



White matter hyperintensities segmentation: a new semi-automated method

Mariangela Iorio¹, Gianfranco Spalletta¹, Chiara Chiapponi¹, Giacomo Luccichenti², Claudia Cacciari¹, Maria D. Orfei¹, Carlo Caltagirone^{1,3} and Fabrizio Piras^{1*}

¹ Neuropsychiatry Laboratory, Department of Clinical and Behavioral Neurology, Istituto di Ricovero e Cura a Carattere Scientifico Santa Lucia Foundation, Rome, Italy

² Department of Radiology, Istituto di Ricovero e Cura a Carattere Scientifico Santa Lucia Foundation, Rome, Italy

³ Department of Neuroscience, Tor Vergata University, Rome, Italy

Edited by:

Lea T. Grinberg, University of California, USA

Reviewed by:

Bogdan O. Popescu, University Hospital Bucharest, Romania
Edson Amaro Jr., Universidade de São Paulo, Brazil

*Correspondence:

Fabrizio Piras, Neuropsychiatry Laboratory, Department of Clinical and Behavioral Neurology, Istituto di Ricovero e Cura a Carattere Scientifico Santa Lucia Foundation, Via Ardeatina 306, Rome 00179, Italy
e-mail: f.piras@hsantalucia.it

White matter hyperintensities (WMH) are brain areas of increased signal on T2-weighted or fluid-attenuated inverse recovery magnetic resonance imaging (MRI) scans. In this study we present a new semi-automated method to measure WMH load that is based on the segmentation of the intensity histogram of fluid-attenuated inversion recovery images. Thirty patients with mild cognitive impairment with variable WMH load were enrolled. The semi-automated WMH segmentation included removal of non-brain tissue, spatial normalization, removal of cerebellum and brain stem, spatial filtering, thresholding to segment probable WMH, manual editing for correction of false positives and negatives, generation of WMH map, and volumetric estimation of the WMH load. Accuracy was quantitatively evaluated by comparing semi-automated and manual WMH segmentations performed by two independent raters. Differences between the two procedures were assessed using Student's *t*-tests and similarity was evaluated using linear regression model and Dice similarity coefficient (DSC). The volumes of the manual and semi-automated segmentations did not statistically differ (*t*-value = -1.79 , $DF = 29$, $p = 0.839$ for rater 1; *t*-value = 1.113 , $DF = 29$, $p = 0.2749$ for rater 2), were highly correlated [$R^2 = 0.921$, $F_{(1,29)} = 155.54$, $p < 0.0001$ for rater 1; $R^2 = 0.935$, $F_{(1,29)} = 402.709$, $p < 0.0001$ for rater 2] and showed a very strong spatial similarity (mean DSC = 0.78 , for rater 1 and 0.77 for rater 2). In conclusion, our semi-automated method to measure the load of WMH is highly reliable and could represent a good tool that could be easily implemented in routinely neuroimaging analyses to map clinical consequences of WMH.

Keywords: white matter hyperintensities, lesion segmentation, MRI, FLAIR, MCI

INTRODUCTION

White matter hyperintensities (WMH) are brain areas of increased signal intensity, appearing on T2-weighted (T2-w) or fluid-attenuated inversion-recovery (FLAIR) magnetic resonance imaging (MRI) scans, that result from localized changes in tissue composition (Barkhof and Scheltens, 2002). The etiology of WMH is not specific and may be secondary to ischemia, demyelinating disorders, hydrocephalus, trauma, inflammatory diseases, radiation injury, amyloidosis, and other causes (Pantoni and Garcia, 1997).

WMH are often observed in elderly participants (de Groot et al., 2000a), and their frequency is greater in individuals with cerebrovascular risk factors such as diabetes and hypertension (de Leeuw et al., 2001). Furthermore, WMH are a common finding in those subjects with neurological illnesses, such as stroke (Fazekas et al., 1993), Parkinson's disease (Marshall et al., 2006), mild cognitive impairment (MCI; DeCarli et al., 2001), Alzheimer's disease (Fazekas et al., 1996; Hirono et al., 2000), and even in patients with primary mental disorders including mood disorders and schizophrenia spectrum disorders (Brown et al., 1995).

Although neuropathological, clinical, and cognitive significance of WMH is unclear (Gunning-Dixon and Raz, 2000) large epidemiological studies provide evidence that WMH have a strong impact on cognitive functioning (Gunning-Dixon and Raz, 2000) and they have been associated with impairment in a number of domains, including psychomotor speed, frontal executive functions (de Groot et al., 2000b; Prins et al., 2004), and explicit memory (de Groot et al., 2000a).

Interestingly, a systematic review and meta-analysis (DeBette and Markus, 2010) provides strong evidence that WMH may be an important predictor of future disease, being associated with an increased risk of stroke, dementia, and mortality. Moreover, it has been shown that WMH often occur in preclinical stages of dementia (such as in MCI patients) and that the presence of WMH may also increase the likelihood of progression from MCI to dementia (Wu et al., 2002).

A challenging issue in studying WMH is their quantification and localization, given their variability and scattered spatial distribution. Two analytic strategies have been used to evaluate WMH on MRI brain images: (1) semi-quantitative rating systems and (2)

quantitative volumetric analyses. The semi-quantitative approach mainly consists in the computation of visual ratings and several scales, markedly different in morphological or anatomical definitions, have been proposed and used in literature (Fazekas et al., 1987; Scheltens et al., 1993; Mäntylä et al., 1997; Wahlund et al., 2001). WMH visual rating scales are relatively quick and easy to perform even across scans and scanners, and they are commonly used in clinical and research settings, thus making them attractive for large epidemiological studies. Unfortunately, they have a number of limitations. Indeed, categorical ratings have a restricted range of values that limit the power of association. Moreover, qualitative scales are often subjective in their interpretation, thus limiting inter-rater and intra-rater reliability (Mäntylä et al., 1997) and consistency within longitudinal studies (Prins et al., 2004).

On the other hand, a number of more recent researches have introduced quantitative methods of measuring WMH severity using computer-based techniques to obtain volumetric measures of WMH burden. These methods vary from manual outlining techniques to fully-automatic WMH detection (Jack et al., 2001; Anbeek et al., 2004; Wen and Sachdev, 2004; Admiraal-Behloul et al., 2005; Wu et al., 2006; Gibson et al., 2010; Cerasa et al., 2012; Simões et al., 2013).

