Published in final edited form as:

Psychiatry Res. 2016 July 30; 253: 7–14. doi:10.1016/j.psychres.2016.05.003.

The effects of white matter disease on the accuracy of automated segmentation

Helmet T. Karim^a, Carmen Andreescu^b, Rebecca L. MacCloud^b, Meryl A. Butters^b, Charles F. Reynolds III^b, Howard J. Aizenstein^{a,b}, and Dana L. Tudorascu^{b,c,d}

^aDepartment of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA

^bDepartment of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

^cDepartment of Internal Medicine, University of Pittsburgh, Pittsburgh, PA, USA

^dDepartment of Biostatistics, Graduate School of Public Health, Pittsburgh, PA, USA

Abstract

Automated segmentation of the brain is challenging in the presence of brain pathologies such as white matter hyperintensities (WMH). A late-life depression population was used to demonstrate the effect of WMH on brain segmentation and normalization. We used an automated algorithm to detect WMH, and either filled them with normal-appearing white-matter (NAWM) intensities or performed a multi-spectral segmentation, and finally compared the standard approach to the WMH filling or multi-spectral segmentation approach using intra-class correlation coefficients (ICC). The presence of WMH affected segmentations for both approaches suggesting that studies investigating structural differences in populations with high WMH should account for WMH. We also investigated how functional data contrasts are affected using normalization between the standard compared to fill and multi-spectral approach. We found that the functional data was not affected. While replication with a larger sample is needed, this study shows that WMH can significantly affect the results of segmentation and these areas are not limited to those affected by WMH. It is clear that to study gray matter differences that some correction should be made to account for WMH. Future studies should investigate which methods for accounting for WMH are most effective.

Keywords

Segmentation; MRI; fMRI; white matter hyperintensities

1. INTRODUCTION

White matter hyperintensities (WMH) are areas in the white matter of the brain that have higher intensity values on a T2-weighted magnetic resonance (MR) scans compared to normal appearing white matter (NAWM). Usually appearing in mid-to-late-life, WMH are attributed to degenerative changes of long penetrating arteries (Breteler et al., 1994),

resulting in demyelination, gliosis and axonal degeneration (Ovbiagele and Saver, 2006). WMH can have a variety of pathologies and are associated with several neurological disorders (e.g. multiple sclerosis). In older populations, the prevalence of white matter disease is highly variable, with reports indicating it affects from 5% to 90% of the population, depending on the study design and study population (Xiong and Mok, 2011). WMH are more common in populations 60 years and older and also are more common in women than in men (de Leeuw et al., 2001).

Although WMH appear as hyperintense on T2-weighted images, they are hypointense (dark) on the T1-weighted images. These dark spots on the T1-weighted images are not typically considered when using automated image processing methods. For instance, studies exploring structural brain changes rely on tissue segmentation algorithms. These algorithms can be affected by the presence of WMH. The process of tissue classification in presence of WMH is often a challenging task. Thus, the presence of WMH can result in tissue misclassification if it is not taken into account in the segmentation algorithm, as it has been reported in Levy et al (Levy-Cooperman et al., 2008). Most neuroimaging software platforms have been developed for populations with healthy brains and therefore there is only a low risk of tissue misclassification since the prevalence of WMH is very low or non-existent in younger populations. Therefore, incorrectly classifying the voxels as gray matter due to lower intensities on the T1 is not a significant issue in healthy subjects.

Further, it is often necessary to coregister all subject volumes into a common template space. This step can be affected by the presence of WMH (Eloyan et al., 2014) as atrophy and WMH can interfere with proper classification of the brain tissue. This interferes with proper coregistration to template space, especially in algorithms that rely on a unified segmentation/normalization algorithm.

WMH filling (Battaglini et al., 2012; Chard et al., 2010; Eloyan et al., 2014; Magon et al., 2014; Sdika and Pelletier, 2009; Valverde et al., 2014) is a previously established method used to alleviate the effect of WMH on segmentation/coregistration algorithms. After identifying WMH (using an automated method), they are filled with intensities of NAWM in the structural image. This allows for these regions to be segmented as white matter instead of being misclassified (usually as gray matter). Another approach utilizes a multi-spectral segmentation by using information from a T1-weighted and T2-weighted image. A WMH prevalence map can be input to identify regions most greatly affected by WMH. This allows for classification of these tissues and models the WMH as a separate tissue class.

