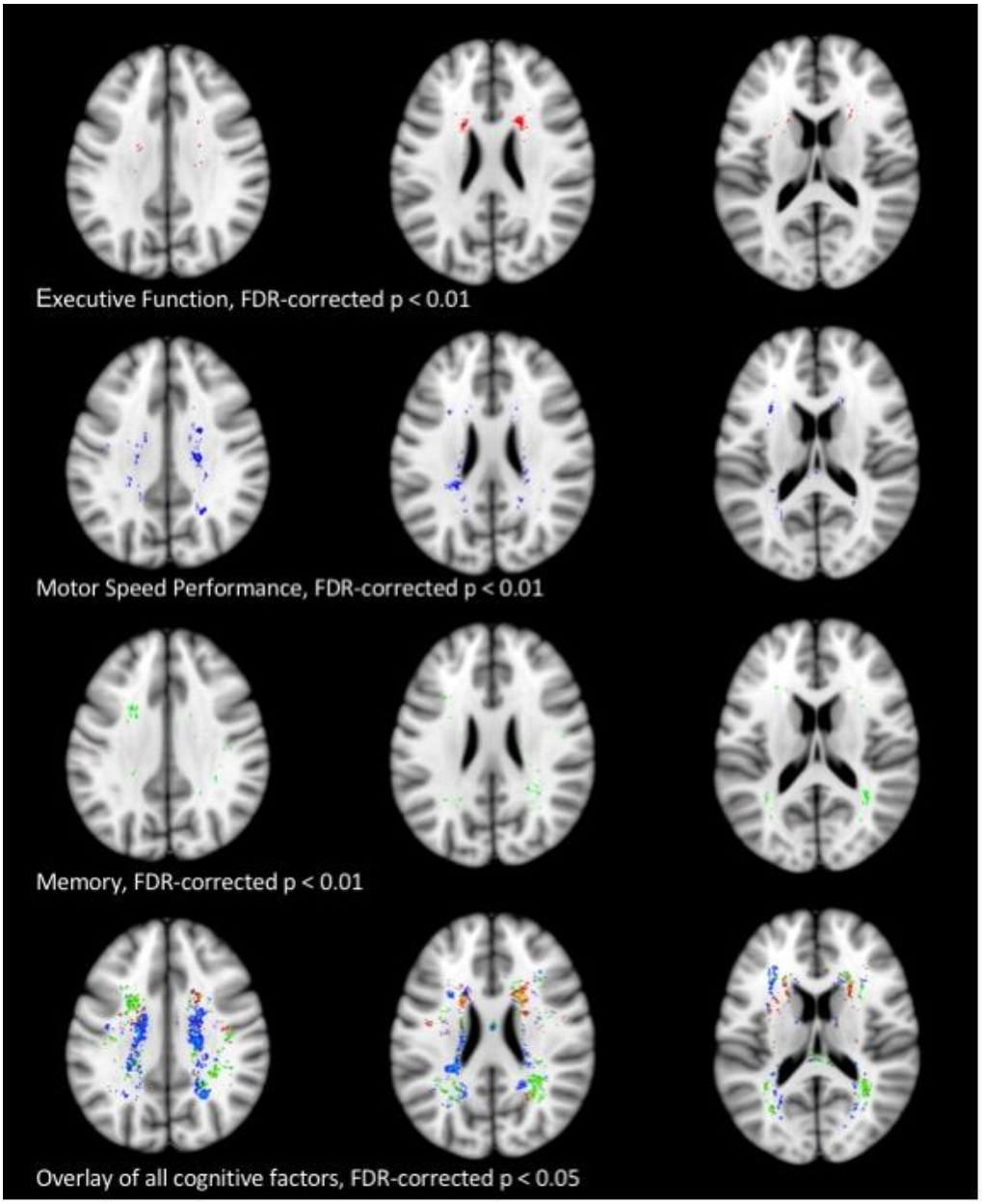
Factor extraction

To reduce the number of cognitive items and to detect the latent cognitive structure, principal axis factoring was conducted on the sixteen cognitive items (Supplementary Table 1). We chose an oblique rotation "oblimin" because we expected the cognitive latent variables not to be entirely independent from each other. Because not all cognitive items were available for all participants, only n = 702 subjects could be included, resulting in a subject to item ratio of 43.875. The Kaiser-Meyer-Olkin (KMO) criterion verified the sampling adequacy for the analysis with an overall measure of sampling adequacy (MSA) = 0.78 and KMO values for individual items between 0.60 and 0.86 (Supplementary Table 1). Four components in the data had an eigenvalue above one (see scree plot, Supplementary Figure 5) and cumulatively explained 47% of the variance. For the exploratory factor analysis, the psych package[6] was used as implemented in R (v3.3.1).[7] Scores were extracted with the regression method. Factor 1, Factor 2, and Factor 3 were roughly normally distributed while Factor 4 was negatively skewed due to the high loadings of visuoconstructive tasks with consistent ceiling effects. Meaningful names for the extracted factors were chosen after considering items with substantial loading (Supplementary Table 2). Factor 1 represented executive function with highest loadings of the Wortschatztest (German vocabulary test), verbal fluency tests, and the Trail Making Test (B/A). Factor 2 included tests related to memory and learning with high loadings of the word recall tests as well as the short term memory test. Factor 3 had a very high loading of the Trail Making Test A, which represented psychomotor speed. Factor 4 comprised all assessments that required visuoconstructive abilities, namely figure drawing tests.

Supplementary Table 2: Items of the exploratory factor analysis: Cognitive items with loadings for each respective factor (substantial loadings > 0.4 are marked bold)

Factor *(meaningful name)*	Factor 1 *("Executive Function")*	Factor 2 *("Memory")*	Factor 3 *("Motor Speed Performance")*	Factor 4 *("Visuoconstructive Abilities")*
Wortschatztest (German vocabulary test)	**0.562**	-0.044	-0.034	0.077
