Measuring white matter lesions on T2 FLAIR MRI: a reliability study

Stephen D. Towler, Catherine C. Price, Sandra M. Mitchell, David J. Libon

Department of Clinical and Health Psychology, University of Florida; Department of Neurology, Drexel University College of Medicine

Aims

**AIM 1:** Evaluate the inter- and intrarater reliability of manually segmenting white matter hyperintensities (WMH) on T2 FLAIR brain MRIs of dementia patients. Each rater’s slice-by-slice segmentation produces a whole-brain binary mask of the WMH areas, from which whole-brain WMH tissue volume is calculated.

**AIM 2:** Evaluate whether these manual measurements of WMH tissue volume are correlated with an expert rater’s subjective visual ratings of the same lesions.

Methods

1) Participants and image acquisition
Fifteen previously acquired whole-brain T2 FLAIR volumes selected from clinical sample of 60 MRIs of individuals with mild to moderate dementia. Mean age 80.1 +/- 5.4 y, education 12.1 +/- 2.9 y, T2 FLAIR from 1.5T Siemens MRI: 26 slices, 5 mm thick, 1 mm gap.

2) Previously acquired subjective visual rating scores
Raters trained to reliability previously rated extent of WMH in each T2 FLAIR using the semi-quantitative Junque LA Scale (Junque, 1990).

3) New WMH measurements using computerized binary segmentation

To create a WMH binary mask from each whole-brain T2 FLAIR MRI, white matter hyperintensities were manually segmented from normal-appearing tissue slice-by-slice. Locally-developed macros and US NIH software (ImageJ: http://rsbweb.nih.gov/ij) reduced manual segmentation to four steps per slice (less than 15 minutes per 26-slice MRI):

1) Adjust brightness and contrast to achieve the subjectively best contrast between WMH and bordering tissue.
2) Use a computer mouse to draw gross boundaries around the hyperintense white matter lesions.
3) Highlight only the hyperintense pixels by adjusting upper- and lower-threshold sliders.
4) Finalize rating for that slice, creating a WMH binary mask that has non-zero values in those areas that are both outlined and highlighted.

Every T2 FLAIR MRI was segmented twice by each rater in a blind pseudo-random order with non-consecutive repeats. WMH tissue volumes were calculated from the resulting WMH binary masks.

Results

**Table 1. High reliability was established within and between raters: Manual segmentations of T2 FLAIR MRI white matter hyperintensities.**

<table>
<thead>
<tr>
<th>Rater 1</th>
<th>Rater 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>rating a</td>
<td>rating b</td>
</tr>
<tr>
<td>DSC = .933</td>
<td>DSC = 1</td>
</tr>
<tr>
<td>DSC = .831</td>
<td>DSC = .827</td>
</tr>
<tr>
<td>DSC = .798</td>
<td>DSC = .792</td>
</tr>
<tr>
<td>DSC = .842</td>
<td></td>
</tr>
</tbody>
</table>

- spatial overlap was high in all repeated ratings: Dice Similarity Coefficient grand mean = 0.84, sd = 0.12. Interpret DSC like kappa (Zijdenbos, 1994).
  - all ICCs > 0.94 (p’ < 0.001)

**Table 2. These manually-measured white matter hyperintensity tissue volumes were also reliable with expert raters’ subjective visual rating scores.**

<table>
<thead>
<tr>
<th>Rater 1</th>
<th>Rater 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>rating a</td>
<td>rating b</td>
</tr>
<tr>
<td>Spearman’s rho = .845</td>
<td>Spearman’s rho = .858</td>
</tr>
<tr>
<td>Spearman’s rho = .849</td>
<td>Spearman’s rho = .831</td>
</tr>
</tbody>
</table>

The white matter lesion volumes measured during the present work correlated with previously collected expert subjective visual ratings of MRI white matter lesions (Junque scores: Spearman’s rho > 0.8, p < 0.01).

Discussion

This work demonstrates a reliable, efficient, high-fidelity method for measuring WMH. This is particularly important for MR studies of dementia patients, as previous work has demonstrated an important double-dissociation in the relationship between WMH severity and focus of cognitive deficits (Price, 2005).

Future Directions

1) Combine 3D lesion maps with maps of brain regions to investigate how the location of white matter lesions relates to cognitive and clinical decline.
2) Combine 3D lesion maps with diffusion imaging (3D MR DTI) to investigate the relationship with WMH and the role of white matter integrity in age-related cognitive and clinical changes.
3) Evaluate, adapt, and train automated WMH measurement techniques. Inter-method reliability will be evaluated with the same approach as interrater reliability in the present work.

Acknowledgements and Refs.


