Visual Rating Scales for Age-Related White Matter Changes (Leukoaraiosis)
Can the Heterogeneity Be Reduced?

Leonardo Pantoni, MD, PhD; Michela Simoni, MD; Giovanni Pracucci, MD; Reinhold Schmidt, MD; Frederik Barkhof, MD, PhD; Domenico Inzitari, MD; for the European Task Force on Age-Related White Matter Changes

Background and Purpose—It has been hypothesized that the use of different visual rating scales partly explains the discordant results of studies investigating risk factors and clinical correlates of age-related cerebral white matter changes (leukoaraiosis). We aimed to compare 6 widely used rating scales for leukoaraiosis and to calculate conversion coefficients of the score of 1 scale in the score of a second scale.

Methods—Two trained raters evaluated 80 pairs of CT and MRI scans using 2 CT and 4 MRI rating scales for white matter changes. Correlations among the scales were evaluated and regression lines were constructed with each of the CT and MRI scale scores as variables.

Results—A high correlation was observed in all the paired comparisons of the 6 scales (Spearman’s \( r \) ranging from 0.85 to 0.96, \( P < 0.0001 \)). Using regression analysis, we determined numeric parameters to transform the score of 1 scale to the corresponding score for each of the remaining scales and relative confidence intervals. The predictive values of these conversions expressed as \( R^2 \) ranged from 0.75 to 0.92.

Conclusions—The present findings support the view that a good correlation exists among the considered visual rating scales for white matter changes. With the limitation that conversion parameters are calculated by applying a linear regression to partly nonlinear scales, their use allows comparison of the results of previous studies that used different scales and to pool data from past and ongoing clinical trials. (Stroke. 2002;33:2827-2833.)

Key Words: leukoaraiosis ■ magnetic resonance imaging ■ tomography, x-ray computed ■ white matter

Cerebral white matter changes (WMCs) are frequently observed on CT and MRI scans of elderly individuals, particularly in patients with vascular risk factors, cerebrovascular diseases, and cognitive and motor impairment.\(^1\)\(^-\)\(^4\) WMCs are seen as bilateral, patchy, or diffuse areas of hypodensity on CT (leukoaraiosis)\(^5\) or hyperintensities on T2-weighted MRI scans involving the periventricular and centrum semiovale white matter. These lesions have irregular margins and do not follow specific vascular territories.

The clinical significance of WMCs is still incompletely elucidated.\(^6\) For example, the contribution of WMCs to cognitive impairment is still a matter of debate.\(^7\) In addition, the pathogenesis of WMCs is under investigation. Today, the most widely accepted opinion is that WMCs represent the radiological appearance of a vascular process linked mainly with cerebral small-vessel changes.\(^8\),\(^9\) Although the association between WMCs and aging has been consistently shown across different studies,\(^8\) conflicting conclusions have been reached as far as other risk factors and clinical correlates are concerned.

A number of causes have been hypothesized to play a role in the contradictory results achieved so far by studies evaluating the frequency, clinical significance, and risk factors of WMCs. Among these causes are (1) differences in subject settings (eg, hospital series versus population cohorts), (2) heterogeneity of patients under evaluation in terms of age and prevalence of risk factors and diseases, and (3) differences in the definition criteria of risk factors. Technical matters could be another possible cause of inconsistency; in fact, WMCs detected by CT or MRI are not exactly exchangeable because MRI is far more sensitive than CT in detecting WMCs.\(^10\),\(^11\) Moreover, among MRI studies, discordances may derive from the use of different pulse sequences and different magnetic field strengths.

It has also been argued that the use of different WMC rating methods may lead to inconsistent conclusions.\(^12\) A method to
make uniform the different visual rating scales would be useful for assessing the factual differences across studies and for conducting meta-analyses of data on the association between WMCs and clinical factors. The aims of our study were to assess the correlation of 6 visual rating scales for WMCs (2 CT and 4 MRI based) chosen among the most widely used and to determine numeric parameters that could allow the conversion of different scale scores and consequently the comparison of the results of different studies.

**Materials and Methods**

Among the WMC rating scales examined in a recent extensive review on the topic,13 we selected 3 scales for CT and 4 scales for MRI evaluation of WMCs. The CT scales were those of Rezek et al,14 van Swieten et al,15 and Blennow et al.16 The MRI scales were those of Fazekas et al,17 Scheltens et al,18 Ylikoski et al,19 and Manolio et al.20 This choice was driven by the fact that these scales were among the most commonly used with validation data available and generally considered good.13,21–22

To evaluate the feasibility of the application of each of the 7 scales on a large series of scans and to become familiar with their use, 2 investigators (L.P., M.S.) with expertise in WMC visual rating scored a first group of 30 pairs of CT and MRI scans. After this trial, we decided not to use the Rezek et al scale for this study because its application resulted to be very time-consuming. Moreover, because basal ganglia and infratentorial lesions were not scored in the remaining 5 scales, we decided to use a modification of the Scheltens et al scale that, unlike the original, did not include scoring of lesions in these 2 locations (see Appendix 1).