Proper classification of brain tissue is required in order for the statistical analysis to be valid in populations with age-specific diseases or other diseases that can severely affect brain tissue. In this study, we used a group of older adults enrolled in a study of late-life depression (LLD). As depression in older adults is associated with the presence of small vessel ischemic disease (Alexopoulos, 2006; Taylor et al., 2013; Tudorascu et al., 2014) this group is likely to have a significant burden of WMH, and thus is well-suited for testing how WMH burden interfere with tissue segmentation and registration. In general, aging populations tend to have higher WMH burden, but patients with LLD often have an even higher WMH burden compared to age-matched healthy controls (Herrmann et al., 2008).

The main purpose of our work was to describe the effect of WMH on segmentation. We however did not investigate the efficacy of the methods used and only investigated the effect of some correction on the results. We also aimed to look at the effect of WMH on functional data coregistration/normalization to a standard (Montreal Neurological Institute, MNI) space. Thus, we examined (via intraclass correlation coefficients, ICC) the differences between the standard approach, which does not account for WMH, and the WMH filling approach as well as the multi-spectral segmentation approach. This will indicate the effect each method has on the previously uncorrected approach and does not indicate accuracy/efficacy.

2. METHODS

2.1. Study Design and Subjects

The subjects participated in a five-year multi-site study of treatment of Late Life Depression (LLD) with an imaging component added to the Pittsburgh site (Lenze et al., 2015). The University of Pittsburgh IRB approved this study. Subjects were included in the LLD study if they were older than 65 years of age, had major depressive disorder that met DSM-IV criteria (First et al., 1997), and had Montgomery-Asberg Depression Rating Scale (MADRS) score greater than 15. Subjects had a mean (std.) adapted Edinburgh handedness index of 17 (6) (right-handed, 1 missing subject with self-reported right-handedness) (Oldfield, 1971). Subjects were excluded if they had a history of mania or psychosis, alcohol/substance abuse, current use of antidepressant or involvement in mental health treatment, dementia or any neurodegenerative disease, prescribed medications used in Alzheimer's disease, and medical conditions with known significant effects on mood (e.g. clinical stroke) and neurologic disorders with known effects on the brain (e.g., multiple sclerosis). After subjects completed informed consent, a baseline MRI scan was acquired. As part of the multi-site study, subjects underwent detailed neurocognitive assessment, a clinical interview, mood and anxiety assessment, and laboratory testing.

A total of 37 participants signed consent, but four were not included for the following reasons: withdrawal due to failure to communicate/failure to comply, or they were deemed (early on) to no longer have Major Depressive Disorder (MDD). Six were not included because they either did not have a structural MPRAGE or FLAIR (see data collection section). A total of 27 subjects were included in this analysis (mean age 67 (SD=6) years).

2.2. MRI/fMRI Data Collection

All scanning was conducted using a 3T Siemens Trio TIM scanner (VB17) with a 12-channel head coil (parallel imaging) located at the Magnetic Resonance Research Center at the University of Pittsburgh. The head was immobilized using vacuum pads to minimize motion artifacts. Multiple imaging sequences were acquired. A map used to identify WMH was collected, specifically an axial T2-weighted fluid-attenuated inversion recovery (FLAIR, TR=9160 ms, TE=90 ms, TI=2500 ms, FA=150°, FOV=256×212, slice thickness=3 mm, slices=48, matrix size=256×212, no gap). An axial T1-weighted 3D sequence was collected (TR = 2300 ms, TE=3.43 ms, TI = 900 ms, FA = 9 deg, FOV=256×224, slice thickness=1 mm, slices=176, matrix size=256×224, no gap). An axial T2*-weighted BOLD interleaved

acquisition using gradient-echo echoplanar imaging (EPI) was also collected (TR=2000 ms, TE=34 ms, FOV=256×256, slices=28, slice thickness=4 mm, FA=90°, matrix size=128×128, no gap, 117 time points) to measure changes in blood oxygen-level dependent (BOLD) response during the task (see next section). Total scan time was one hour. Structural scans covered from the bottom of the cerebellum to the top of the head, whereas functional scans only covered from midway of the cerebellum to the top of the head.

2.3. Faces/Shapes Task

The study used the faces-shapes task to study the effect of emotional responses to faces (Hariri et al., 2003). We have further used it to study the effect of WMH on functional coregistration due to its activation of some subcortical structures that may be affected by WMH. During the task, subjects are required to select one of two facial expressions (angry/fearful and neutral) that match that of a simultaneously presented target expression. As a control task, subjects matched one of two geometric shapes with a control shape. The matching task (5 blocks) was interleaved with 4 blocks of the experimental faces task. Each block lasted 24 seconds and contains 6 trials lasting 4 seconds each.