Manual outlining techniques are referred to as region-of-interest (ROI) methodologies, in which the tracer inspects the scan using visualization softwares and manually draws WMH areas. The volume of each region is then calculated from the section thickness and the number of voxels included in the traced area. The values of all sections are then added together to make a total WMH volume. Manual outlining procedures, although accurate, have several limitations such as labor intensiveness, time consume, subjectiveness, and error-proneness. In addition, manual detection suffers from intra-observer and inter-observer variability (Smith et al., 2002).

In recent years, major progresses have been made on the development of semi- or fully-automated segmentations of WMH. These techniques are based on computer algorithms developed to replace human eye, allowing the collection of quantitative volumetric data on WMH (Jack et al., 2001; Anbeek et al., 2004; Wen and Sachdev, 2004; Admiraal-Behloul et al., 2005; Wu et al., 2006; Gibson et al., 2010; Cerasa et al., 2012; Simões et al., 2013). These techniques are more objective and free from user bias compared to the visual WMH ratings (Payne et al., 2002).

However, such methods have different accuracy, computational speed, and complexity. The majority of these approaches relies on multimodal data, thus being based on different MRI sequences, including T2-w, proton density (PD), T1-weighted (T1-w), inversion recovery (IR) and, more widely, FLAIR images. Innovative procedures to detect WMH have been developed based on Markov random field model (Schwarz et al., 2009), k-nearest neighbor (Anbeek et al., 2004; Wen and Sachdev, 2004), and neural classification (Dyrby et al., 2008), which require training images with WMH labels. The segmentation accuracy of these methods depends on the representative training data that may be difficult to select due to the heterogeneous nature of WMH. Differently, Admiraal-Behloul et al. (2005) proposed a fuzzy inference system to classify the WMH based on both anatomical

locations and intensity values from three different MR images (T2-w, PD, and FLAIR) without the need of training samples. Moreover, some methods automatically or semi-automatically segment the WMH relying exclusively on FLAIR images by defining a cut-off threshold on the images (Jack et al., 2001; Wu et al., 2006; Gibson et al., 2010; Simões et al., 2013). In this context, Jack et al. (2001) and Gibson et al. (2010) employed empirical thresholds before applying linear fitting or fuzzy clustering to segment WMH. Similarly, Wen and Sachdev (2004) used the mean and standard deviation (SD) of gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) intensities to estimate the threshold for WMH and they used a WM probability map to identify the most likely WM regions. Recently, Simões et al. (2013) proposed a WMH segmentation method based on a modified context-sensitive Gaussian mixture model to determine voxel class probabilities, followed by a false positive correction step, where common FLAIR artifacts are eliminated from the segmentation.

All the quantitative methods here discussed have some limitations. Indeed, complex automatic or semi-automatic computer-based segmentation procedures require a large amount of technical resources that may not be available in clinical settings. Furthermore, multi-spectral approaches are not always accessible in clinical practice since the acquisition of all these images is cost-intensive and requires long processing time (Shen et al., 2008). Conversely, FLAIR-based approaches may, in sporadic cases, overestimate the WMH load due to FLAIR typical high intensity appearance in cortical areas, such as the septum pellucidum, and low artifacts in the fourth ventricle where a large percentage of false positive is detected (Wang et al., 2012). Such limitations make these methods scarcely applicable in large clinical contexts and barely replicable across sites.

In this study, we present a semi-automated method for WMH load measurement that overcomes the above presented limits. We applied the method on a group of MCI patients with various WMH load. The aims of our study are: (i) to develop a semi-automated algorithm for the detection, quantification, and localization of WMH using only FLAIR images and T1-w at 3T which is fully reproducible in different research contexts and clinical populations; and (ii) to evaluate the accuracy of the method by comparing it with manual segmentation.

MATERIALS AND METHODS

SUBJECTS

We recruited in our memory clinic in Rome, Italy, 30 patients with amnesic mild cognitive impairment (a-MCI; 21 males, 9 females; mean age \pm SD = 72 ± 6 years, range 56–85; mean education = 9 ± 4 years; mean mini mental state examination (MMSE; Folstein et al., 1975; score = 26 ± 2) with a variable degree of WMH load. Specific inclusion criteria for a-MCI were (1) diagnostic evidence of a-MCI consistent with Petersen guidelines (Petersen et al., 1999), i.e., (i) complaint of defective memory, (ii) normal activities of daily living, (iii) abnormal memory function for age, and (iv) absence of dementia; and (2) suitability for MRI scanning.

Exclusion criteria were (1) major medical illnesses and autoimmune-inflammatory disease; (2) co-morbidity of primary

psychiatric or neurological disorders, and (3) any other significant mental or neurological disorder. Medical and psychiatric histories were obtained from each subject, and all patients underwent a series of standard clinical examinations, including physical, neurological and mental status examinations, neurocognitive tests, and brain MRI (described below). All patients underwent the first diagnostic assessment and none were taking acetylcholinesterase inhibitors or psychotropic drugs (antidepressants, benzodiazepines, and antipsychotics). The study was approved and undertaken in accordance with the guidance of our local Ethics Committee and written consent was obtained from all participants.

MRI ACQUISITION

All 30 participants underwent the same imaging protocol, which included 2D FLAIR and whole-brain high-resolution T1-w images, using a 3T Allegra MR imager (Siemens, Erlangen, Germany) with a standard quadrature head coil. All planar sequences acquisitions were obtained in the plane of the AC–PC line. Particular care was taken to center the subjects in the head coil and to restrain their movements with cushions and adhesive medical tape.

Two-dimensional FLAIR images were obtained in the axial plane (TE = 109 ms, TR = 8500 ms, slice thickness = 5 mm, slices = 24, matrix = 188 × 256, phase FOV = 0.73). Whole-brain T1-w images were obtained in the sagittal plane using a modified driven equilibrium Fourier transform (MDEFT, TE = 2.4 ms, TR = 7.92 ms, flip angle = 15°, voxel-size = 1 mm × 1 mm × 1 mm). All scans were visually inspected by a neuroradiologist (GL) with high expertise in clinical neuroimaging in order to exclude images with poor quality and motion artifacts.

