Intersession Repeatability of Contrast Sensitivity Scores in Age-Related Macular Degeneration

Praveen J. Patel,1 Fred K. Chen,1 Gary S. Rubin,2,3 and Adnan Tufail1

PURPOSE. To describe the intersession repeatability of contrast sensitivity (CS) measurement using Pelli-Robson charts in patients with age-related macular degeneration.

METHODS. Repeatability was calculated from three measurements of CS over a 12-week period using a standardized protocol in 107 nontreated eyes of 107 patients with age-related macular degeneration who were enrolled in an ongoing clinical trial.

RESULTS. Data from 91 patients were included in the analysis, with a 95% coefficient of repeatability of 7 letters (0.35 log CS), ranging from 6 letters for 32 eyes with drusen only to 8 letters for patients with late AMD (macular scars or geographic atrophy). Three (3%) of these stable patients had an apparent six or more letter reduction in contrast sensitivity at the week 1 visit compared with baseline.

CONCLUSIONS. There is a high intersession test–retest variability of Pelli-Robson CS scores in patients with AMD, with implications for AMD clinical trial design. Although a change criterion of six or more letters may be an adequate end point in clinical trials for patients with early AMD, a larger change criterion may be necessary for clinical trials of patients with late AMD. (Invest Ophthal Vis Sci. 2009;50:2621–2625) DOI:10.1167/iovs.08-2407

Investigators have used a range of functional outcome measures in clinical trials to assess the efficacy of new treatments for age-related macular degeneration (AMD), including the measurement of contrast sensitivity (CS).1,2 Although high-contrast distance visual acuity is an important outcome measure, several studies have shown that CS measures can provide additional information relating to functional abilities including mobility and orientation.3,4 In addition, studies have shown that after adjustment for visual acuity, individuals with better CS scores were three to five times less likely to report difficulty with vision-related tasks.5 To distinguish clinical change from measurement error at successive patient visits both in clinical trials and clinical practice, it is important to establish the repeatability of the clinical method being used. In patients with neovascular (n)AMD, it is difficult to establish the intersession repeatability of CS measurement, because of disease progression and prompt treatment. However, we can use the untreated fellow eye of patients receiving active therapy for nAMD to assess the intersession repeatability of CS scores in this patient group.

These eyes represent a clinical spectrum of disease ranging from eyes with drusen and good CS to eyes with macular scars and poor CS. Previous studies reporting the repeatability and reproducibility of the Pelli-Robson CS chart have included normal patients6–11 and those with low vision5,9,11,12 (summarized in Table 1). In these studies of repeatability and reproducibility, a range of different patient groups and different scoring methods were used for the test and 95% limits of agreement ranging from 0.14 (equivalent to approximately 3 letters) to 0.48 (equivalent to approximately 10 letters) logarithm of CS (log CS) were reported in low-vision patients. In an initial report, Pelli et al.13 advocated a line-by-line method of scoring, but subsequent investigators have suggested that letter-by-letter scoring achieves better repeatability in view of the reduced test–retest variability with this method of scoring (coefficient of repeatability; CR of 0.20 log CS for letter-by-letter scoring compared to 0.45 log CS for line-by-line scoring).16 To our knowledge, this is the first report of the intersession repeatability of CS measurement in a large cohort of patients with AMD.

METHODS

Subjects

Data from the untreated eye of patients enrolled in the Avastin (bevacizumab) for Choroidal Neovascularization (ABC) Trial (Patel P), et al. JOVS 2007;48:ARVO E-Abstract 4536) were used in this repeatability study. This ongoing clinical trial is a prospective, double-masked, randomized, controlled trial investigating the safety and efficacy of intravitreous bevacizumab (Avastin, Genentech, Inc., South San Francisco, CA) in the treatment of nAMD. All patients had consented to visual acuity testing, and the research complied with the tenets of the Declaration of Helsinki. In addition, approval for the research had been obtained both from the ABC Trial Steering Committee and the Research Governance Committee of Moorfields Eye Hospital. CS data collected at baseline, week 1, and week 12 visits from a total of 107 eyes of 107 patients were available for analysis. These eyes had a spectrum of disease from drusen to geographic atrophy and macular scars due to nAMD. For analysis purposes, eyes were classified into three AMD subgroups: eyes with early AMD with small to intermediate (<125 μm) drusen only; eyes with early AMD with large drusen (≥125 μm with or without pigment changes); and eyes with late AMD (central geographic atrophy or advanced CNV with subfoveal fibrosis not amenable to further treatment). Patients with noncentral geographic atrophy were included in the early AMD large drusen with or without pigment change category. None of these eyes had active, treatable nAMD and clinical examination, fluorescein angiography, and optical coherence tomography (OCT) imaging were used to detect disease progression at each visit.