Before the beginning of each block, a brief instruction (“match emotion” or “match form”) is presented for 2 seconds. On the faces trials, 12 different images are used, six per block, three of each gender, all derived from a standard set of pictures of facial affect (Ekman et al., 1975). In the control condition, six different sets of geometric forms are used, each set including two geometric shapes. During imaging, subjects responded with button presses, allowing for determination of accuracy and reaction time. Stimulus presentation and response recording were controlled using the E-Prime software package (Psychology Software Tools, Inc., Pittsburgh, 2002).

2.4. Data Processing

All of our preprocessing used Statistical Parametric Mapping software (SPM 12) (<http://www.fil.ion.ucl.ac.uk/spm/>) as well as Python and Matlab programming platforms. The following sections are ordered temporally, thus each section precedes the following in terms of order of analysis.

The following steps generate gray matter density maps (or modulated segmentations) and contrast maps for the faces-shapes task. We compare an algorithm that segments/normalizes the MPAGE image (standard) to (1) an algorithm that fills the MPAGE with NAWM intensities and then segments/normalizes and (2) an algorithm that uses a multi-spectral approach (uses the MPAGE and FLAIR and the WMH as a spatial prior) to segment/normalize.

2.4.1. Coregistration of Functional/Structural Images and Functional Analysis

—The goal of this step is to coregister both the MPAGE and FLAIR to the functional data and then to perform slice-time and motion correction. The FLAIR was first linearly coregistered to the MPAGE. No functional volumes were discarded as the scanner already removed the dummy/stabilization volumes. Functional volumes were slice time corrected,

then realigned to correct for motion over time. A skull-stripped MPRAGE (which was done via human eye using ITK-SNAP) was then used to coregister the (not skull-stripped) MPRAGE and FLAIR to the mean functional volume. The skull stripping improves coregistration to the functional image, as it does not have a prominent skull signal.

Slice-time correction was done prior to motion-correction with the reference slice being the (temporally) middle slice with sinc interpolation. Motion correction used rigid transformations to correct to the mean functional image with mutual information as the similarity metric and 4th degree spline interpolation. Structural to functional coregistrations used rigid transformations with normalized mutual information as the similarity metric. The transformations were not applied until the normalization step.

2.4.2. Structural Segmentation—The coregistered MPRAGEs were segmented into three different tissue types: gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF), as well as 3 non-brain classes. The segmentation was implemented in SPM12 (Ashburner and Friston, 2005). The default segmentation that is implemented in SPM12 is a unified segmentation framework based on an iterative algorithm that performs tissue segmentation as well as registration and intensity non-uniformity (bias) correction. The segmentation algorithm models image intensities as a mixture of Gaussian distributions assigning a mixture for each tissue class. In this unified segmentation model SPM12 uses tissue probability maps (TPMs) as a priori information for tissue classes. The segmentation automatically segments the input image into GM, WM, and CSF. However, the classification is probabilistic as a probability value of belonging to each of the classes is assigned to each voxel in the output images. When generating binary images, voxels corresponding to greater tissue probability in the maps are counted as members of that particular class. All segmentation outputs had an output resolution of 2mm isotropic.

In this study, all of the segmentations implemented in SPM12 were performed using the default atlas and parameters for this version. This segmentation also outputs a deformation field that can be used to normalize any image (that was coregistered to the MPRAGE) to MNI space. This deformation field will be used in 2.4.3 and 2.4.4.

2.4.3. WMH Segmentation and Normalization—The WMH mask was obtained for each subject from the coregistered T2-weighted FLAIR image using an automated method for quantification and localization of WMH (Wu et al., 2006). The WMH quantification was done using a fuzzy connected algorithm. The automated WMH segmentation method is an iterative algorithm that involves an automated selection of “seeds” of possible WMH lesions and fuzzy connectedness, which clusters voxels based on their adjacency and affinity, to segment WMH lesions around the seeds. The fully automated WMH segmentation system was implemented in C++ and ITK. The WMH segmentation was normalized to MNI space (2mm isotropic) using the deformation field that was generated in 2.4.2.