A representative axial slice of FLAIR images for each patient is shown in **Figure 1**.

SEMI-AUTOMATED WMH SEGMENTATION

The semi-automated WMH segmentation algorithm was performed using FSL 4.1¹ and the VBM8 toolbox² implemented in SPM 8³ running in Matlab 2007b (MathWorks, Natick, MA, USA) and it consisted of three major steps:

- (1) Image pre-processing, including non-brain tissue removal, spatial normalization into MNI space, removal of cerebellum and brain stem, and spatial filtering;
- (2) Automated detection of WMH in FLAIR images and, if necessary, subsequent manual editing for removal of false positives;
- (3) Post-processing, including the generation of WMH map and volumetric estimation of the WMH load at the individual and sample levels.

STEP 1: PRE-PROCESSING

Image preprocessing included skull-stripping of the FLAIR images to restrict our analyses to brain tissue only reducing the risk of false

positives and to improve the accuracy and efficiency of WMH segmentation.

Removal of non-brain tissues

For each subject, the FLAIR image was co-registered to the T1-weighted image using the FLIRT tool, integrated within FSL software through full affine alignment with nearest-neighbor resampling (correlation ratio cost function). Then, VBM8 with standard options was used to segment T1-weighted images and obtain a whole binary brain mask by summing the segmented GM, WM, and CSF in native space. The mask was then multiplied by the co-registered FLAIR and T1-w images to remove non-brain tissue. The average time for this step was 15 min per subject.

Spatial normalization of the skull-stripped FLAIR images into MNI space

The T1-w skull-stripped images were then linearly registered to MNI152 T1 1 mm brain template, and the transformation matrix was applied to the skull-stripped FLAIR images. Average time to complete the task for each subject was 1 min.

Removal of cerebellum and brain stem

Since WMH are rare in cerebellum and brain stem (Wen and Sachdev, 2004), we created a mask of these brain regions in MNI space to exclude them from the analysis. The mask was created using the Harvard–Oxford subcortical and the MNI structural atlases implemented in FSL 4.1. For each subject, the mask was then multiplied for the skull-stripped and normalized FLAIR image. The time to complete the task for each subject was negligible. In order to remove residual inhomogeneities, the normalized, skull-stripped FLAIR images were smoothed with a 2-mm full width at half maximum Gaussian kernel (Ashburner and Friston, 2000). This procedure was performed using FSL `fslmaths` utility. The average time to complete the step for each subject was 1 min.

Step 1 was performed using an in-house shell script which automatically concatenates all three substeps for all subjects.

STEP 2: DETECTION OF WMH ON FLAIR IMAGES

Automatic identification of WMH on the intensity histogram of FLAIR images

For each subject we used FSL `fslstats` utility to calculate the mean and SD of the intensity of all brain voxels. In our procedure, after several trials, we optimized the process by choosing as threshold value the intensity mean +1.5 SD. This threshold was computed and applied to the FLAIR image for each subject, thus generating a map (of likely WMH) excluding all voxels having an intensity below such threshold. The time to complete the task was 1 min per subject and was implemented in a “for” cycle within a shell script.

Correction of false positives and negatives

Each generated WMH map was visually inspected and compared to the original FLAIR image in order to identify false classifications (positive or negative). These were then manually corrected (as shown in **Figure 2**, middle row) by a trained neuropsychologist

¹<http://www.fmrib.ox.ac.uk/fsl/>

²<http://dbm.neuro.uni-jena.de/vbm.html>

³www.fil.ion.ucl.ac.uk/spm

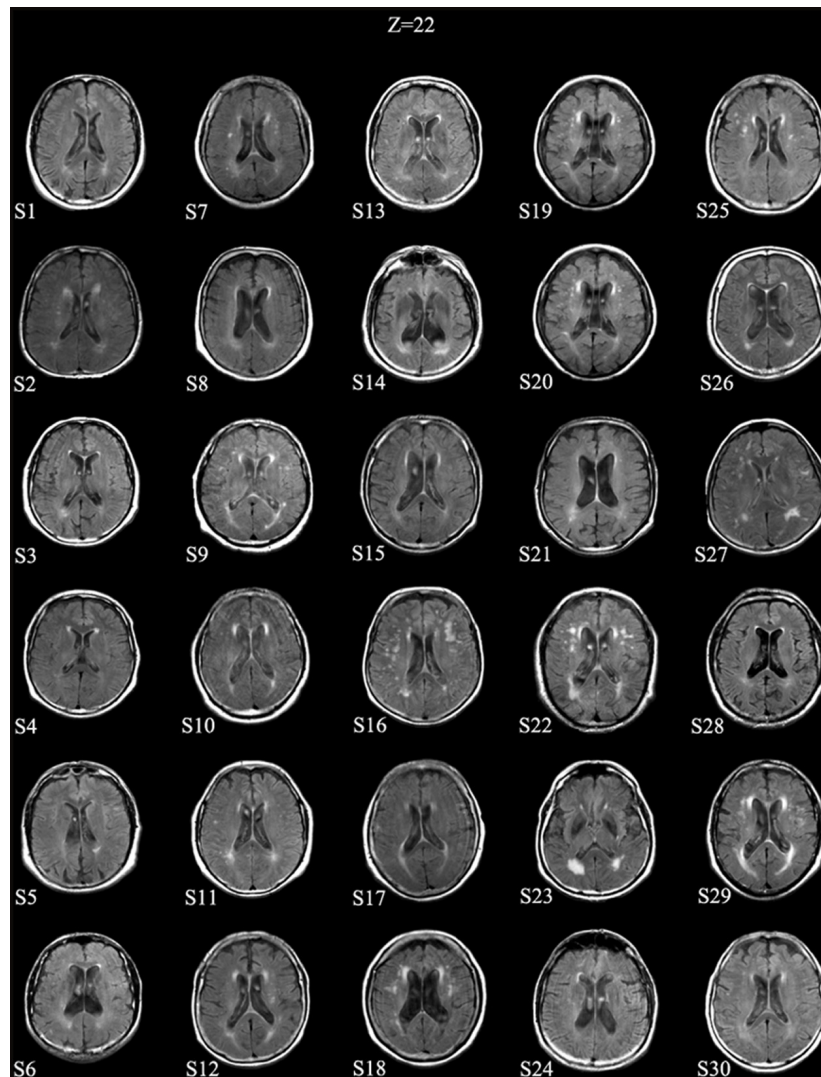


FIGURE 1 | Patients' MRI scans. A representative axial slice (at approximate MNI Z-level of 22) of FLAIR images for each patient is shown. Images are in neurological convention (left is left).