For the scope of our study, we selected 80 scans of subjects >60 years of age who underwent both a CT and an MRI study at a maximum of 3 weeks apart. This time interval was chosen to minimize the likelihood of major brain changes between the 2 scan examinations. In addition, no cerebrovascular event was to occur in the interval between the CT and the MRI study. Scans showing large cortical or border-zone infarcts were excluded because these lesions could influence the evaluation of WMCs. Because the aim of the study was to evaluate images, except for the above-mentioned ones, no other clinical criteria for the inclusion of the scans in the study were required. Scans were obtained through a search in the archives of our department and progressive evaluation of subjects admitted as inpatients or outpatients to our department. Attention was paid to include scans with different degrees of WMCs to minimize ceiling or floor effects on scoring. Most of the evaluated scans belonged to patients with cognitive disorders, gait impairment, or vascular risk factors, particularly hypertension.

The same 2 observers who performed the training trial did all the scan ratings. The 6 scales were applied in succession by the 2 raters together, and consensus on each scan rating score was reached. Because MRI is known to be more sensitive than CT in detecting WMCs, CT scans were always evaluated before MRI scans to avoid rating bias. Considering MRI, fluid-attenuated inversion recovery sequences were chosen for the rating when available. In the remaining cases, T2-weighted and/or proton density images were used. All CT scans were performed on a Somatom Plus-4 or Somatom CR machine; 60 MRI studies were acquired on a Philips Gyroscan T5-NT (0.5 T), and the remaining 20 were acquired on a Philips Gyroscan ACS-NT (1.5 T).

**Statistical Analysis**

Correlations between each couple of rating scales were assessed. For scales scoring periventricular and deep WMCs separately (those of Fazekas et al, Ylikoski et al, and Scheltens et al), we summed the 2 subscores to obtain a total score to be compared with that obtained on the remaining scales (those of Blennow et al, van Swieten et al, and Manolio et al). The level of significance of the relation between the rating scores of each of the 6 considered scales was tested by Spearman’s nonparametric correlation.

A linear regression model was applied for each couple of scale scores to transform the score of 1 scale to the score obtained on the second scale of the couple. The predictive value of the model was expressed as $R^2$, ranging from 0 to 1.

**Results**

The analysis of the correlations between each of the 6 rating scales is reported in the Figure. Overall, a high correlation between all the rating scales was present (Table 1).

Because the linear regression model has a clear limitation in the conversion of nonlinear rating scores such as those of the considered scales, we also present descriptive comparisons of the corresponding rating scores on the 6 scales (Table 2). Because some scales cover a wide range of scores and our observations were limited in number, we pooled some of the scores to have a sufficient number of observations for each comparison.

The regressions between the considered pairs of scales are shown in Table 3. The application of 2 parameters to a linear equation allows the conversion of the score obtained on a single scale into the score of another scale. For example, given the score of 2 on the van Swieten et al scale, the converted Blennow et al scale total score is as follows: $Y = 0.07 + 0.69 \times 2 = 1.45$. The predictive values of the conversions were generally high, with $R^2$ ranging from 0.75 to 0.92.

