Association of Multiple Sclerosis Related Cognitive Impairment with an MRI-Derived Composite Score

Flavia Nelson1*, Aziz H. Poonawalla2,3, Sushmita Datta1, Rosa C. Banuelos1, Mohammad H. Rahbar4,5, Francisco Perez2, Jerry S. Wolinsky1 and Ponnada A. Narayana1

1Department of Neurology, University of Texas Health Science Center-Houston, Houston, TX, USA
2Department of Radiology, University of Wisconsin-Madison, Madison, WI, USA
3Department of Diagnostic and Interventional Imaging, University of Texas Health Science Center-Houston, Houston, TX, USA
4Biostatistics, Epidemiology and Research Design Component, Center for Clinical and Translational Sciences, University of Texas Health Science Center-Houston, Houston, TX, USA
5Division of Clinical and Translational Sciences, Department of Internal Medicine, University of Texas Health Science Center-Houston, Houston, TX, USA

*Corresponding author: Flavia Nelson, MD, 6431 Fannin Street, MSB 7.220, Houston, TX USA, Tel: (713) 500-7058; Fax: (713)500-7040; E-mail: Flavia.M.Nelson@uth.tmc.edu

Received date: Aug 28, 2014, Accepted date: Oct 28, 2014, Publication date: Oct 30, 2014

Abstract

Background: MRI-derived metrics such as atrophy, burden of disease (BOD), and cortical lesions (CL) have independently been reported to be associated with Multiple Sclerosis (MS) related cognitive impairment (CI). Composite scores combining some of these individual metrics have also been shown to improve correlations with MS-related physical disability; however the value of a composite score for MS-related CI has not yet been evaluated. In this study we assessed the relationship between CI and a quantitative composite score constructed from MRI-derived metrics to include total white matter lesion volume, CL number, and normalized cerebrospinal fluid (nCSF), a measure of brain atrophy.

Methods: Thirty three (n=33) patients underwent neuropsychological testing and were classified into CI groups using a 4-point severity scale (0, 1, 2, 3) where zero indicates non-impaired, “1” represents borderline, “2” represents mild, and “3” represents moderate CI. Images obtained at 3T were segmented into tissue and lesion compartments from which BOD and nCSF were quantitatively measured. CL number was identified visually by consensus. BOD, nCSF, and CL number were then transformed to z-scores: zBOD, znCSF and zCL, and a composite score “Z3” was constructed from these measures. The associations between Z3 and CI and the individual transformed measures (z-scores) and CI were evaluated via ordinal logistic regression.

Results: Z3 was significantly associated with CI (OR=1.443, 95% Confidence Interval: 1.048-1.987, p=0.024) with a slightly larger OR than any individual measure. Of the individual measures, only BOD had a significant association with CI (OR=1.064, 95% Confidence Interval: 1.008-1.123, p=0.025). No significant association was found between atrophy or CL and CI.

Conclusion: The Z3 score is associated with increased CI severity. This association was mainly driven by BOD. Larger studies are needed to assess the potential advantage of a composite score over measures of white matter lesion burden alone.

Keywords: Multiple sclerosis; MRI; Cognitive impairment; Cortical lesions; Atrophy; Burden of disease; Composite score

Abbreviations:

MRI: Magnetic Resonance Imaging; MS: Multiple Sclerosis; CI: Cognitive Impairment; CL: Cortical Lesions; BOD: Burden of Disease; CNS: Central Nervous System; Gd: Gadolinium; nCSF: normalized CSF; EDSS: Expanded Disability Status Scale; NP: Neuropsychological; DIR: Double inversion recovery; PSIR: Phase sensitive inversion recovery; OR: Odds ratio