A prevalence map was generated to demonstrate the group’s distribution of WMHs. This was done by computing (for each voxel) the number of subjects for whom the voxel was classified as WMH divided by the total number of subjects in the sample.

2.4.4. Functional Image Normalization and General Linear Models—The slice-time and motion corrected fMRI data was normalized based on the deformation field in 2.4.2. The fMRI data was spatially smoothed with a Gaussian kernel with full-width half-maximum (FWHM) of 8 mm.

A general linear model was run modeling the effects of faces, shapes, the mean of the scan, and 6 motion terms. Faces and shapes were convolved with a canonical hemodynamic response function. An AR(1) term and high-pass filter was used (1/128 Hz). A contrast showing the effect of faces greater than shapes was generated for each subject.

2.4.5. MPAGE WMH Filling and Structural Segmentation—On the coregistered MPAGE, the WMH were filled with NAWM intensities from the MPAGE (averaged from each individual slice), and the filled MPAGE was segmented again (see 2.4.2). This was done by first (for each slice) calculating the mean and standard deviation (divided by two) of each slice for NAWM intensities. We then generated a normal distribution with those features and randomly sampled from such a distribution to get the values used to perform the filling (adapted from (Valverde et al., 2014)). Filling WMH with NAWM intensities reduce the probability of incorrectly classifying a voxel as gray matter (Eloyan et al., 2014). Normal-appearing white matter was defined as voxels that were classified as white matter but not classified as WMH.

The complete process consisted of the following steps. For each subject: 1) A mask of WMH was created; 2) The average values of NAWM were computed for each slice; 3) The WMH masked area was filled with the average value of NAWM for that specific slice; 4) The structural segmentation was performed again which generated a deformation field that was used in 2.4.6. These segmentations were confirmed by human eye.

2.4.6. Functional Image Normalization via Filled MPAGE and GLM—The slice-time and motion corrected fMRI data was normalized based on the deformation field in 2.4.5. Then spatial smoothing and a general linear model were performed (see 2.4.4).

2.4.7. MPAGE Multi-Spectral Structural Segmentation: WMH Priors—The multi-spectral approach uses both the coregistered MPAGE and FLAIR as inputs into SPM12's segmentation. The WMH segmentation is used as a spatial prior in the algorithm. In this segmentation, the information from the FLAIR and MPAGE are used to segment gray, white, CSF, (plus 3 non-brain tissues) and also considers the WMH as a separate class (thus 7 probability maps are generated). This segmentation output a deformation field that was used in 2.4.8.

2.4.8. Functional Image Normalization via Multi-Spectral Approach and GLM—The slice-time and motion corrected fMRI data was normalized based on the deformation field in 2.4.7. Then spatial smoothing and a general linear model were performed (see 2.4.4).

2.4.9. Processing Outputs—From section 2.4.2 a modulated (probability map is multiplied by the Jacobian) gray matter density map is output (standard). A filled version of this map was output in section 2.4.5 and a multi-spectral version of this map was output in

section 2.4.7. These will be used to generate structural intra-class correlation coefficients (ICC's).

Similarly, a faces-shapes contrast was generated using the standard normalization in section 2.4.4 (standard). A filled version of this contrast map was output in section 2.4.6 and a multi-spectral version was output in section 2.4.8. These were used to generate functional intra-class correlation coefficients (ICC's).

2.5. Intraclass Correlation Coefficients (ICC's) and Assessment of Segmentation/Normalization

The intra-class correlation coefficient (ICC(A,1): degree of absolute agreement among measurements or criterion-referenced reliability) was computed for each voxel to compare the standard gray matter and contrast outputs to the filled and multi-spectral versions (McGraw and Wong, 1996). ICC measures the reliability between two ratings, but it can also be used to measure the agreement between measurements. We performed ICC's to compare the standard outputs with filled and multi-spectral approaches (for both functional and structural data). At each voxel we computed an ICC measure:

$$ICC = \frac{MSR - MSE}{MSR + (k - 1) * MSE + k * \frac{MSC - MSE}{n}}$$

$$MSR = \frac{n * \sum (\bar{y}_{i.} - \bar{\bar{y}})^2}{k - 1}$$

$$MSC = \frac{k * \sum (\bar{y}_{.j} - \bar{\bar{y}})^2}{n - 1}$$

$$MSE = \frac{\sum (y_{ij} - \bar{\bar{y}})^2 - n * \sum (\bar{y}_{i.} - \bar{\bar{y}})^2 - k * \sum (\bar{y}_{.j} - \bar{\bar{y}})^2}{n - 1}$$

