

**On-line Table: Inter-reader agreement**

| Variable                    | Reader <sup>a</sup> |           |                      | Agreement  |              |                    |
|-----------------------------|---------------------|-----------|----------------------|------------|--------------|--------------------|
|                             | Reader 1            | Reader 2  | P Value <sup>b</sup> | $\kappa^c$ | (95% CI)     | Agree <sup>d</sup> |
| Based on group B images     |                     |           |                      |            |              |                    |
| Overall image T2WIs         |                     |           |                      |            |              |                    |
| Overall image-quality grade | 3.7 ± 0.8           | 4.2 ± 0.5 | .011                 | 0.09       | (-0.10-0.28) | 68%                |
| Fat-saturation grade        |                     |           |                      |            |              |                    |
| Maxillary region            | 4.8 ± 0.4           | 4.5 ± 0.6 | .008                 | 0.07       | (-0.17-0.30) | 94%                |
| Mandibular region           | 4.6 ± 0.6           | 4.3 ± 0.6 | .041                 | 0.05       | (-0.19-0.30) | 87%                |
| Lower neck region           | 4.7 ± 0.5           | 4.6 ± 0.6 | .81                  | 0.14       | (-0.16-0.44) | 94%                |
| Post-Gad T1WIs              |                     |           |                      |            |              |                    |
| Overall image-quality grade | 3.7 ± 0.7           | 4.3 ± 0.6 | .003                 | 0.06       | (-0.13-0.24) | 65%                |
| Fat-saturation grade        |                     |           |                      |            |              |                    |
| Maxillary region            | 5.0 ± 0.0           | 4.6 ± 0.6 | .003                 | -          | -            | 97%                |
| Mandibular region           | 4.9 ± 0.2           | 4.5 ± 0.5 | .001                 | 0.15       | (-0.05-0.35) | 100%               |
| Lower neck region           | 5.0 ± 0.2           | 4.6 ± 0.6 | .005                 | 0.12       | (-0.10-0.33) | 97%                |
| Based on group A images     |                     |           |                      |            |              |                    |
| STIR images                 |                     |           |                      |            |              |                    |
| Overall image-quality grade | 3.4 ± 0.7           | 3.8 ± 0.8 | .007                 | 0.23       | (-0.06-0.53) | 52%                |
| Fat-saturation grade        |                     |           |                      |            |              |                    |
| Maxillary region            | 4.2 ± 0.6           | 4.4 ± 0.7 | .14                  | 0.25       | (-0.01-0.51) | 91%                |
| Mandibular region           | 4.1 ± 0.7           | 4.0 ± 0.8 | .51                  | 0.30       | (0.05-0.56)  | 82%                |
| Lower neck region           | 4.3 ± 0.6           | 4.3 ± 0.6 | .86                  | -0.07      | (-0.33-0.19) | 84%                |
| Post-Gad T1WIs              |                     |           |                      |            |              |                    |
| Overall image-quality grade | 2.8 ± 0.8           | 2.3 ± 0.5 | .001                 | 0.18       | (-0.03-0.39) | 39%                |
| Fat-saturation grade        |                     |           |                      |            |              |                    |
| Maxillary region            | 4.2 ± 0.8           | 3.4 ± 0.8 | <.001                | 0.25       | (0.06-0.43)  | 52%                |
| Mandibular region           | 3.2 ± 0.7           | 2.4 ± 0.5 | <.001                | 0.10       | (0.00-0.20)  | 25%                |
| Lower neck region           | 1.5 ± 0.8           | 1.4 ± 0.6 | .53                  | 0.58       | (0.39-0.77)  | 87%                |

<sup>a</sup> Values are mean ± SD or No. (%). All quality grades are on a 1-5 scale, with 5 being the best quality.

<sup>b</sup> Wilcoxon signed rank test for an average difference between the readers.

<sup>c</sup> Cohen weighted  $\kappa$  (linear weights) for quality grades and Cohen unweighted  $\kappa$  for susceptibility artifacts.

<sup>d</sup> Percentage of subjects on which both readers agree, after grouping the 5-point scale as 1-2, 3, and 4-5.