Introduction

MRI is the most sensitive noninvasive imaging modality for visualization and characterization of MS lesions. There is a general enthusiasm for using MRI metrics as biomarkers in clinical trials and treatment evaluation [1,2]. However, MRI measures have not been accepted as biomarkers by regulatory authorities, in part due to the lack of understanding of their relationship with clinical outcomes [3]. Currently there is no consensus on whether any single MRI derived measure (such as white matter lesion load, enhancing lesion volume, black hole volume or global/regional atrophy, to name a few) is sufficient to serve broadly as an imaging biomarker in MS [4].
Cognitive Impairment (CI) is an important cause of MS-related disability, contributing to the great interest of the MS community for better diagnostic and monitoring tools for CI. Individual MRI markers of tissue damage such as global atrophy and T2 lesion load have been shown to have significant correlations with CI [5-10]. Other MRI-based measures such as cortical, thalamic and callosal atrophy and their correlations with CI, have also been investigated in the literature, with varying results [11-15]. Other novel imaging techniques, such as double inversion recovery (DIR) which suppresses the white matter and the cerebrospinal fluid (CSF) signal providing a clear visualization of the cortical ribbon and phase sensitive inversion recovery (PSIR) which provides a sharp, delineation between gray and white matter, have been applied to the study of CI. Recent studies on the application of these techniques report an increase in the detection sensitivity of cortical lesions (CL) and have established a role of CI in the pathophysiology of CI [16-20]. Despite irrefutable progress in this field, none of the aforementioned studies have provided enough evidence to definitively conclude that a single MRI measure can potentially be used as an MRI risk indicator of CI.

Since one of the major strengths of MRI is its multi-modal nature, a promising approach is to construct a CI biomarker from a combination of MRI-derived measures, each capturing a different aspect of the disease’s clinical course [21-25]. However, different MRI measures have different dimensions, such as lesion volume (measured in cc), tissue atrophy (measured as percent/ratio/fraction), or CL counts, reported numerically. In addition, the numerical values of each measure may differ widely, with tissue volumes several orders of magnitude larger than lesion volumes. These limitations may be overcome by using a z-score transformation where each measure is standardized creating a dimensionless measure. After the z-score transformation, different measures may simply be added together to form a composite measure, providing a more complete representation of the disease process than with any single MRI derived measure alone. The Z score assumes that the directions of the deviations are consistent in indicating a worse condition among all the measures in the composite score, and that the contributions of all z-scores are equally weighted.

MRI derived composite scores in MS

The use of composite scores has precedent in the MS literature. Quantitative measures derived from MRI such as the z-score have previously been investigated and correlated with measures of clinical disability like the Expanded Disability Status Scale (EDSS) [26]. A composite measure of this type (Z4) was introduced in 2000 as part of the results of the North American Linomide Trial [27] and was shown to distinguish treatment groups at 3 months. The composite score Z4 included volume of Gd-enhancing tissue (reflecting active inflammation), total lesion volume (reflecting heterogeneous changes in tissue), T1 hypointense lesion volume (reflecting severe tissue alterations), and normalized cerebrospinal fluid (a measure of global atrophy). This composite score was described as follows: Z4=z (Gd lesions)+z(T1+T2 lesions)+z(T1 lesions)+z(nCSF). All of these metrics were measured as volume except for normalized cerebrospinal fluid (nCSF) which is measured as percent/ratio/fraction. Analysis showed that patients who worsened clinically also showed an increase in Z4, supporting the notion that a composite score of multiple MRI measures has clinical relevance [27]. More recently, the same Z4 score was applied to cohorts from the Teriflunomide Multiple Sclerosis Oral trial (TEMSO), who were teriflunomide doses of both 7 and 14 mg reduced the mean change from baseline in Z4 over time compared to placebo [28,29].

In the present work, to investigate the relationship of combined MRI-derived metrics with CI, we used a similar composite score the Z3 score which incorporates 1) the CL number, 2) T1+T2 lesion volume representing burden of disease (BOD), and 3) nCSF, representing global atrophy, all scores directionally compatible; larger being worse. We investigated the association between Z3 and CI severity using a 4-point impairment scale ranging from no impairment to moderate impairment, based on a composite battery of neuropsychological (NP) testing. We hypothesized that the Z3 score will be a better indicator of CI than any of the individual measures.