(FP) with a strong expertise in lesion analyses using the image processing software MRICron⁴. In particular, wrong identifications of WMH were deleted through the *eraser* tool of MRICron, while areas of WM lesions that were omitted by the algorithm were re-included. Correction of false positives was necessary for two patients while false negatives were present in nine patients (including the two where false positives were detected) and mainly consisted of underestimation of periventricular WMH volume. Manual correction required an average time of 8 min per patient.

STEP 3: POST-PROCESSING

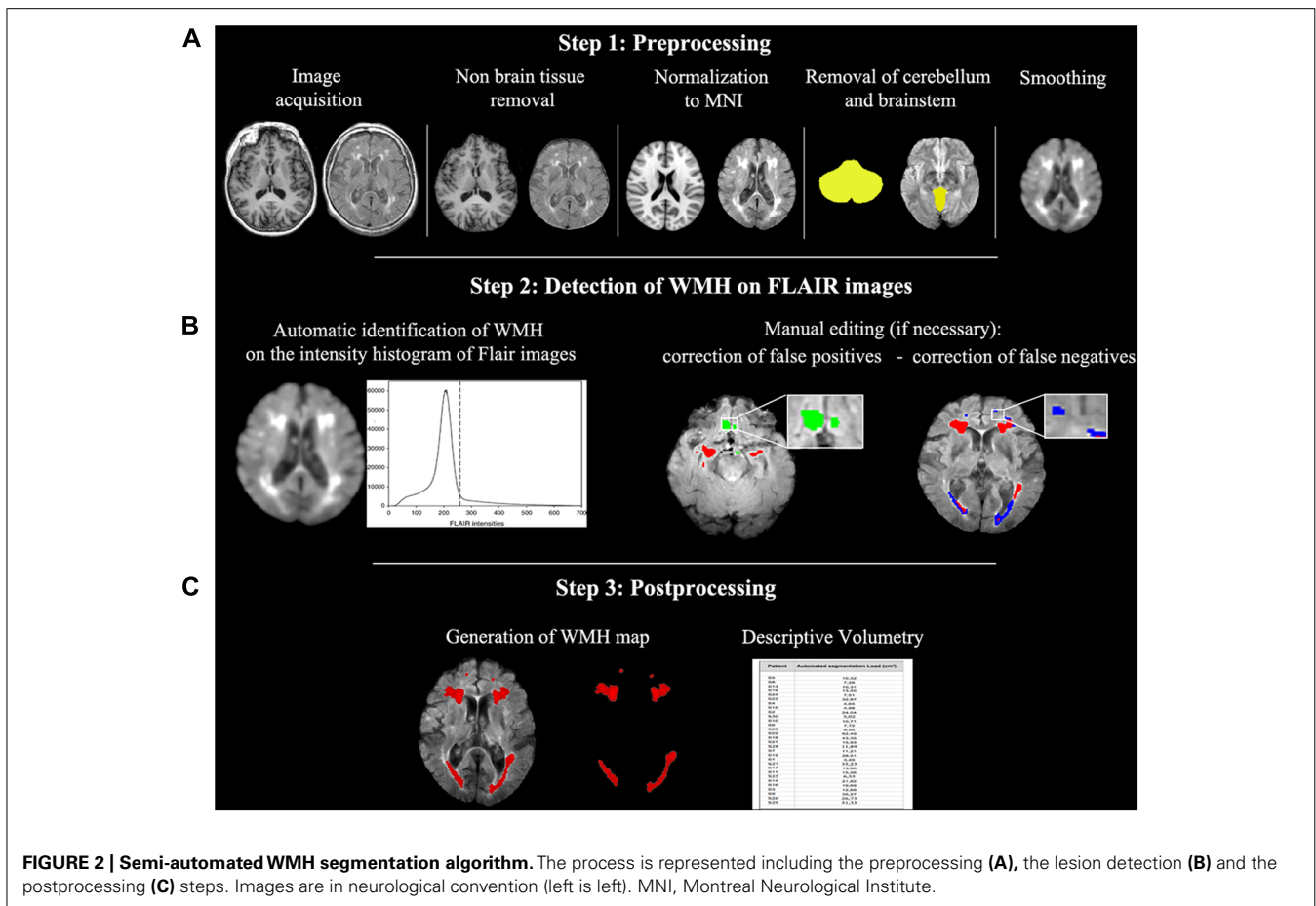
The final output provided by the system is a binary image in which a voxel is valued 1 if it is considered a WMH, 0 otherwise. Using these binary masks, for each subject, WMH volumes (expressed in

cm³) were calculated automatically using FSL *fsfstats* utility, again through an automatic shell script developed in-house.

MANUAL SEGMENTATION OF WMH

The manual segmentation of WMH on FLAIR images was performed by an expert neuroradiologist (Giacomo Luccichenti) and a trained clinician (Claudia Cacciari), expert in lesion segmentation, who were not aware of the results of the semi-automated procedure. Manual segmentation was delineated on the standard registered FLAIR images using MRICron software by tracing the lesion outline with a mouse-controlled interface. This process resulted in the definition of binary images, considered as *gold standard*. For each subject, WMH volumes (expressed in cm³) were calculated automatically using FSL *fsfstats* utility. The mean time to complete the task for each subject was 2 h and 32 min.

⁴<http://www.mccauslandcenter.sc.edu/mricro/mricron/>



STATISTICAL ANALYSES

Statistical analyses were performed with Statview software. The inter-rater reliability was calculated using the Spearman correlation coefficient. The differences between volumetric data derived from semi-automated and manual segmentations were evaluated using Student's *t*-test while a linear regression model was computed to assess the relationship between the two datasets.