Patients with disease progression were excluded from analysis, as the assessment of repeatability of measurements taken over 12 weeks assumes no clinical or subclinical change in disease status over this period that would cause true change in CS. Although a reasonable assumption in this cohort, the repeatability of CS was also reported for the baseline and week 1 visits only, to exclude the effect of subclinical disease progression over 12 weeks with a statistical analysis of the difference in mean CS scores between the three visits. Coefficients of

1Moorfields Eye Hospital, London, United Kingdom; the 2Department of Ophthalmology, London, United Kingdom; and the 3Biomedical Research Centre for Ophthalmology, London, United Kingdom.

Received for publication June 9, 2008; revised August 28 and October 30, 2008; accepted April 15, 2009.

Disclosure: P.J. Patel, None; F.K. Chen, None; G.S. Rubin, None; A. Tufail, None.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked ‘advertisement’ in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Adnan Tufail, Moorfields Eye Hospital, 162 City Road, London, EC1V 2PD UK; adnan.tufail@moorfields.nhs.uk.

Copyright © Association for Research in Vision and Ophthalmology

2621
trials performed the measurements during the study period and were this method of scoring. Six optometrists accredited for work in clinical
Pelli-Robson chart (Clement Clarke Inc., Columbus, OH) at a distance performed the CS measurement in a standardized protocol using the
the spherical and cylindrical components of the refractive error to give for refraction. Subjective refraction was then performed refining both
were worn, retinoscopy or autorefraction was used as a starting point
At the baseline visit, the patient’s distance-viewing spectacle prescrip-
All patients were refracted at each visit using a standardized protocol.
change in CS based on the change criteria of six or more letters or nine
repeatability were also calculated for different AMD subtypes. Finally, the number of these stable eyes with AMD showing an apparent change in CS based on the change criteria of six or more letters or nine or more letters was calculated.

**Refraction and CS Measurement**

All patients were refracted at each visit using a standardized protocol. At the baseline visit, the patient’s distance-viewing spectacle prescription was measured with a lensometer, and the result was used as the beginning approximate refraction. If no spectacles for distance vision were worn, retinoscopy or autorefraction was used as a starting point for refraction. Subjective refraction was then performed refining both the spherical and cylindrical components of the refractive error to give the final correction (at 4 m) for both eyes.

After refraction, optometrists accredited for work in clinical trials performed the CS measurement in a standardized protocol using the Pelli-Robson chart (Clement Clarke Inc., Columbus, OH) at a distance of 1 m and chart luminance of 80 to 120 cd/m². The right eye was tested followed by the left eye on charts 1 and 2, respectively, with +0.75 D added to the patient’s refraction. The patient was asked to name each letter on the chart, starting with the highest-contrast letters on the upper left-hand corner and reading horizontally across the entire line. As low-contrast letters can take some time to appear, the patient was given instructions to keep looking and not give up too soon. The optometrist circled each letter read correctly and crossed out each letter read incorrectly, with letters not attempted left unmarked. The test was stopped when the patient failed to correctly identify two or more letters in a triplet. The letter-by-letter scoring advocated by Elliot et al. was used to score the test, in view of the better repeatability of this method of scoring. Six optometrists accredited for work in clinical trials performed the measurements during the study period and were masked to previous measurements of CS when undertaking measurements.

**Statistical Analysis**

The mean CS for the entire cohort of patients at each visit was calculated, and any difference between CS scores across the three visits was analyzed by Friedman test, as Kolmogorov-Smirnov statistics showed that the CS scores were not normally distributed (P < 0.0005). In line with methods outlined by Bland and Altman, the standard deviation of CS measurements for individual patients was calculated and plotted against the mean score for each patient (for all three measurements). The mean intrasubject standard deviation (s<sub>w</sub>) was used to calculate the coefficient of repeatability (CR), defined by Bland and Altman as 1.96 × √(2s<sup>2</sup><sub>w</sub>) or 2.77s<sub>w</sub>. The difference between two measurements for the same subject is expected to be less than the coefficient of repeatability for 95% of pairs of observations. The term s<sup>2</sup><sub>w</sub> is the within-subject residual mean square in the one-way ANOVA table. This method was used for the entire cohort of patients, with additional values calculated for the three AMD diagnostic categories. To exclude the potential effect of disease progression, revised coefficients of repeatability were calculated for patients for the baseline and week 1 measurements only.

