Quantification of White Matter Alterations: A Reliability Analysis

Catherine Price¹,², Ilona Schmalfuss²,³,⁴, Christopher Sistrom³

University of Florida Departments of Clinical & Health Psychology¹, Anesthesiology², Radiology³, North Florida-South Georgia Veteran Administration Radiology Department⁴

Rationale

White Matter Alterations (WMAs) are also termed ‘subcortical hyperintensities’, ‘white matter lesions’, and ‘unidentified bright objects’. These names typify their bright and diffuse appearance in the white matter on T2 weighted MRI images. The clinical significance of WMAs has become of increasing interest as WMAs increase in frequency with advancing age and are associated with a variety of disorders.

Problem: Published studies vary in the 1) type of MRI sequence used to visualize WMAs and 2) type of measurement scale. Most frequently used are T2 weighted images and visual rating scales.

Purpose: To examine differences in the reliability of raters’ WMA measurements according to

1) type of MRI sequence (Fluid Attenuated Inversion Recovery (FLAIR) versus conventional T2-weighted) and

2) measurement technique (semi-automated pixel thresholding versus traditional visual rating scales).

Hypothesis: Stronger (higher) rater reliabilities will be obtained when WMAs are measured on FLAIR images and with a semi-automated computerized thresholding technique.

Participant Sample

Following IRB approval, 140 out of 3,382 MRI scans obtained in a 12 month time frame were selected for rating using a set inclusion and exclusion criteria.

Inclusion criteria: age 60 or older, presence of WMA, T2 and FLAIR images available, adequate quality

Exclusion criteria: prior brain surgery, infarctions, tumor

In order to obtain a sample of brains with a range of WMA severity, all 140 brains were graded as mild, moderate, severe by a neuroradiologist (IS).

After removing header information, 15 of the 140 scans (5 mild, 5 moderate, 5 severe) were randomly selected, separated by sequence (FLAIR, T2) and randomized for repeat scoring.

Rater Procedures

Two raters blind to severity rank independently conducted 90 ratings of randomly organized unidentified scans.

- 3 times for each FLAIR scan
- 3 times for the matching T2 scan

For these ratings, two technique were used:
- Pixel Thresholding Technique.
- Rotterdam Visual Rating Scale.

Type of MRI Sequences Rated

FLAIR

T2 weighted

Rating Measurement Techniques

Pixel Thresholding

Visual Rating Scale

Scion Image is a derivative of NIH Image. Both are image processing freewares (rsbweb.nih.gov).

Results

<table>
<thead>
<tr>
<th>Intra-class correlation (ICC) coefficients</th>
<th>mean</th>
<th>error variance</th>
<th>patient variance</th>
<th>reliability (ICC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLAIR_pixel</td>
<td>9487</td>
<td>0.455</td>
<td>48.5</td>
<td>0.99</td>
</tr>
<tr>
<td>FLAR_visual</td>
<td>5366</td>
<td>2.49</td>
<td>33.7</td>
<td>0.93</td>
</tr>
<tr>
<td>T2_pixel</td>
<td>8989</td>
<td>1.76</td>
<td>41.0</td>
<td>0.96</td>
</tr>
<tr>
<td>T2_visual</td>
<td>7222</td>
<td>17.3</td>
<td>31.4</td>
<td>0.64</td>
</tr>
</tbody>
</table>

*Depicts individual rater scores for 3 time periods. The order in which the patients are shown is the SAME for each plot.

Conclusion

• The Pixel Threshold method applied to FLAIR images is the most reliable approach for the analysis of WMAs.

• Moreover, the Pixel Threshold method provided high reliability within and between the T2 and FLAIR sequences. Consequently, both FLAIR and T2 images are equally appropriate to use when a reliable quantification method is applied.

• Applying a Visual Rating technique to T2 images yielded the lowest reliability.