Furthermore, we assessed the similarities between the two datasets in terms of both volumetric and spatial agreement by computing the Dice similarity coefficient (DSC; Dice, 1945), which is a measure of sets agreement, given by the formula:

$$DSC(\text{Seg}, \text{Ref}) = \frac{2|\text{Seg} \cap \text{Ref}|}{(|\text{Seg}| + |\text{Ref}|)},$$

where Seg is the automatically segmented image and Ref is the manually segmented image. DSC reflects differences in locations more strongly than differences in size (Zijdenbos et al., 1994). Indeed, this formula represents the size of the union of the Seg and Ref sets divided by the average size of the two sets. The DSC is commonly used when comparing datasets and, in particular, it has been widely used to assess similarities between automatic and manual segmentation procedures (Anbeek et al., 2004; Gibson et al., 2010; Cerasa et al., 2012; Simões et al., 2013). Generally, a DSC value of 0 indicates no spatial overlap and a value of 1

indicates perfect spatial alignment. Usually, DSC values of 0.7 and higher suggest good agreement between two delineations (Bartko, 1991).

RESULTS

The semi-automated procedure of WMH quantification was successfully completed in all cases. Hyperintense signals on FLAIR images were present in all subjects, but their extent and distribution varied considerably (mean volume = $18.63 \pm 14.81 \text{ cm}^3$). The mean (\pm SD) volume of WMH from the two manual raters was $16.74 (\pm 13.89) \text{ cm}^3$ and $19.5 (\pm 16.29) \text{ cm}^3$ per subject (see Table 1).

The two raters performed similarly in manual lesion segmentation ($r = 0.976, p < 0.0001$). Further, no statistical differences were found in the comparison between semi-automated and manual rater 1 WMH segmentation volumes (t -value = $-1.79, \text{DF} = 29, p = 0.839$) as well as between semi-automated and manual rater 2 WMH segmentation volumes (t -value = $1.113, \text{DF} = 29, p = 0.2749$).

Furthermore, as shown in Figure 3, the WMH volumes from the semi-automated segmentation method were highly correlated with volumes obtained through the manual method with a $r = 0.921, R^2 = 0.847, F_{(1,29)} = 155.540, p < 0.0001$ for the manual rater 1 and $r = 0.967, r^2 = 0.935, F_{(1,29)} = 402.709, p < 0.0001$ for the manual rater 2.

Table 1 | Results of semi-automated and manual segmentation of white matter hyperintensities.

Patient	WMH segmentation load (cm ³)		
	Manual rater 1	Manual rater 2	Semi-automated
S1	3.65	3.78	3.49
S2	26.38	26.43	26.95
S3	10.64	14.63	12.69
S4	5.38	5.25	4.65
S5	10.84	10.98	9.49
S6	7.97	6.73	7.72
S7	12.51	14.75	11.21
S8	7.54	7.75	7.28
S9	37.22	37.39	35.27
S10	10.52	12.57	10.71
S11	9.55	12.74	15.26
S12	27.53	33.27	25.83
S13	9.05	8.15	10.31
S14	14.32	16.92	17.64
S15	4.67	4.98	5.22
S16	20.87	22.53	19.69
S17	8.40	10.19	13.90
S18	29.22	37.54	43.35
S19	12.27	15.60	13.40
S20	6.09	8.67	9.35
S21	13.24	20.27	15.93
S22	43.11	63.30	62.49
S23	31.76	37.84	32.87
S24	7.99	8.42	7.51
S25	6.56	7.01	6.55
S26	19.12	22.67	26.73
S27	24.46	28.96	35.23
S28	9.02	11.27	11.89
S28	65.70	68.34	51.33
S30	6.62	5.98	5.02
Mean	16.74	19.50	18.63
SD	13.89	16.29	14.81

The second, the third and the fourth columns respectively show the white matter hyperintensities (WMH) volumes from the two manual labels and the semi-automated segmentation. SD, standard deviation.

Finally, the two WMH segmentation procedures showed a very strong spatial similarity, with high DSC (manual rater 1 mean = 0.78, SD = 0.10; manual rater 2 mean = 0.77, SD = 0.14; see **Table 2**).

DISCUSSION

In the present study, we showed our semi-automated procedure for the detection, localization, and quantification of WMH on

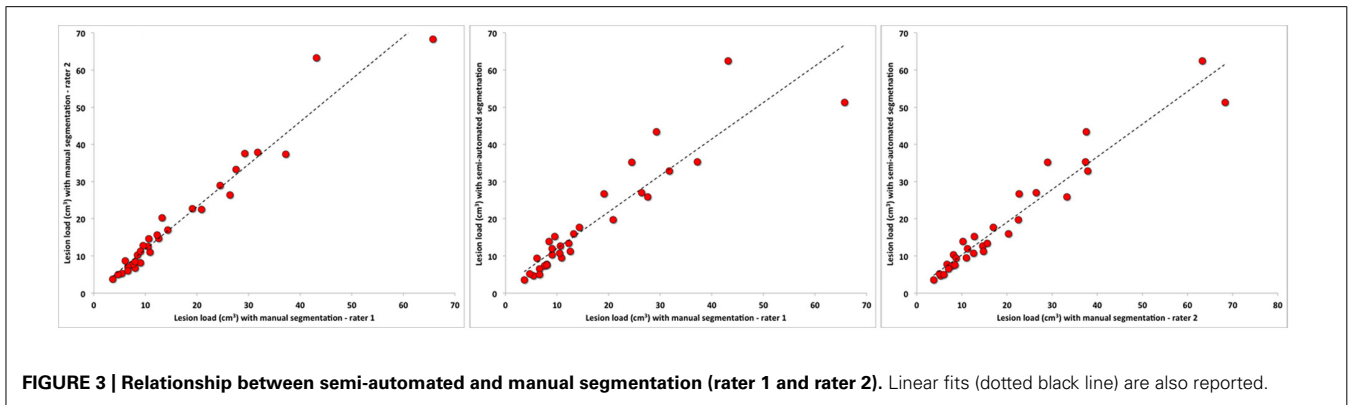
Table 2 | Relationship between semi-automated and manual segmentation of white matter hyperintensities.