Materials and Methods

Subjects

This was a single center, prospective study, approved by our institutional review board (HSC-MS-05-0333-Cortical lesion detection and correlation with cognitive impairment in multiple sclerosis). Volunteers included 39 MS patients enrolled in the study at The University of Texas Health Science Center at Houston; 6 subjects were excluded from the analysis due to inadequate image acquisition and one due to severe CI scores consistent with dementia and suggestive of a concomitant pathology. The analyzed cohort was comprised of 30 relapsing-remitting, 2 secondary-progressive, and 1 primary-progressive MS patients with mean EDSS of 3.2 (range 1.0-6.0), who were drawn from an existing database of patients who were tested for evidence of CI within one year prior to the MRI scan. Each patient was evaluated by the same neuropsychologist (FP) who has extensive expertise in detection of MS related CI, the same neuropsychological battery was used in all subjects. Written informed consent was obtained from every participant prior to enrollment and subject eligibility was established based on NP test results as well as inclusion/exclusion criteria as follows:

Inclusion criteria: 1) clinically definite MS by the McDonald criteria [30], 2) NP evaluation performed within one year from enrollment, 3) age 18 to 60 (to avoid confounding effects of age-related cognitive deficits), 4) EDSS score 6 or less, 5) brain MRI consistent with the clinical diagnosis of MS.

Exclusion criteria: 1) contraindications for MRI (pacemakers, metal implants, claustrophobia etc.), 2) history of clinical relapse or a high dose steroid treatment in the 3 months prior to the study, 3) history of drug or alcohol abuse, 4) acute or uncontrolled depression within 3 months from testing, and 5) history of treatment with immunosuppressive treatments.

Once eligibility was established by the neuropsychologist, recruitment was done via a telephone call or an invitation letter. Eligible patients were referred for one imaging session and patient eligibility was confirmed prior to enrollment. All personnel except for the neuropsychologist were blinded to the subjects’ cognitive status and the neuropsychologist was blinded to the MRI findings. Each patient enrolled attended a 2 hour session that included: 1) collection of demographics and disease history, 2) neurological exam/EDSS score assessment, and 3) MRI of the brain without gadolinium.
Neuropsychological assessment

A comprehensive battery of standardized NP tests that evaluated a broad range of cognitive domains was administered by a board certified neuropsychologist with extensive expertise in MS-related CI. A complete list of tests grouped by cognitive function is summarized in Table 1.

<table>
<thead>
<tr>
<th>Cognitive function</th>
<th>Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td>Wechsler Memory Scale-R (WMS-R)</td>
</tr>
<tr>
<td></td>
<td>Auditory-verbal learning test</td>
</tr>
<tr>
<td>Ability and achievement</td>
<td>Wechsler Adult Intelligence Scale—III (WAIS—III)</td>
</tr>
<tr>
<td>Verbal functions and language skills</td>
<td>Aphasia screening test</td>
</tr>
<tr>
<td></td>
<td>Drawings on command</td>
</tr>
<tr>
<td></td>
<td>Controlled oral word association (COWA)</td>
</tr>
<tr>
<td>Motor function</td>
<td>Sensory-Perceptual examination</td>
</tr>
<tr>
<td></td>
<td>Finger tapping test (FTT)</td>
</tr>
<tr>
<td></td>
<td>Grooved Pegboard test (GPT)</td>
</tr>
<tr>
<td></td>
<td>Hand dynamometer or gripped strength test</td>
</tr>
<tr>
<td>Attention concentration and tracking (working memory)</td>
<td>Paced Auditory Serial Addition Test (PASAT)</td>
</tr>
<tr>
<td></td>
<td>Stroop Neuropsychological Screening Test</td>
</tr>
<tr>
<td></td>
<td>Trial Making Test</td>
</tr>
<tr>
<td>Executive function</td>
<td>Wisconsin Card Sorting Test (WCST)</td>
</tr>
<tr>
<td>Personal adjustment and emotional functioning</td>
<td>Minnesota Multiphasic Personality Inventory-2 (MMPI)</td>
</tr>
</tbody>
</table>