where MSR is the mean squares of regression (rows: subjects), MSC is the mean squares of regression (columns: standard and filled or standard and multi-spectral), MSE is the mean square error, k is the number of columns (2 values in this analysis), and n is the number of subjects. At each voxel, we have a matrix Y that is n by k (i.e. 27 subjects by 2 methods: either standard and WMH fill or multi-spectral approach). The bars indicate a mean and the subscripts i or j indicate the direction (mean across method and mean across subjects, respectively), and double bars indicate the global mean. Note that MSR and MSC numerators are sum of regressions (SSR) and sum of columns (SSC), respectively, and that MSE's numerator is sum of total (SST) minus SSR and SSC.

This resulted in two maps where each voxel contains one ICC value. The displayed values at each voxel are equal to 1-voxel (ICC), thus values that perfectly agree are actually equal to zero. This was done for illustration purposes, since higher ICC values (>0.8) (McGraw and Wong, 1996) indicate good agreement (while we are interested in areas of low agreement). The map displays higher (1-ICC) values in the areas of the brain where the WMH are expected.

We also performed an analysis where we investigated global ICC as well as local ICC. For the global ICC, we calculated the mean and standard deviation (across voxels) of the ICC (comparing the standard to the fill and multi-spectral approach) in the gray matter (MNI tissue probability map probability 0.3), white matter (MNI tissue probability map probability 0.3 excluding areas labeled as WMH), and WMH (prevalence map 0.3) for both structural and functional ICC's. We also investigated local ICCs in four regions (left and right): amygdala, caudate, fusiform, and insula. All regions except the caudate are regions significantly activated by the faces-shapes task. The caudate, amygdala, and insula have varying levels of WMH burden (caudate usually being the greatest). These were extracted for both functional and structural ICC. The regions are from the automated anatomical labeling atlas (AAL) (Tzourio-Mazoyer et al., 2002).

3. RESULTS

Figure 1 shows the WMH prevalence map (threshold at values greater than 0.3 or 30% prevalence in the sample). There is high prevalence in the white matter near the ventricles as well as other subcortical structures (e.g. caudate). Structural agreement was calculated to investigate whether gray matter density estimates were impacted by either WMH filling or multi-spectral approach for segmentation. Structural agreement (1-ICC) between the gray matter density estimates of the standard approach and the WMH fill approach as well as the multi-spectral approach are shown in figure 2 (left/right, respectively) threshold at values greater than 0.6 (to show areas of lowest agreement). Structural agreement was impacted in areas of high WMH burden. However both methods showed that by performing some correction areas outside the WMH are also affected. The multi-spectral approach affects the gray matter density estimates to an even greater extent spatially.

Functional agreement was calculated to investigate whether contrast values (faces-shapes) were impacted by either WMH filling or multi-spectral approach for normalization. Functional agreement (1-ICC) between the contrast faces-shapes normalized by the deformation field of the standard method and the WMH fill method as well as the multi-spectral approach are shown in figure 3 (left/right, respectively), threshold at values greater than 0.01. Unlike structural agreement, functional agreement was fairly high (ranging from 0–0.09 and 0–0.2 for the two methods). Thus, contrast values were not affected: however, similar to the structural agreement, functional agreement was, in general, lower in the multi-spectral approach compared to the WMH filling approach.

Global and local (4 ROI's) structural/functional agreement is plotted [mean and (standard deviation)] in figure 4. Global structural agreement for gray/white matter and WMH was, in general, higher for the WMH filling approach [0.05 (0.09), 0.13 (0.23), and 0.56 (0.33),

respectively] than the multi-spectral approach [0.31 (0.19), 0.30 (0.26), and 0.77 (0.26), respectively], shown in figure 4A. WMH had lowest agreement compared to gray/white matter. Global functional agreement was relatively similar for gray/white matter and WMH for both the WMH filling approach [$1*10^{-3}$ ($1.7*10^{-3}$), $1.8*10^{-3}$ ($2.9*10^{-3}$), and $7.3*10^{-3}$ ($9.6*10^{-3}$), respectively] and the multi-spectral approach [$5.2*10^{-3}$ (0.01), $4.4*10^{-3}$ ($7.2*10^{-3}$), and $2.7*10^{-3}$ ($2.7*10^{-3}$), respectively], shown in figure 4B. They show high agreement.