Patient	DSC		
	Manual rater 1 Manual rater 2	Manual rater 1 Semi-automated	Manual rater 2 Semi-automated
S1	0.83	0.81	0.96
S2	0.79	0.70	0.70
S3	0.84	0.88	0.76
S4	0.74	0.66	0.64
S5	0.75	0.71	0.93
S6	0.72	0.70	0.85
S7	0.74	0.73	0.67
S8	0.75	0.74	0.97
S9	0.97	0.93	0.96
S10	0.80	0.67	0.59
S11	0.83	0.74	0.66
S12	0.73	0.68	0.87
S13	0.71	0.67	0.85
S14	0.72	0.70	0.68
S15	0.79	0.79	0.93
S16	0.91	0.94	0.93
S17	0.84	0.70	0.63
S18	0.82	0.71	0.61
S19	0.85	0.91	0.83
S20	0.78	0.72	0.61
S21	0.74	0.84	0.66
S22	0.80	0.80	0.67
S23	0.89	0.95	0.91
S24	0.95	0.96	0.94
S25	0.92	0.92	0.89
S26	0.87	0.79	0.71
S27	0.78	0.68	0.59
S28	0.87	0.84	0.76
S28	0.81	0.78	0.74
S30	0.70	0.71	0.57
Mean	0.81	0.78	0.77
SD	0.07	0.10	0.14

The second column shows the Dice similarity coefficient (DSC) between two manual labels. The third and the fourth columns respectively show the DSC between the manual rater 1 and the semi-automated segmentation as well as between the manual rater 2 and the semi-automated segmentation. SD, standard deviation.

FLAIR images applicable to a wide range of patients with various diseases. This procedure is based on FLAIR and T1-w images (the latter are used for preprocessing purposes only, see **Figure 2**).

Results indicate that the algorithm performed remarkably well, compared to the gold-standard (manual segmentation by experts), since no statistical differences between the two outputs were found and a very high similarity emerged, both in terms of volumetric



load and spatial location. This is an outstanding outcome, since the semi-automated procedure requires a time consumption which is approximately six times lower than the manual approach.

Other automated procedures developed to classify and quantify WMH have used a variety of classification approaches, including Markov random field model (Schwarz et al., 2009), k-nearest neighbor (Anbeek et al., 2004; Wen and Sachdev, 2004), neural classification (Dyrby et al., 2008), modified Gaussian mixture model (GMM) that incorporates neighborhood information (Simões et al., 2013) and threshold cut-offs (Jack et al., 2001; Gibson et al., 2010). Otherwise, our approach combines conservative voxel intensity thresholding with several key components that need further discussion.

First, we incorporated specific steps without any human intervention including the removal of non-brain tissue and of areas where WMH are unlikely (i.e., the cerebellum and the brainstem) in order to improve the accuracy of the method and overcome the limitations of subjectivity of the raters, thus ensuring test-retest reliability. Furthermore, these steps, which are not included in the majority of other methodologies, reduce the probability to include false positives and decrease the time spent to remove them, thus resulting in faster processing and more accurate maps of WMH.

Second, the transformation of FLAIR images from native to standard space (MNI), allows later statistical group comparisons and/or voxel-based statistics since the WMH images are in a common coordinates system. This feature enables a quantitative estimation and localization of WMH that can be further processed using statistical voxel-based image analysis softwares, which may be of importance for the study of the consequences of WM changes on cognition and behavior (Soderlund et al., 2006; Cacciari et al., 2010; Spalletta et al., 2013). Indeed, many studies suggested that the clinical significance of WMH is related not only to their volume but also to their localization (Chen et al., 2006). Indeed, since the distribution of WMH in specific functional regions may be closely related to specific types of cognitive impairment, clinical symptoms may be linked to the regional distribution of WMH. Moreover, an important implication of this automated procedure is its implementation in follow-up studies with repeated MRI scans. Allowing coregistration among different MRI sessions, our method will facilitate a more precise detection of eventual new WMH occurring at

follow up, as well as of differences in their size and localization. This, particularly, would help in generating models of disease progression.

Moreover, our automated approach offers the possibility of a fully streamlined methodology with little user intervention and utilizes freely available software (e.g., FSL and SPM), which can be easily implemented on most desktop computers of standard power and are of routine use in most laboratories equipped to carry out neuroimaging research. This makes our approach reasonably accessible. Relatively little image manipulation is required and the quantification approach is procedurally straightforward. Additionally, little knowledge in computer programming allows to automate the steps and greatly reduce the processing time using in-house shell scripts which automatically concatenate all steps for all subjects.

A possible limitation of the presented methodology lies in the low number of brain slices used as inputs. In fact, the analyzed FLAIR images were acquired with a slice thickness of 5 mm which may be a poor resolution for accurate volumetric quantification and might induce interpolation errors in the registration processes. The MRI scanner used in the present study is relatively old and does not allow fast high-resolution 3D FLAIR imaging. A higher image resolution could improve the WMH quantification, and the registration accuracy, which would lead to more precise WMH localization; however, it would lead to an increase of acquisition time thus decreasing clinical practicability. However, modern MRI systems allow fast 3D FLAIR acquisition which would surely improve image resolution and, in turn, WMH segmentation.

In summary, we presented an automated detection and segmentation approach for WMH using T1-w and FLAIR images. By evaluating our method, we demonstrated that this segmentation approach is robust and reliable and we suggest that it has potential use in clinical studies.

Future works should consider the presented method to map cognitive consequences of WMH and their progression in time through large (and possibly multicenter) cross-sectional and longitudinal studies.

ACKNOWLEDGMENT

Grant support: Italian Ministry of Health grants RF07-08 and RC 06-07-08-09/A.

AUTHOR CONTRIBUTIONS

Mariangela Iorio designed the study, analyzed imaging data, contributed to the writing of the manuscript; Gianfranco Spalletta designed the study, analyzed data, and contributed to the writing of the manuscript; Chiara Chiapponi analyzed imaging data and contributed to the writing of the manuscript; Giacomo Lucichenti analyzed imaging data and contributed to the writing of the manuscript; Claudia Cacciari analyzed imaging data; Maria Donata Orfei wrote the draft of the manuscript; Carlo Calta-girone contributed to the writing of the manuscript; Fabrizio Piras designed the study, analyzed imaging data, contributed to the writing of the manuscript.