Table 1: Neuropsychological testing.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Plane</th>
<th>TR msec</th>
<th>TE msec</th>
<th>TI msec</th>
<th>Image matrix</th>
<th>FOV mm</th>
<th>Slice mm</th>
<th>SENSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSIR</td>
<td>axial</td>
<td>4300</td>
<td>13</td>
<td>400</td>
<td>256x256</td>
<td>240</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>DIR</td>
<td>axial</td>
<td>15000</td>
<td>25</td>
<td>3400/325</td>
<td>512x512</td>
<td>240</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>T1W</td>
<td>axial</td>
<td>600</td>
<td>9.2</td>
<td>-</td>
<td>256x256</td>
<td>240</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Dual-echo FSE</td>
<td>axial</td>
<td>6800</td>
<td>10/90</td>
<td>-</td>
<td>256x256</td>
<td>240</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>FLAIR</td>
<td>axial</td>
<td>10,000</td>
<td>80</td>
<td>2600</td>
<td>256x256</td>
<td>240</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2: Image acquisition parameters for MRI pulse sequences in the MS imaging protocol.

The clinical interpretation of the NP test results for each individual patient was based on an integrated approach which takes into account standard test scores consistent with brain dysfunction, test score patterns, and the qualitative aspects of the examination. Based on an extensive assessment of executive and memory functions, CI was defined as patterns of impairment in two or more tests in comparison with established normative data. Individual test results were classified as non-impaired if the standard score did not deviate from the normative data. Borderline impairment was classified as below 1 standard deviation (SD) from the normative data. Mild impairment was classified as 1.5 SD below the normative data. Moderate impairment was classified as 2 SD below the mean. The specific classification of each patient was made by a board-certified neuropsychologist with extensive experience in the evaluation of cognitive functions in patients with MS. Since CI in MS is generally not homogeneous and there is great variability from patient to patient, the neuropsychologist used his clinical judgment in classifying each patient following the criteria established. Individual results were classified on a 4-point scale (0, 1, 2, 3) where zero indicates non-impaired (normal), “1” represents borderline, “2” represents mild, and “3” represents moderate CI.

MR imaging

Each patient underwent an MRI of the brain according to a protocol that included standard T1 and T2 weighted scans and two optimized sequences for detection of CI; DIR, and PSIR, for acquisition parameters please see Table 2.

Quantitative MRI Analysis

Image analysis, including segmentation for tissue quantization was performed using integrated operator assisted automated image analysis (MRI-AP) [31] and database management (CTS) software packages developed at our institution. In brief, all images acquired in different series (FLAIR, dual echo, T1-images etc.) were retrospectively aligned. Images were stripped of the extra-meningeal tissues, filtered using the anisotropic diffusion filter, corrected for the bias field, and intensity normalized [31]. Segmentation of GM, WM, and T2-hyperintense lesion volume were based on the dual echo and FLAIR images using the hidden Markov random field expectation-maximization (HMRF-EM) algorithm and Parzen window classification [31].

The measure of whole brain atrophy used was nCSF [27]. T1-hypointense lesions were automatically identified and quantified using...
the procedure described by Datta et al., 2006 [32]. BOD was calculated as the sum of T2-hyperintense and T1-hypointense component lesion volumes, obtained from tissue segmentation (Figure 1).

Figure 1: Example of automated segmentation defining T2 (salmon) and T1 (red) lesion volumes, as well as gray and white matter and CSF.

Cortical lesions

CL were identified on DIR and validated on PSIR as previously detailed [33]. Lesions were classified using DIR/PSIR as (a) purely intracortical (IC, total confinement within the cortical ribbon) and (b) mixed (MX, originating in the cortex but with at least 25% subcortical extension); for the purpose of this analysis both types (IC and MX) were considered cortical. Lesion identification and classification was done by 2 experienced raters (FN and AP) by consensus and without knowledge of the NP test results.