Local structural agreement for the WMH filling approach for the left/right amygdala [0.09/0.13 (0.09/0.14)], caudate [0.38/0.45 (0.29/0.33)], fusiform gyrus [0.03/0.05 (0.03/0.08)], and insula [0.10/0.08 (0.18/0.12)] as well as for the multi-spectral approach for the left/right amygdala [0.46/0.36 (0.22/0.25)], caudate [0.52/0.53 (0.31/0.35)], fusiform gyrus [0.28/0.28 (0.20/0.21)], and insula [0.24/0.22 (0.19/0.18)] varied greatly depending on their WMH spatial proximity. Caudate was greatly affected as it is spatially close to the ventricles and areas affected by WMH's. In general, the multi-spectral approach had lower agreement than the fill approach (compared to the standard approach).

Local functional agreement for the WMH filling approach left/right amygdala [$1*10^{-3}$ / $1.2*10^{-3}$ ($1*10^{-3}$ / $1*10^{-3}$)], caudate [$5.8*10^{-3}$ /0.01 ($5.6*10^{-3}$ /0.01)], fusiform gyrus [$1.0*10^{-3}$ / $1*10^{-3}$ ($1.0*10^{-3}$ / $1*10^{-3}$)], and insula [$1.8*10^{-3}$ / $2.3*10^{-3}$ ($1.9*10^{-3}$ / $3.9*10^{-3}$)] as well as for the multi-spectral approach for the left/right amygdala [$1.9*10^{-3}$ / $2.4*10^{-3}$ ($2.4*10^{-3}$ / $2.1*10^{-3}$)], caudate [$2.7*10^{-3}$ / $3.0*10^{-3}$ ($2.0*10^{-3}$ / $2.1*10^{-3}$)], fusiform gyrus [$5.5*10^{-3}$ / $4.1*10^{-3}$ (0.01/0.01)], and insula [$2.0*10^{-3}$ / $2.8*10^{-3}$ ($1.3*10^{-3}$ / $1.7*10^{-3}$)] was very high.

4. DISCUSSION

In aged populations, the presence of WMH is often unaccounted for in both structural and functional brain imaging studies. This can lead to improper segmentation/normalization (to a standard space) that can result in findings that would otherwise be accounted for by WMH. In this study, we demonstrate that accounting for WMH may affect the results of structural but not functional MRI analyses.

Previous studies have found that WMH often lead to misclassification (Levy-Cooperman et al., 2008). A previous study showed that WMH filling could be a useful way to account for WMH when performing segmentation (Eloyan et al., 2014). This is one method for dealing WMH that is relatively simple, however there are other methods that are much more complex but may yield a better result. Multi-spectral segmentation performs a segmentation using multiple scans (e.g. MPRAGE and FLAIR) that can improve certain aspects of the segmentation. In this study we performed comparisons (using ICCs) between standard structural segmentations and WMH filled structural segmentations as well as multi-spectral structural segmentations. We also looked at the effect of standard normalization and WMH filled normalization as well as multi-spectral normalization on fMRI contrasts (faces-shapes) by performing ICCs in a similar manner. ICCs demonstrate the effect each method had on the original uncorrected approach and do not measure efficacy.

WMH affected the segmentation/normalization of structural images, but had little to no effect on the contrast values from the functional data. This indicates that it may be more important to account for WMH in some way when studying gray matter density (or structural measures) than when studying functional measures. This was, in general, true for both of the approaches used.

In regions that were classified as WMH the structural agreement (for both approaches) was lower compared to the gray matter or white matter (excluding WMH). The structural ICCs indicate that, if left uncorrected, areas outside of WMH are also affected during the segmentation/normalization procedure. This may result in spurious cortical and subcortical gray matter differences, which could be accounted for by WMH. Importantly, the multi-spectral approach seems to also affect regions of cortical gray matter. Accounting for WMH may be even more important when considering investigating certain regions, especially subcortical regions that are spatially closer to the ventricles (e.g. caudate). Our results show that for the caudate, agreement between the standard approach and either of the corrected approaches is affected. We also show that other regions (e.g. amygdala) are more affected when using the multi-spectral approach. These may also explain some results in the literature that rely on total gray matter, as, without correction, total gray matter will be greater in populations affected by WMH (as they are misclassified as gray matter).). Therefore, it is possible (in conditions such as depression) that the total gray matter differences between two groups (e.g. healthy vs. depressed) may be due to differences WMH burden rather than total gray matter.