**RESULTS**

Data from 14 patients were excluded from the analysis because they had CS too poor to assess with the Pelli-Robson chart at 1 m (<2 letters read correctly at one or more visits) and data from a further 2 patients were excluded because of the presence of CNV. A total of 273 CS measurements from 91 patients

<table>
<thead>
<tr>
<th>Sample Size and Diagnosis</th>
<th>Age (y) with Range</th>
<th>Letter or Line Scoring</th>
<th>CR (log CS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lovie-Kitchen and Brown&lt;sup&gt;18&lt;/sup&gt;</td>
<td>78 normal patients</td>
<td>21–68</td>
<td>Letter</td>
</tr>
<tr>
<td>Kiser et al.&lt;sup&gt;15&lt;/sup&gt;</td>
<td>17 patients with advanced eye disease</td>
<td>20–90</td>
<td>Letter</td>
</tr>
<tr>
<td>Elliott et al.&lt;sup&gt;7&lt;/sup&gt;</td>
<td>72 normal patients</td>
<td>30 young (mean 22.5 ± 4.3 y) 42 old (mean 70.2 ± 6.7 y)</td>
<td>Line</td>
</tr>
<tr>
<td>Haymes and Chen&lt;sup&gt;14&lt;/sup&gt;</td>
<td>22 low-vision patients</td>
<td>51–92</td>
<td>Letter</td>
</tr>
<tr>
<td>Dougherty et al.&lt;sup&gt;9&lt;/sup&gt;</td>
<td>20 young, normal 17 older, normal 17 older, low vision</td>
<td>22–86 (mean 24 for young normal; 59 for older normal; 58 for low-vision group)</td>
<td>Letter</td>
</tr>
<tr>
<td>Haymes et al.&lt;sup&gt;12&lt;/sup&gt;</td>
<td>47 normal 17 AMD</td>
<td>22–77 normal (mean 48) 58–83 AMD group (mean, 73 y)</td>
<td>Letter</td>
</tr>
<tr>
<td>Rubin&lt;sup&gt;10&lt;/sup&gt;</td>
<td>24 normal 23 retinal disease</td>
<td>21–79 for normal 16–85 for patients</td>
<td>Letter</td>
</tr>
<tr>
<td>Thayaparan et al.&lt;sup&gt;11&lt;/sup&gt;</td>
<td>53 patients and normal subjects</td>
<td>Not provided</td>
<td>Letter</td>
</tr>
</tbody>
</table>

<sup>*</sup> Converted from Early Treatment of Diabetic Retinopathy visual acuity score.
were therefore included in the analysis. There were 38 male patients and 53 female patients with 39 left eyes and 52 right eyes. The number of eyes in each AMD category, visual acuity, and mean age of patients are shown in Table 2. The mean time (±SD) in days between the baseline and week 1 visual acuity measurements was 7 (±3) days. The mean interval between baseline and week 12 visits was 84 (±9) days. Of the 24 patients with late AMD, 19 had a macular scar due to previous nAMD and 5 had geographic atrophy.

The mean CS score of patients at each visit is shown in Table 3 with no significant difference between visits (P = 0.24). The plot of intrasubject SD against mean CS score is shown in Figure 1, with a modest trend to increased variability for lower CS scores. The plot of intrasubject SD against mean CS score for the data from the baseline and week 1 measurements was similar (not shown). The CR for patients is shown in Table 4, with measurements from patients with small to intermediate drusen being the most repeatable (CR = 6 letters) and measurements from patients with late AMD the least repeatable (CR = 8 letters). The overall CR for this cohort of patients with stable AMD for all visits was 7 letters (for all three measurements over 12 weeks and 6 letters for the baseline and week 1 visits only). There appears to be increased variability in the CS score ranging between 10 and 25 letters (0.35–1.10 log CS).