**Discussion**

We have found a very high correlation among 6 rating scales for the visual evaluation of WMCs. We have also been able to calculate numeric parameters to obtain the score of 1 scale once that of another scale is known. To the best of our knowledge, this is the first attempt to develop a procedure to homogenize the results of studies on WMCs that used different visual rating scales. The study of the correlation between CT- and MRI-detected WMCs is one of the identified topics on which a joint action by the International Working Group on Harmonization of Dementia Drug Guidelines focuses.28 The appraisal of WMCs represents an important step in the evaluation of patients affected by dementia or cerebrovascular diseases and in the field of age-related brain abnormalities. More recently, WMC severity has also been incorporated as part of the inclusion criteria of therapeutic trials focused on vascular dementia,23–26 and WMC load has been proposed as an outcome measure.29 Therefore, it appears essential that the methods for the evaluation of WMCs be as accurate and reproducible as possible. In this regard, it has been postulated that computerized measurement of the extent of WMCs could be the optimal solution.27 However, the use of these methods is currently limited because they are time-consuming and expensive and require specific technical equipment.28 Moreover, computerized methods are challenging to use in large studies, eg, population-based studies, or in multicentric clinical trials. On the other hand, visual rating is fast and can be applied to images of different quality obtained on different scanners because the rater can correct for variations in image contrast, resolution, and to some extent, even for differences in angulation.28 Another way of homogenizing studies on WMCs is to have a unique scale applicable to both CT and MRI. A European collaborative group has recently developed such a scale.29 This option, however, will not allow comparison of past or ongoing studies that used different scales.
The results of our study may have application not only in the evaluation and pooling of already acquired data but also in future investigations, particularly those conducted on a large scale that may use MRI and CT indiscriminately, according to local expertise and available facilities. In fact, although CT is widely recognized to be less sensitive to WMCs than MRI, it is still the most easily accessible neuroimaging procedure in many centers. However, we should point out that finding a correlation between different rating scales does not prove the validity of the scales themselves. Because the main goal of our study was to evaluate the comparability of different scales, we did not make correlation with clinical findings, nor did we assess the validity of the visual rating scales compared with a quantitative measure of WMC load. Therefore, no definite conclusion can be derived in relation to the influence of different visual rating scales on the assessment of clinical correlates of WMCs and on the validity of the scales in respect to the pathological extension of WMCs.

Possible shortcomings of our study need to be taken into account. One major limitation of visual rating scale for WMCs is the interrater variability. In our study, we decided to overcome
provided the best instructions and illustrative examples. Future studies aiming at applying this study procedure to other scales should take this aspect into consideration either by pretrial assessment of the interrater agreement or by assessment of scans by consensus.

Another caution is that the conclusions of our study can be applied only to the scales that we took into consideration. All the rating scales compared in this study have a similar approach to WMCs; some are even very close to each other. The extension of the calculation of conversion parameters to other scales requires previous demonstration of a good correlation among scale scores. From this point of view, our procedure may serve as an example. A third limitation concerns the predictive values of our conversion parameters. $R^2$ ranged between 0.75 and 0.92, meaning that in the worst case, application of the linear equation to transform the score of a scale into another reaches incorrect conclusions in one fourth of the conversions. During the study, we did not observe an improvement in the $R^2$ values because the number of the observations increased once the first 50 pairs of scans were rated. This would mean that the predictive values depend on intrinsic characteristics of the model rather than on the number of scans. On the other hand, predictive values such as those observed in our study can be considered acceptable, especially when considering the high correlations among scales. Finally, because the scale scores were not linear, the regression analysis represents an obvious simplification and thus presents limitations. The descriptive analyses we reported with the equivalent scores on different scales represent an attempt to solve this issue. However, given that some of the scales have very wide score ranges, only with a much larger number of observations for each score could a conclusion with statistical significance be reached.

From the results of this study, we conclude that a very good correlation is found among the considered scales. Further studies are needed to evaluate whether this conclusion applies to other rating scales for WMCs and to assess whether the correlation with clinical findings is influenced by the application of these scales. Conversion factors derived from our study might be applied for pooling data from past studies on age-related WMCs. Moreover, the use of these conversion factors may prove useful in the identification of subgroups of patients (eg, those with small-vessel disease-related WMCs) in ongoing clinical trials testing new drugs for vascular dementia that used CT or MRI.

**Appendix 1**

**Visual Rating Scales for CT and MRI WMCs**

**Van Swieten et al (Minimum, 0; Maximum, 4)**
The severity of WMCs on 3 subsequent CT slices (figures provided in the original article) is graded separately for the regions anterior and posterior to the central sulcus: 0 =no lesion, 1 =partly involving the white matter, and 2 =extending up to the subcortical region. The scores for the 2 regions have to be added together.

**Blennow et al (Minimum, 0; Maximum, 3)**
Extension and severity of CT WMCs are rated. The final score is the mean value between the extension and severity scores.

**Extent of WMCs**
Scores are as follows: 0 =no decrease in the attenuation of white matter; 1 =decreased attenuation of white matter at the margins at the frontal and occipital horns of the lateral ventricles; 2 =decreased attenuation of white matter at the margins around the frontal and occipital horns of the lateral ventricles with some extension toward the centrum semiovale; and 3 =decreased attenuation of white matter extending around the whole lateral ventricles and coalescing in the centrum semiovale.

**Severity of WMCs**
Scores are as follows: 0 =none, 1 =mild, 2 =moderate, and 3 =marked decrease in the attenuation of white matter.

**Fazekas et al (Total Score Minimum, 0; Maximum, 6)**
Periventricular and deep WMCs are rated separately. A total score is obtainable by summing the 2 partial scores.