Functional agreement was very high, and was not as strongly affected either globally or locally. The functional agreement was relatively lower when using the multi-spectral approach compared to the WMH filling approach. Furthermore, the WMH filling approach seemed to only affect functional agreement near WMH, whereas the multi-spectral approach affected mostly gray matter. We did not limit this analysis to regions that were significantly activated by the task and thus the agreement in certain regions (especially white matter) may be agreement between two values closer to zero (e.g. noise rather than task contrast).

This study had several limitations. It had a limited sample size and only one task was used to look at the effect of WMH on parameter estimates in functional MR studies. We suggest that future studies should investigate this effect with tasks that directly activate structures like the caudate, which are more greatly affected by WMH. Finally, our sample was limited to LLD patients. Thus, these results may not generalize to a different population, especially healthy aging populations. Further analyses of these effects are needed. Future studies should also investigate the efficacy of WMH filling compared to the multi-spectral segmentation approach.

In conclusion, this study demonstrated that accounting for WMH (using a simpler WMH filling approach or a more complex multi-spectral approach) modifies the results of structural MR analysis, a conclusion that warrants changes in the current standard analysis pipelines for MRI data. It is clear from these results that accounting for WMH can significantly affect the outcome of gray matter density estimates and thus is needed in future studies where WMH are a significant issue (e.g. when studying aging populations).

Acknowledgments

We would like to thank Ani Eloyan, PhD for her helpful comments during the editing of the manuscript. This work was supported by 5R01 MH076079, NIMH R01 MH076079, 5R01 AG033575, MH 086686, P30 MH90333, R01 080240.

References

- Alexopoulos GS. The vascular depression hypothesis: 10 years later. *Biol Psychiatry*. 2006; 60:1304–1305. [PubMed: 17157096]
- Ashburner J, Friston KJ. Unified segmentation. *Neuroimage*. 2005; 26:839–851. [PubMed: 15955494]
- Battaglini M, Jenkinson M, De Stefano N. Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Hum Brain Mapp*. 2012; 33:2062–2071. [PubMed: 21882300]
- Breteler MM, van Swieten JC, Bots ML, Grobbee DE, Claus JJ, van den Hout JH, van Harskamp F, Tanghe HL, de Jong PT, van Gijn J, et al. Cerebral white matter lesions, vascular risk factors, and cognitive function in a population-based study: the Rotterdam Study. *Neurology*. 1994; 44:1246–1252. [PubMed: 8035924]
- Chard DT, Jackson JS, Miller DH, Wheeler-Kingshott CA. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. *J Magn Reson Imaging*. 2010; 32:223–228. [PubMed: 20575080]
- de Leeuw FE, de Groot JC, Achten E, Oudkerk M, Ramos LM, Heijboer R, Hofman A, Jolles J, van Gijn J, Breteler MM. Prevalence of cerebral white matter lesions in elderly people: a population based magnetic resonance imaging study. The Rotterdam Scan Study. *J Neurol Neurosurg Psychiatry*. 2001; 70:9–14. [PubMed: 11118240]
- Ekman P, Friesen WV, Press CP. Pictures of facial affect. consulting psychologists press. 1975
- Eloyan A, Shou H, Shinohara RT, Sweeney EM, Nebel MB, Cuzzocreo JL, Calabresi PA, Reich DS, Lindquist MA, Crainiceanu CM. Health effects of lesion localization in multiple sclerosis: spatial registration and confounding adjustment. *PLoS One*. 2014; 9:e107263. [PubMed: 25233361]
- First MB, Spitzer RL, Gibbon M, Williams JB. User's guide for the Structured clinical interview for DSM-IV axis I disorders SCID-I: clinician version. American Psychiatric Pub. 1997
- Hariri AR, Mattay VS, Tessitore A, Fera F, Weinberger DR. Neocortical modulation of the amygdala response to fearful stimuli. *Biol Psychiatry*. 2003; 53:494–501. [PubMed: 12644354]
- Herrmann LL, Le Masurier M, Ebmeier KP. White matter hyperintensities in late life depression: a systematic review. *J Neurol Neurosurg Psychiatry*. 2008; 79:619–624. [PubMed: 17717021]
- Lenze EJ, Mulsant BH, Blumberger DM, Karp JF, Newcomer JW, Anderson SJ, Dew MA, Butters MA, Stack JA, Begley AE, Reynolds CF 3rd. Efficacy, safety, and tolerability of augmentation pharmacotherapy with aripiprazole for treatment-resistant depression in late life: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2015
- Levy-Cooperman N, Ramirez J, Lobaugh NJ, Black SE. Misclassified tissue volumes in Alzheimer disease patients with white matter hyperintensities: importance of lesion segmentation procedures for volumetric analysis. *Stroke*. 2008; 39:1134–1141. [PubMed: 18323507]
- Magon S, Gaetano L, Chakravarty MM, Lerch JP, Naegelin Y, Stippich C, Kappos L, Radue EW, Sprenger T. White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: a longitudinal study. *BMC Neurosci*. 2014; 15:106. [PubMed: 25200127]
- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological methods*. 1996; 1:30.
- Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*. 1971; 9:97–113. [PubMed: 5146491]
- Ovbiagele B, Saver JL. Cerebral white matter hyperintensities on MRI: Current concepts and therapeutic implications. *Cerebrovasc Dis*. 2006; 22:83–90. [PubMed: 16685119]
- Sdika M, Pelletier D. Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. *Hum Brain Mapp*. 2009; 30:1060–1067. [PubMed: 18412131]