Statistical analysis/Composite Z-score calculation

Individual Z-scores for each of the measures CL, BOD (T2+T1), and nCSF were computed by subtracting the sample mean of the measure from the original score and dividing it by its sample standard deviation. The composite z-score Z3 was then calculated as the sum of the three individual scores: Z3=zBOD+znCSF+zCL.

Factor analysis of the three Z-scores was performed to justify equal weights used in the composite score. The factor loadings from factor analysis were 0.66, 0.63, and 0.50 for the standardized variables zCL, znCSF, and zBOD, respectively. The similarity in size of the factor loadings indicates that equal weights are reasonable for these three variables in the Z3 composite score. Based on our factor analysis, the composite score accounts for about 91% of the variation in these variables. The outcome variable CI is a measure of the severity or degree of disease and consists of the categories non-impaired, borderline, mild, and moderate as described earlier. In our analyses we treated CI as an ordinal variable by using ordinal logistic regression. Initially, univariable ordinal logistic regression was carried out to determine which variables were associated with CI. Specifically, the demographic variables gender and age, the MRI-based metrics CL, nCSF, and BOD, and the composite score Z3, were individually included in the logistic models as the independent variable. Multivariable ordinal logistic regression models were also fit to control for potential confounding due to demographic variables age and gender. For both univariable and multivariable ordinal logistic regression models, proportional odds were assumed. Statistical analysis was performed using the SAS 9.3 statistical software (SAS Inc., Cary, NC).

Results

Of the 33 patients, 24.2 % (n=8) were found to be non-impaired (normal), 24.2% (n=8) were found to be borderline, and 51.5 % were found to be in the mild (n=10) and moderate (n=7) range of impairment (30.3% and 21.2%, respectively). Univariable ordinal logistic regression models showed that the individual variables associated with CI were age (p=0.030), BOD (p=0.037) and the composite score Z3 (p=0.052) the latter marginally significant. For the
association between the composite score Z3 and CI, a unit increase in Z3 translates to an odds increase of 1.352 of having greater CI. The associations of CI to gender, CL, and nCSF were not significant (all \(p>0.100\)). Details regarding measures of associations are summarized in Table 3.

### Table 3: Factors associated with the Outcome Variable CI based on Univariable Ordinal Logistic Regression Models.

The validity of the proportional odds assumption was tested and confirmed for all univariable and multivariable models at 5% level of significance.

After adjusting for age, the composite score Z3 was significantly associated with CI (OR=1.443, 95% Confidence Interval: 1.048-1.987, \(p=0.024\)). This association was slightly stronger than that of any of the individual measures, of which only BOD was significant (OR=1.064, 95% Confidence Interval: 1.008, 1.028, \(p=0.025\)), additional details are provided in Table 4.

### Table 4: Association between Z scores and the Outcome Variable CI after adjusting for age as a potential confounding variable in separate Multivariable Ordinal Logistic Regression Models.

We further explored the association of CI and Z3 by comparing non-impaired (normal) to each of the CI impairment categories. Relative to the referent category, CI=0 (non-impaired) the age-adjusted OR of increased CI severity corresponding to one unit increase in the composite score Z3 were 1.443, 2.083, and 3.007 for CI=1, CI=2, and CI=3, respectively. Additional details are presented in Table 5.

### Table 5: Association between Z3 with outcome variable CI via Ordinal Logistic Regression (comparing Baseline CI=0 to each CI Severity Level).