**Periventricular Hyperintensities**
Scores are as follows: 0 =absence, 1 ="caps" or pencil-thin lining, 2 =smooth "halo," and 3 =irregular periventricular hyperintensities extending into the deep white matter.

**Deep White Matter Hyperintense Signals**
Scores are as follows: 0 =absence, 1 =punctuate foci, 2 =beginning confluence of foci, and 3 =large confluent areas.

**Modified Scheltens et al (Minimum, 0; Maximum, 30)**

**Periventricular Hyperintensities (Minimum, 0; Maximum 6)**
Scoring is as follows: caps, occipital 0/1/2 and frontal 0/1/2; bands, lateral ventricles 0/1/2 (0 =absent, 1 =≤5 mm, 2 =≤6 mm and ≤10 mm).

**White Matter Hyperintensities (Minimum, 0; Maximum, 24)**
Scoring is as follows: frontal 0/1/2/3/4/5/6, parietal 0/1/2/3/4/5/6, occipital 0/1/2/3/4/5/6, temporal 0/1/2/3/4/5/6 (0 =no abnormalities,
Ylikoski et al (Total Score Minimum, 0; Maximum 48)
WMCs located at 4 locations (frontal horns, body of the ventricles, trigones, and occipital horns) are rated separately on each hemisphere.

Periventricular Leukoaraiosis (Minimum, 0; Maximum, 24)
Scoring is as follows: 0=no hyperintensity; 1=punctuate, small foci (mild); 2=beginning confluent (moderate); and 3=nodular band, extending hyperintensity (severe).

Centrum Semiovale Leukoaraiosis, Including Watershed Areas (Minimum, 0; Maximum, 24)
Scoring is as follows: 0=no hyperintensity; 1=punctuate, small foci (mild); 2=beginning confluent (moderate); and 3=large confluent areas (severe). Total leukoaraiosis score is periventricular leukoaraiosis score plus centrum semiovale leukoaraiosis score equal to 0 to 48.

Manolio et al (Minimum, 0; Maximum, 9)
The total volume of white matter lesions is evaluated with template images and text descriptions. Periventricular and subcortical regions are not rated separately. A text description includes 9 grades: 0=no white matter signal abnormalities; 1=discontinuous periventricular rim or minimal “dots” of subcortical white matter; 2=thin, continuous periven-
TABLE 3. Linear Regressions Between the Scores of the 6 Evaluated Rating Scales