- Taylor WD, Aizenstein HJ, Alexopoulos GS. The vascular depression hypothesis: mechanisms linking vascular disease with depression. *Mol Psychiatry*. 2013; 18:963–974. [PubMed: 23439482]
- Tudorascu DL, Rosano C, Venkatraman VK, MacCloud RL, Harris T, Yaffe K, Newman AB, Aizenstein HJ. Multimodal MRI markers support a model of small vessel ischemia for depressive symptoms in very old adults. *Psychiatry Res*. 2014; 224:73–80. [PubMed: 25205441]
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 2002; 15:273–289. [PubMed: 11771995]
- Valverde S, Oliver A, Llado X. A white matter lesion-filling approach to improve brain tissue volume measurements. *Neuroimage Clin*. 2014; 6:86–92. [PubMed: 25379419]
- Wu M, Rosano C, Butters M, Whyte E, Nable M, Crooks R, Meltzer CC, Reynolds CF 3rd, Aizenstein HJ. A fully automated method for quantifying and localizing white matter hyperintensities on MR images. *Psychiatry Res*. 2006; 148:133–142. [PubMed: 17097277]
- Xiong YY, Mok V. Age-related white matter changes. *J Aging Res*. 2011; 2011:617927. [PubMed: 21876810]

Figure 1.

WMH prevalence map in this sample (calculated at each voxel as number of subjects divided by the number of subjects) overlaid on a single subject in MNI space. The image is threshold at prevalence greater than 0.3 (greater than 30% of the sample). It can be seen that the highest prevalence occurs near the ventricles and certain parts of the white matter.

Figure 2.

Structural agreement (1-ICC) maps overlaid on a single subject in MNI space threshold at ICC greater than 0.6. Structural agreement compares the gray matter density maps output by the segmentation. Low values indicate high agreement and high values indicate low agreement (see color bar for reference). (Left) Structural ICC between the standard approach and the WMH fill approach and (Right) structural ICC between the standard approach and the multi-spectral approach.

Figure 3.

Functional agreement (1-ICC) maps overlaid on a single subject in MNI space threshold at ICC greater than 0.01. Functional agreement compares the contrast maps (faces-shapes) normalized via the different methods. Low values indicate high agreement and high values indicate low agreement (see color bar for reference). (Left) Functional ICC between the standard approach and the WMH fill approach for normalization and (Right) functional ICC between the standard approach and the multi-spectral approach for normalization. NOTE: Color bar range between methods varies (0–0.09 and 0–0.2 for fill and multi-spectral approach, respectively). This is to maximize contrast (due to fairly high agreement, closer to zero, in the fill method relative to the multi-spectral approach).

Figure 4.

Global and local structural/functional agreement for ICC between standard approach and fill approach (red) as well as multi-spectral approach (blue). Mean ICC and standard deviations are plotted. Values plotted are 1-ICC (high values indicate low agreement). (A) Global structural agreement in the gray matter, white matter (excluding WMH), and WMH (defined as voxels greater than 30% WMH prevalence in this sample). (B) Global functional agreement in the gray matter, white matter, and WMH. Note the range of values (0–0.015) indicating high agreement. (C) Local structural agreement in the left/right amygdala, caudate, fusiform, and insula. (D) Local functional agreement in the regions listed. Note the range of values (0–0.02) indicating high agreement.