Phonetic Verbal Fluency	**0.561**	0.005	0.161	-0.062
Semantic Verbal Fluency	**0.556**	0.055	0.064	-0.013
SIDAM Working Memory	0.399	0.016	0.057	0.046
SIDAM Language	0.381	0.026	-0.053	0.008
SIDAM Long Term Memory	0.298	-0.058	-0.063	0.207
Boston Naming Test	0.295	-0.031	0.079	0.068
Delayed Recall Words	-0.026	**0.829**	0.032	0.011
Recognition Words	-0.042	**0.630**	-0.067	-0.041
Immediate Recall Words	0.087	**0.574**	0.041	0.021
SIDAM Short Term Memory	0.216	0.262	-0.020	0.110
CERAD Figure Drawing	-0.039	-0.013	-0.013	**0.702**
SIDAM Figure Drawing	0.001	0.008	0.010	**0.536**
Delayed Figure Drawing	0.151	0.087	0.139	0.396
TMT A	0.066	0.049	**0.784**	0.018
TMT B/A	**0.413**	0.120	-0.421	0.031

Abbreviations: CERAD = Consortium to Establish a Registry for Alzheimer's Disease; SIDAM = the Structured Interview for Diagnosis of Dementia of Alzheimer type; TMT = Trail Making Test



Supplementary Figure 6: Voxel wise lesion symptom mapping with the non-parametric Brunner-Munzel test; Z-values are shown for executive function depicted in red, for memory depicted in green, and for motor speed performance depicted in blue; All major significant clusters overlap with the the results depicted in Figure 2

References

1. Shiee N, Bazin P-L, Ozturk A, et al. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *Neuroimage* 2010; 49: 1524–35.
2. Fazekas F, Chawluk JB, Alavi A, et al. MR Signal Abnormalities at 1 . 5 T in Alzheimer ' s Dementia and Normal Aging deficiency. *Am J Roentgenol* 1987; 149: 351–356.
3. Fazekas F, Chawluk JB, Alavi A, et al. MR Signal Abnormalities at 1 . 5 T in Alzheimer ' s Dementia and Normal Aging deficiency. *AJNR Am J Neuroradiol* 1987; 149: 351–56.
4. Beggs CB, Magnano C, Shepherd SJ, et al. Dirty-Appearing White Matter in the Brain is Associated with Altered Cerebrospinal Fluid Pulsatility and Hypertension in Individuals without Neurologic Disease. *J Neuroimaging* 2016; 26: 136–143.
5. Styner M, Lee J, Chin B, et al. 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation. *Midas* 2008; 1–6.
6. Revelle W. Package 'psych' - Procedures for Psychological, Psychometric and Personality Research. *R Packag* 2015; 1–358.
7. R Core team. R Core Team. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing , Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.* 2015; 55: 275–286.