**Table 3. Mean CS Scores at Each Visit**

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Eyes (n)</th>
<th>Baseline</th>
<th>Week 1</th>
<th>Week 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early AMD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small/intermediate drusen</td>
<td>32</td>
<td>27 (4)</td>
<td>28 (4)</td>
<td>27 (4)</td>
</tr>
<tr>
<td>Large drusen with or</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>without pigment change</td>
<td>35</td>
<td>24 (5)</td>
<td>24 (4)</td>
<td>23 (5)</td>
</tr>
<tr>
<td>Late AMD</td>
<td>24</td>
<td>19 (8)</td>
<td>19 (8)</td>
<td>18 (7)</td>
</tr>
<tr>
<td>Total</td>
<td>91</td>
<td>25 (7)</td>
<td>25 (7)</td>
<td>25 (7)</td>
</tr>
</tbody>
</table>

Data are the Mean CS score in letters (SD).

Table 5 shows the percentage of these patients with stable AMD who were “incorrectly” classified as having improving or decreasing CS when different change criteria were used between the week 1 and baseline visits. In total, seven (8%) of these patients with stable AMD show an apparent change in CS (three decreasing and four improving CS) when the change criterion of 6 letters or more between baseline and week 1 measurements was used. There were 35 (38%) eyes with a less than 3-letter change at both 1 and 12 weeks when compared to the baseline measurement. In 76 (84%) eyes, CS did not change by 6 or more letters at both 1 week and 12 weeks.

**Table 4. CR for CS Scores by Diagnosis**

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>All Visits</th>
<th>Baseline and Week 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small/intermediate drusen</td>
<td>6 (0.30)</td>
<td>6 (0.30)</td>
</tr>
<tr>
<td>Large drusen with or</td>
<td>7 (0.35)</td>
<td>6 (0.30)</td>
</tr>
<tr>
<td>without pigment change</td>
<td>8 (0.40)</td>
<td>7 (0.35)</td>
</tr>
<tr>
<td>Late AMD</td>
<td>7 (0.35)</td>
<td>6 (0.30)</td>
</tr>
</tbody>
</table>

Data are CR in letters (log CS).

**DISCUSSION**

CS measurement is an important part of the functional assessment of patients with AMD and has been used as an outcome measure in several of the pivotal trials for the treatment of nAMD. Research also suggests that CS can provide valuable information about visual function in addition to visual acuity assessment. It is therefore important to establish the intersession repeatability of CS measurements to distinguish clinical change from measurement variability with longitudinal follow-up of patients, especially in patients with AMD, in view of the extensive research with new pharmacologic therapies in this field.

Several research groups have investigated the repeatability of CS measurement using the Pelli-Robson charts in clinical use in normal subjects and in patients with low vision (see Table 1 for summary) with Bland-Altman 95% CR ranging from 0.14 log CS (3 letters) to 0.18 log CS (~4 letters) in normal subjects and 0.2 (4 letters) to 0.48 (~10 letters) in low-vision patients.

According to data from a previous study, CS in normal subjects may range from 38 to 31 letters (1.9–1.5 log CS) in addition, scores may be at the lower end of this range in normal elderly subjects. The mean CS score in the early AMD cohort was below this normal range and is consistent with a previous study of CS in patients with drusen. This may reflect the effect of early macular dysfunction on CS.

In a 1998 study, Rubin estimated the repeatability of the Pelli-Robson CS chart in 66 normal subjects and 68 patients by

**Table 5. Number of Patients with Clinically Stable AMD Showing Apparent Change in CS between Week 1 and Baseline Visits Using Different Change Criteria**

<table>
<thead>
<tr>
<th>Change Criterion (CS letters)</th>
<th>Decreased CS</th>
<th>Improved CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥3</td>
<td>12 (13)</td>
<td>17 (19)</td>
</tr>
<tr>
<td>≥6</td>
<td>3 (3)</td>
<td>5 (5)</td>
</tr>
<tr>
<td>≥9</td>
<td>0 (0)</td>
<td>3 (3)</td>
</tr>
</tbody>
</table>

Data are the number (percentage) of clinically stable patients showing apparent CS change at week 1 visit compared to baseline.
testing individuals twice, usually in the same session. This study reported an excellent test-retest agreement with an intraclass correlation of 0.86 for patients with a range of diagnoses (including glaucoma, age-related maculopathy, and cataract) and an intraclass correlation coefficient of 0.98 for normal individuals.

