The Effect of Optical Defocus on the Test–Retest Variability of Visual Acuity Measurements

Daniel A. Rosser,1,2 Ian E. Murdoch,1,2 and Simon N. Cousens3

PURPOSE. To determine the effect of optical defocus on the test–retest variability (TRV) of visual acuity measurements in normal subjects.

METHODS. Normal subjects underwent repeated visual acuity measurement with optical defocus of 0, 0.50, and 1.00 D. All measurements were taken using the Early Treatment Diabetic Retinopathy Study (ETDRS) version of the Bailey-Lovie logMAR chart. TRV was quantified in terms of its 95% range, both empirically and using the approach of Bland and Altman.

RESULTS. According to the Bland and Altman approach, the estimated 95% TRV ranges were ±0.11 logarithm of the minimum angle of resolution (logMAR) for 0-D defocus, ±0.18 logMAR for 0.50-D defocus, and ±0.25 logMAR for 1.00-D defocus.

CONCLUSIONS. Optical defocus has a considerable effect on the TRV of visual acuity measurements. These findings have important implications for both clinical practice and clinical research. Uncorrected refractive errors as small as 0.50 D may compromise the detection of visual change in individuals, and contribute to unnecessarily large sample sizes in clinical trials in which visual acuity is used as a primary outcome measure.

Invest Ophthalmol Vis Sci. 2004;45:1076–1079. DOI: 10.1167/iovs.03-1320

Visual acuity is the primary measure of visual function in both clinical practice and clinical research. It is well established that, when visual acuity is measured repeatedly on a given individual, the acuity score tends to vary, even in the absence of true clinical change.1–10 This variability is a form of measurement noise and is referred to hereafter as test–retest variability (TRV). Our ability to detect true clinical change using visual acuity data is inversely related to the TRV of the test used. One way of quantifying TRV is in terms of its 95% range—that is, the range of values in which, 95 times in a 100, the differences between repeated measures of acuity conducted in the absence of any change will lie. Published estimates of 95% TRV ranges vary from ±0.07 logarithm of the minimum angle of resolution (logMAR)2–10 to ±0.33 logMAR.7 Some of this variation may be explained by differences in test design and/or scoring method. The Snellen chart’s numerous limitations have been well described.11,12 Most pertinent with respect to TRV is the large-scale increment, which is predominantly a consequence of using a line-by-line scoring method. An excessively large-scale increment is known to be associated with increased TRV.2,5,10 Charts based on the design principles of Bailey and Lovie12 eliminate some of the extraneous sources of variability inherent in Snellen measurements. They also allow the use of interpolated scoring methods that result in a smaller scale increment and, in turn, lower TRV.

If we consider only those published estimates of TRV made using Bailey-Lovie type charts and interpolated scoring, the range of reported values is reduced, but still substantial (2.7-fold variation, Table 1). What are the causes of this large variation? No association between age of subjects and TRV was found in the studies of Beck et al.17 and Lovie-Kitchen and Brown.18 Reeves et al.5 suggested three other potential sources of variation in TRV: The presence or absence of disease, whether or not the measurements were conducted in a single session, and the presence or absence of uncorrected refractive error. Subjects with cataract appear at both extremes of TRV in Table 1, suggesting that presence of cataract is not a major determinant of TRV. Three of the four studies that estimated TRV over two visits reported 95% ranges of less than ±0.10 logMAR, which is not consistent with a hypothesis of increasing TRV with longer periods between examinations.

Although there has been speculation that a relationship exists between uncorrected refractive error-optical defocus and the TRV of visual acuity measurements,8 the empirical data are ambiguous. Siderov and Tiu19 reported a 95% range of ±0.16 logMAR for both aided and unaided measurements, whereas Elliott and Sheridan7 found TRV in unaided normal subjects to be three times that observed when a full correction was worn (95% ranges of ±0.21 and ±0.07 logMAR, respectively). Neither study gave details of the refractive errors in their subjects. Elliott and Sheridan7 give the mean unaided visual acuity of their population as 6/9 suggesting that the overall degree of ametropia was not high. Siderov and Tiu,19 in contrast, recruited from a university optometry clinic, and it is possible that the extent of ametropia was greater in their sample.