REFERENCES

- Admiraal-Behloul, F., Van Den Heuvel, D. M., Olofsen, H., Van Osch, M. J., Van Der Grond, J., Van Buchem, M. A., et al. (2005). Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. *Neuroimage* 28, 607–617. doi: 10.1016/j.neuroimage.2005.06.061
- Anbeek, P., Vincken, K. L., Van Osch, M. J., Bisschops, R. H., and Van Der Grond, J. (2004). Probabilistic segmentation of white matter lesions in MR imaging. *Neuroimage* 21, 1037–1044. doi: 10.1016/j.neuroimage.2003.10.012
- Ashburner, J., and Friston, K. J. (2000). Voxel-based morphometry: the methods. *Neuroimage* 11, 805–821. doi: 10.1006/nimg.2000.0582
- Barkhof, F., and Scheltens, P. (2002). Imaging of white matter lesions. *Cerebrovasc. Dis.* 13(Suppl. 2), 21–30. doi: 10.1159/000049146
- Bartko, J. J. (1991). Measurement and reliability: statistical thinking considerations. *Schizophr. Bull.* 17, 483–489. doi: 10.1093/schbul/17.3.483
- Brown, F. W., Lewine, R. R. J., and Hudgins, P. A. (1995). White matter hyperintensity signals associated with vascular risk factors in schizophrenia. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* 19, 39–45. doi: 10.1016/0278-5846(94)00102-N
- Cacciari, C., Moraschi, M., Di Paola, M., Cherubini, A., Orfei, M. D., Giove, F., et al. (2010). White matter microstructure and apathy level in amnesic mild cognitive impairment. *J. Alzheimers Dis.* 20, 501–507. doi: 10.3233/JAD-2010-1384
- Cerasa, A., Bilotta, E., Augimeri, A., Cherubini, A., Pantano, P., Zito, G., et al. (2012). Cellular neural network methodology for the automated segmentation of multiple sclerosis lesions. *J. Neurosci. Methods* 203, 193–199. doi: 10.1016/j.jneumeth.2011.08.047
- Chen, Y. F., Wang, H., Chu, Y., Huang, Y. C., and Su, M. Y. (2006). Regional quantification of white matter hyperintensity in normal aging, mild cognitive impairment, and Alzheimer's disease. *Dement. Geriatr. Cogn. Disord.* 22, 177–184. doi: 10.1159/000094785
- de Groot, J. C., de Leeuw, F. E., Oudkerk, M., Hofman, A., Jolles, J., and Breteler, M. M. (2000a). Cerebral white matter lesions and depressive symptoms in elderly adults. *Arch. Gen. Psychiatry* 57, 1071–1076. doi: 10.1001/archpsyc.57.11.1071
- de Groot, J. C., de Leeuw, F. E., Oudkerk, M., Van Gijn, J., Hofman, A., Jolles, J., et al. (2000b). Cerebral white matter lesions and cognitive function: the Rotterdam scan study. *Ann. Neurol.* 47, 145–151. doi: 10.1002/1531-8249(200002)47:2<145::AID-ANA3>3.3.CO;2-G
- de Leeuw, F. E., de Groot, J. C., Achten, E., Oudkerk, M., Ramos, L. M., Heijboer, R., et al. (2001). Prevalence of cerebral white matter lesions in elderly people: a population based magnetic resonance imaging study. The Rotterdam scan study. *J. Neurol. Neurosurg. Psychiatry* 70, 9–14. doi: 10.1136/jnnp.70.1.9
- Debette, S., and Markus, H. S. (2010). The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *Br. Med. J.* 341, c3666. doi: 10.1136/bmj.c3666
- DeCarli, C., Miller, B. L., Swan, G. E., Reed, T., Wolf, P. A., and Carmelli, D. (2001). Cerebrovascular and brain morphologic correlates of mild cognitive impairment in the National Heart, Lung, and Blood Institute Twin Study. *Arch. Neurol.* 58, 643–647. doi: 10.1001/archneur.58.4.643
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology* 26, 297–302. doi: 10.2307/1932409
- Dyrby, T. B., Rostrop, E., Baare, W. F., Van Straaten, E. C., Barkhof, F., Vrenken, H., et al. (2008). Segmentation of age-related white matter changes in a clinical multi-center study. *Neuroimage* 41, 335–345. doi: 10.1016/j.neuroimage.2008.02.024
- Fazekas, F., Chawluk, J. B., Alavi, A., Hurtig, H. I., and Zimmerman, R. A. (1987). MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *Am. J. Roentgenol.* 149, 351–356. doi: 10.2214/ajr.149.2.351
- Fazekas, F., Kapeller, P., Schmidt, R., Offenbacher, H., Payer, F., and Fazekas, G. (1996). The relation of cerebral magnetic resonance signal hyperintensities to Alzheimer's disease. *J. Neurol. Sci.* 142, 121–125. doi: 10.1016/0022-510X(96)00169-4
- Fazekas, F., Kleinert, R., Offenbacher, H., Schmidt, R., Kleinert, G., Payer, F., et al. (1993). Pathologic correlates of incidental MRI white matter signal hyperintensities. *Neurology* 43, 1683–1689. doi: 10.1212/WNL.43.9.1683
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatry Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6
- Gibson, E., Gao, F. Q., Black, S. E., and Lobaugh, N. J. (2010). Automatic segmentation of white matter hyperintensities in the elderly using FLAIR images at 3T. *J. Magn. Reson. Imaging* 31, 1311–1322. doi: 10.1002/jmri.22004
- Gunning-Dixon, F. M., and Raz, N. (2000). The cognitive correlates of white matter abnormalities in normal aging: a quantitative review. *Neuropsychology* 14, 224–232. doi: 10.1037/0894-4105.14.2.224
- Hirono, N., Kitagaki, H., Kazui, H., Hashimoto, M., and Mori, E. (2000). Impact of white matter changes on clinical manifestation of Alzheimer's disease: a quantitative study. *Stroke* 31, 2182–2188. doi: 10.1161/01.STR.31.9.2182
- Jack, C. R. Jr., O'Brien, P. C., Rettman, D. W., Shiung, M. M., Xu, Y., et al. (2001). FLAIR histogram segmentation for measurement of leukoaraiosis volume. *J. Magn. Reson. Imaging* 14, 668–676. doi: 10.1002/jmri.10011
- Mäntylä, R., Erkinjuntti, T., Salonen, O., Aronen, H. J., Peltonen, T., Pohjasvaara, T., et al. (1997). Variable agreement between visual rating scales for white matter hyperintensities on MRI comparison of 13 rating scales in a poststroke cohort. *Stroke* 28, 1614–1623. doi: 10.1161/01.STR.28.8.1614
- Marshall, G. A., Shchelchkov, E., Kaufer, D. I., Ivanco, L. S., and Bohnen, N. I. (2006). White matter hyperintensities and cortical acetylcholinesterase activity in parkinsonian dementia. *Acta Neurol. Scand.* 113, 87–91. doi: 10.1111/j.1600-0404.2005.00553.x
- Pantoni, L., and Garcia, J. H. (1997). Pathogenesis of leukoaraiosis: a review. *Stroke* 28, 652–659. doi: 10.1161/01.STR.28.3.652
- Payne, M. E., Fetzter, D. L., MacFall, J. R., Provenzale, J. M., Byrum, C. E., and Krishnan, K. R. (2002). Development of a semi-automated method for quantification of MRI gray and white matter lesions in geriatric subjects. *Psychiatry Res.* 115, 63–77. doi: 10.1016/S0925-4927(02)00009-4
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., and Kokmen, E. (1999). Mild cognitive impairment: clinical characterization and outcome. *Arch. Neurol.* 56, 303–308. doi: 10.1001/archneur.56.3.303
- Prins, N. D., Van Straaten, E. C., Van Dijk, E. J., Simoni, M., Van Schijndel, R. A., Vrooman, H. A., et al. (2004). Measuring progression of cerebral white matter lesions on MRI: visual rating and volumetrics. *Neurology* 62, 1533–1539. doi: 10.1212/01.WNL.0000123264.40498.B6
- Scheltens, P., Barkhof, F., Leys, D., Pruvo, J. P., Nauta, J. J., Vermersch, P., et al. (1993). A semiquantitative rating scale for the assessment of signal hyperintensities on magnetic resonance imaging. *J. Neurol. Sci.* 114, 7–12. doi: 10.1016/0022-510X(93)90041-V
- Schwarz, C., Fletcher, E., Decarli, C., and Carmichael, O. (2009). Fully-automated white matter hyperintensity detection with anatomical prior knowledge and without FLAIR. *Inf. Process. Med. Imaging* 239–251. doi: 10.1007/978-3-642-02498-6_20
- Shen, S., Szameitat, A. J., and Sterr, A. (2008). Detection of infarct lesions from single MRI modality using inconsistency between voxel intensity and spatial location: a 3-D automatic approach. *IEEE Trans. Inf. Technol. Biomed.* 12, 532–540. doi: 10.1109/TITB.2007.911310
- Simões, R., Mönninghoff, C., Dlugaj, M., Weimar, C., Wanke, I., van Cappellen van Walsum, A. M., et al. (2013). Automatic segmentation of cerebral white matter hyperintensities using only 3D FLAIR images. *Magn. Reson. Imaging* 31, 1182–1189. doi: 10.1016/j.mri.2012.12.004
- Smith, S. M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P. M., Federico, A., et al. (2002). Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage* 17, 479–489. doi: 10.1006/nimg.2002.1040
- Soderlund, H., Nilsson, L. G., Berger, K., Breteler, M. M., Dufouil, C., Fuhrer, R., et al. (2006). Cerebral changes on MRI and cognitive function: the CASCADE study. *Neurobiol. Aging* 27, 16–23. doi: 10.1016/j.neurobiolaging.2004.12.008