We report a CR of 7 letters (0.35 log CS) for this cohort of 91 patients with stable AMD with an intraclass correlation of 0.87. CS measurement was most repeatable for the subgroups with early AMD (CR = 6–7 letters; 0.30–0.35 log CS) and least repeatable for patients with late AMD (CR = 8 letters; 0.40 log CS). There also appeared to be increased variability at the moderate range of CS loss (15–25 letters; 0.5–1.25 log CS), although there were only a few patients in this range of CS scores. It is difficult to draw definitive conclusions from our study about the relationship between CS and repeatability in patients with AMD in view of the relatively small number of patients with poor CS scores (<15 letters). A larger sample size in the lower range of CS scores is needed to investigate this further. Sources of intersession measurement variability include variability in measurement method, change in disease state, and patient factors. As in all clinical trials, we attempted to minimize variability in measurement method by using a standardized protocol and ensuring adherence to the protocol. In addition, we adopted a letter-by-letter scoring method that has been shown to be more repeatable than the original line-by-line scoring method. To address the potential effect of subclinical disease progression on CS over an interval of 12 weeks, we measured the repeatability both for the three visits over this period and for the initial two measurements separated by 1 week. The CR scores are similar for these two test–retest intervals, and there was no significant difference in mean CS scores over the 12 weeks, suggesting little disease progression over this period. In addition, no patient with data included in the analysis was noted to have disease progression on clinical examination, fluorescein angiography, or OCT imaging. Patient factors leading to variability in measurement include patient fatigue and changes in fixation between tests. In addition, patient age and early macular dysfunction may lead to more variable CS scores in patients with early AMD than in normal young subjects. The existence of multiple and variable preferred retinal loci for fixation in late AMD22 may contribute to the test–retest variability of CS measurement in this subgroup. This additional mechanism may underlie the trend to increased variability of scores in the late AMD subgroup, although in view of the small number of patients with late AMD included in this study, further studies are needed to explore whether these patients do indeed have greater variability in CS measurement than do patients with early AMD.

In this stable cohort of 91 patients with AMD, it is interesting to note that 7 had a change in CS of 6 or more CS letters when retested 1 week after the initial assessment (Table 5). This finding and the CR of 8 letters in the late AMD subgroup suggest that it may be necessary to apply a more generous change criterion for patients with late AMD. However, in view of the small number of patients in this subgroup, further studies are needed to better define the change criterion for CS in patients with late AMD. Furthermore, a change of less than 6 (in patients with small to intermediate drusen) or 8 letters (in patients with late AMD) in the mean CS score cannot be considered clinically meaningful even if statistically significant.

The advantages of this study include the large sample size and the use of a standardized protocol in a clinical trial setting to measure CS to minimize interobserver variability. In addition, we report results using data from three visits separated by 12 weeks before reporting revised data for two measurement visits separated by 1 week only in case of disease progression over the 12-week period. We excluded patients with active CNV or with CS too poor to measure with the Pelli-Robson chart. One limitation is the fact that this study was not designed specifically as a repeatability study but used data from an ongoing clinical trial. However, this may also be viewed as a strength; although studies specifically designed to measure repeatability may achieve an extremely high degree of repeatability in researchers dedicated to this goal, the values may not be easily applied to other settings. The optometrists in this study adhered to a standardized clinical trial protocol in performing measurements but without the knowledge that the repeatability of measurements would be formally assessed. This situation better reflects a true clinical trial setting in which many measurements are taken by several observers following set protocols but often without formal assessment of intersession repeatability.

In summary, this is the first study to our knowledge to report the intersession repeatability of CS in a large cohort of patients with age-related macular degeneration in a clinical trials setting. The values obtained in this study may be used to guide future clinical trial design and provides an insight into the repeatability of CS measurement in this important group of patients.

References


**ERRATUM**


In Figure 5, only the right-hand image belonged to the 20-year-old subject. The image in the unaccommodated state was from the 35-year-old subject. The correct figure is below.

![Figure 5](http://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/932960/) Example of two cross-sectional slices of the 3D MRI measurements of the 20-year-old subject in the unaccommodated and fully accommodated states.