<table>
<thead>
<tr>
<th>X (Known Score)</th>
<th>Y (Unknown Score)</th>
<th>a</th>
<th>b</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van Swieten et al</td>
<td>Blennow et al</td>
<td>0.07</td>
<td>0.686</td>
<td></td>
</tr>
<tr>
<td>Blennow et al</td>
<td>Van Swieten et al</td>
<td>0.0571</td>
<td>1.343</td>
<td>0.922</td>
</tr>
<tr>
<td>Van Swieten et al</td>
<td>Scheltens et al</td>
<td>4.526</td>
<td>5.057</td>
<td></td>
</tr>
<tr>
<td>Scheltens et al</td>
<td>Van Swieten et al</td>
<td>-0.325</td>
<td>0.158</td>
<td>0.799</td>
</tr>
<tr>
<td>Van Swieten et al</td>
<td>Ylikoski et al</td>
<td>10.523</td>
<td>7.072</td>
<td></td>
</tr>
<tr>
<td>Ylikoski et al</td>
<td>Van Swieten et al</td>
<td>-0.764</td>
<td>0.112</td>
<td>0.789</td>
</tr>
<tr>
<td>Van Swieten et al</td>
<td>Fazekas et al</td>
<td>1.334</td>
<td>1.092</td>
<td></td>
</tr>
<tr>
<td>Fazekas et al</td>
<td>Van Swieten et al</td>
<td>-0.666</td>
<td>0.755</td>
<td>0.824</td>
</tr>
<tr>
<td>Van Swieten et al</td>
<td>Manolio et al</td>
<td>1.144</td>
<td>1.526</td>
<td></td>
</tr>
<tr>
<td>Manolio et al</td>
<td>Van Swieten et al</td>
<td>-0.239</td>
<td>0.531</td>
<td>0.810</td>
</tr>
<tr>
<td>Scheltens et al</td>
<td>Blennow et al</td>
<td>-0.172</td>
<td>0.110</td>
<td></td>
</tr>
<tr>
<td>Blennow et al</td>
<td>Scheltens et al</td>
<td>4.696</td>
<td>6.878</td>
<td>0.755</td>
</tr>
<tr>
<td>Scheltens et al</td>
<td>Ylikoski et al</td>
<td>5.177</td>
<td>1.330</td>
<td></td>
</tr>
<tr>
<td>Ylikoski et al</td>
<td>Scheltens et al</td>
<td>-1.942</td>
<td>0.671</td>
<td>0.893</td>
</tr>
<tr>
<td>Scheltens et al</td>
<td>Fazekas et al</td>
<td>0.642</td>
<td>0.196</td>
<td></td>
</tr>
<tr>
<td>Fazekas et al</td>
<td>Scheltens et al</td>
<td>-0.642</td>
<td>4.338</td>
<td>0.850</td>
</tr>
<tr>
<td>Scheltens et al</td>
<td>Manolio et al</td>
<td>0.152</td>
<td>0.276</td>
<td></td>
</tr>
<tr>
<td>Manolio et al</td>
<td>Scheltens et al</td>
<td>1.731</td>
<td>3.072</td>
<td>0.847</td>
</tr>
<tr>
<td>Ylikoski et al</td>
<td>Blennow et al</td>
<td>-0.515</td>
<td>0.079</td>
<td></td>
</tr>
<tr>
<td>Blennow et al</td>
<td>Ylikoski et al</td>
<td>10.492</td>
<td>9.810</td>
<td>0.776</td>
</tr>
<tr>
<td>Ylikoski et al</td>
<td>Fazekas et al</td>
<td>0.0484</td>
<td>0.140</td>
<td></td>
</tr>
<tr>
<td>Fazekas et al</td>
<td>Ylikoski et al</td>
<td>2.997</td>
<td>6.153</td>
<td>0.864</td>
</tr>
<tr>
<td>Ylikoski et al</td>
<td>Manolio et al</td>
<td>-0.737</td>
<td>0.200</td>
<td></td>
</tr>
<tr>
<td>Manolio et al</td>
<td>Ylikoski et al</td>
<td>6.159</td>
<td>4.406</td>
<td>0.880</td>
</tr>
<tr>
<td>Fazekas et al</td>
<td>Blennow et al</td>
<td>-0.445</td>
<td>0.535</td>
<td></td>
</tr>
<tr>
<td>Blennow et al</td>
<td>Fazekas et al</td>
<td>1.330</td>
<td>1.514</td>
<td>0.810</td>
</tr>
<tr>
<td>Fazekas et al</td>
<td>Manolio et al</td>
<td>-0.466</td>
<td>1.324</td>
<td></td>
</tr>
<tr>
<td>Manolio et al</td>
<td>Fazekas et al</td>
<td>0.717</td>
<td>0.667</td>
<td>0.882</td>
</tr>
<tr>
<td>Manolio et al</td>
<td>Blennow et al</td>
<td>-0.162</td>
<td>0.381</td>
<td></td>
</tr>
<tr>
<td>Blennow et al</td>
<td>Manolio et al</td>
<td>1.102</td>
<td>2.142</td>
<td>0.816</td>
</tr>
</tbody>
</table>

Data shown in this table can be applied to the following linear equation: 

\[ Y = a + bx \]

where a = constant and b = regression coefficient. The predictive value of the conversion is expressed as \( R^2 \).

Appendix 2

European Task Force on Age-Related White Matter Changes: Country Coordinators
J. Bogousslavsky (Lausanne, Switzerland), N. Bornstein (Tel Aviv, Israel), T. del Ser (Madrid, Spain), J. De Reuck (Gent, Belgium), B. Einarssohn (Reykjavik, Iceland), T. Erkinjuntti (Helsinki, Finland), F. Fazekas (Graz, Austria), J. Ferro (Lisboa, Portugal), A. Klimkowicz (Cracow, Poland), D. Leys (Lille, France), H. Markus (London, UK), Z. Nagy (Budapest, Hungary), D. Russell (Oslo, Norway), P. Scheltens (Amsterdam, the Netherlands), L.-O. Wahlund (Huddinge, Sweden), and G. Waldemar (Copenhagen, Denmark).

References


Visual Rating Scales for Age-Related White Matter Changes (Leukoaraiosis): Can the Heterogeneity Be Reduced?
Leonardo Pantoni, Michela Simoni, Giovanni Pracucci, Reinhold Schmidt, Frederik Barkhof and Domenico Inzitari
for the European Task Force on Age-Related White Matter Changes

Stroke. 2002;33:2827-2833; originally published online October 31, 2002;
doi: 10.1161/01.STR.0000038424.70926.5E
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2002 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/33/12/2827

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/