- Spalletta, G., Fagioli, S., Caltagirone, C., and Piras, F. (2013). Brain microstructure of subclinical apathy phenomenology in healthy individuals. *Hum Brain Mapp.* 34, 3193–3203. doi: 10.1002/hbm.22137
- Wahlund, L. O., Barkhof, F., Fazekas, F., Bronge, L., Augustin, M., Sjo-gren, M., et al. (2001). A new rating scale for age-related white matter changes applicable to MRI and CT. *Stroke* 32, 1318–1322. doi: 10.1161/01.STR.32.6.
- Wang, Y., Catindig, J. A., Hilal, S., Soon, H. W., Ting, E., Wong, T. Y., et al. (2012). Multi-stage segmentation of white matter hyperintensity, cortical and lacunar infarcts. *Neuroimage* 60, 2379–2388. doi: 10.1016/j.neuroimage.2012.02.034
- Wen, W., and Sachdev, P. (2004). The topography of white matter hyperintensities on brain MRI in healthy 60- to 64-year-old individuals. *Neuroimage* 22, 144–154. doi: 10.1016/j.neuroimage.2003.12.027
- Wu, C. C., Mungas, D., Petkov, C. I., Eberling, J. L., Zrelak, P. A., Buono-core, M. H., et al. (2002). Brain structure and cognition in a community sample of elderly Latinos. *Neurology* 59, 383–391. doi: 10.1212/WNL.59.3.383
- Wu, M., Rosano, C., Butters, M., Whyte, E., Nable, M., Crooks, R., et al. (2006). A fully automated method for quantifying and localizing white matter hyperintensities on MR images. *Psychiatry Res.* 148, 133–142. doi: 10.1016/j.psychres.2006.09.003
- Zijdenbos, A. P., Dawant, B. M., Margolin, R. A., and Palmer, A. C. (1994). Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE Trans. Med. Imaging* 13, 716–724. doi: 10.1109/42.363096

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 July 2013; accepted: 30 October 2013; published online: 02 December 2013.

Citation: Iorio M, Spalletta G, Chiapponi C, Luccichenti G, Cacciari C, Orfei MD, Caltagirone C and Piras F (2013) White matter hyperintensities segmentation: a new semi-automated method. *Front. Aging Neurosci.* 5:76. doi: 10.3389/fnagi.2013.00076 This article was submitted to the journal *Frontiers in Aging Neuroscience*.

Copyright © 2013 Iorio, Spalletta, Chiapponi, Luccichenti, Cacciari, Orfei, Caltagirone and Piras. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.