**Discussion**

The literature on the relationship between CI and individual MRI metrics has expanded rapidly in recent years. In addition to improving our understanding of the pathophysiology of the disease, these studies suggest that CI is unlikely a consequence of damage to one specific area of the CNS (such as cortex, thalamus, white matter, etc.) and that CI can be associated with different individual MRI measures. Based on this, the assumption that combining multiple MRI derived metrics would potentially strengthen the association with CI was entertained. This study demonstrates that when the most common measures of brain tissue damage such as atrophy and BOD (T1 + T2 lesion load), and a frequent measure associated with CI, CL number, are combined by means of an MRI composite score (Z3), the relationship between CI and brain tissue damage could potentially be better represented than with single individual metrics. These results were unexpected and their interpretation needs to be tempered by the limitations of the study which are mainly the small cohort size and the significant variability in the number of CL found in each individual patient which ranged from 0 to 63. The observed variability of CL between MS subjects is consistent with previous post mortem histopathological studies by Brownell and Hughes [34]. The location of CL was not incorporated in the analyses also due to the significant variability in number and location of CL among individuals. Another consideration is our limited ability to detect CI with current imaging techniques; it is estimated that detection rate with a 1.5T magnet and the DIR technique is 30% of what is found in post-mortem studies [35]. Although our study used a 3T magnet which increases detection over 1.5T, when it comes to cortical damage we still are likely under-detecting the true cortical lesion load [36]. Therefore the lack of a significant association between CI and CL observed in this analysis could reflect technological limitations. In a previous study by our group, a significant correlation was found between CI and CI, this significance was not reproduced in the current analysis. This could be explained by the facts that the previous analysis included a larger number of subjects but more likely due to the use of a different statistical approach. Briefly, in the previous study Poisson regression models were first used to evaluate for evidence of an association between CI and each different type of lesion. Negative binomial regression was also done and preferred as it explained the variation in the number of lesions better than the Poisson model. A binary response variable was used in the original analysis whereas in the
current one we used an ordinal variable to take into account the severity of CI. In the previous study, the CI categories (normal, borderline, mild, moderate) were collapsed into two groups: normal and impaired. This approach could also explain the discrepancy in significance with our current findings.

In conclusion, although the associations between Z3 and CI and BOD and CI did not vary significantly, likely explained by the notion that the significance of Z3 was mostly driven by BOD, the benefit of composite scores is that they describe the disease state in a more comprehensive manner than a single metric alone and is better suited for multicenter clinical trials by virtue of being a normalized metric. A composite score has the potential for a more robust representation of the relation with CI even if minimal, and possibly more so if applied to larger data sets or patients with a more significant cortical lesion load than the ones evaluated in this cohort. Last but not least, an ideal study of this kind would include other metrics not explored in this analysis such as thalamic and cortical atrophy, both of which have been associated with CI. Of interest, atrophy metrics alone which are considered to have some of the strongest relationships with CI were not found to be significant in our study.

Efforts to advance our understanding of the pathophysiological basis of cognitive dysfunction in MS require continual refinement. A multimodal approach including advanced imaging techniques such as DTI and fMRI may shed more light on the importance of the integrity of neural networks to the preservation of cognitive function in MS. Such studies are currently underway at our institution.

Acknowledgements

We would like to thank Vipul Kumar “Vips” Patel for his instrumental assistance during image acquisition. We also acknowledge the support provided by the Biostatistics/ Epidemiology/ Research Design (BERD) component of the Center for Clinical and Translational Sciences (CCTS) for this project. CCTS is mainly funded by the NIH Centers for Translational Science Award (NIH CTSA) grant (UL1 RR024148), awarded to University of Texas Health Science Center at Houston in 2006 by the National Center for Research Resources (NCRR) and its renewal (UL1 TR000371) by the National Center for Advancing Translational Sciences (NCATS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NCRR or the NCATS.

Funding

This work was primarily supported by NIH (grant number RO1EB002095-04S1, to FN RO1EB002095 to PAN) and by the Clayton Foundation for Research (to JSW). The purchase of the 3T scanner was partially supported by NIH (grant number S10 RR19186 to PAN).

Conflicts of interest

The authors declare they have no conflicts of interest.

References