To our knowledge, the only published study in which the effects of optical defocus have been systematically investigated in a way that sheds light on its relationship with TRV is that of Carkeet et al.19 They assessed the effect of two discrete levels of optical defocus on the shape of the frequency-of-seeing curve as determined by Probit analysis.20 The results showed that optical defocus was associated with a flattening of the frequency-of-seeing curve. Based on this finding, Carkeet et al.19 commented that the 95% TRV range would be expected to be larger under conditions of defocus than in subjects with good correction.

The purpose of this study was to investigate whether a relationship between optical defocus and TRV of visual acuity measurements exists in normal subjects.

From the 1Institute of Ophthalmology, University College London, London, United Kingdom; 2Moorfields Eye Hospital, London, United Kingdom; 3London School of Hygiene and Tropical Medicine, London, United Kingdom.

Supported by the Trustees of Moorfields Eye Hospital National Health Service, Grant MURI 1011.

Submitted for publication December 8, 2003; revised January 6, 2004; accepted January 11, 2004.

Disclosure: D.A. Rosser, None; I.E. Murdoch, None; S.N. Cousens, 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: Dan A. Rosser, Department of Epidemiology and International Eye Health, Institute of Ophthalmology, University College London, 11-13 Bath Street, London EC1V 9EL, UK; dan.rosser@nltworld.com.

1076
Materials and Methods

Subjects

Normal subjects were recruited from the staff of Moorfields Eye Hospital (London, UK). The tenets of the Declaration of Helsinki were adhered to in full. Inclusion criteria were age 50 years or less, hyperopia not exceeding +0.50 D (mean sphere) and myopia not exceeding –10.00 D (mean sphere), astigmatism not exceeding 1.50 D, absence of any ocular abnormality including media opacity, no history of ocular abnormality including amblyopia, no history of regular use of the ETDRS logMAR chart, and acuity better than +0.20 logMAR (Snellen equivalent 6/9.5).

Testing Procedure

Before testing, each subject underwent formal subjective refraction with binocular balancing using the intermediate contrast technique of Humphris and Woodruff, and an end point of maximum plus/minus minus for maximum visual acuity. The chart used for refraction was for that purpose only. One eye only of each subject was assessed. When both eyes met the inclusion criteria, the right eye was used as the study eye. Each subject then underwent two acuity measurements with each of three refractive corrections: full refractive correction (0 D optical defocus), full refractive correction plus +0.50 D defocus, and full refractive correction plus +1.00 D defocus.

Each subject therefore underwent six acuity measurements using the ETDRS logMAR acuity chart in a session lasting approximately 15 to 20 minutes. This chart is based on the design suggested by Bailey and Lovie, but incorporates the recommendations of the National Academy of Sciences-National Research Council (NAS-NRC). The chart has been described in detail by Ferris et al. Three versions of the ETDRS logMAR chart were used for the measurements at 0 D defocus, +0.50 D defocus, and +1.00 D defocus: the refraction chart, chart 1, and chart 2, respectively. The refraction chart was used for the measurements at zero defocus because the control of letter difficulty is less rigorous than in charts 1 and 2. Accordingly, if the refraction chart was associated with a higher degree of TRV than charts 1 and 2, it would lessen any measured effect of defocus on TRV rather than exacerbate it. To limit memory effects, subjects were required to read the chart forward for one measurement and backward for the other. The order of the six measurements was randomized (with the sole restriction of the same chart never being used for consecutive measurements) to limit learning or fatigue effects as potential confounders. All charts were viewed from a distance of 6.5 m. This results in a range of effective letter sizes from +0.80 to +1.00 logMAR (Snellen equivalent 0.579–6/1.9) as opposed to the range at the standard 4-m distance of +1.00 to –3.0 logMAR (Snellen equivalent 6/60–6/3). The 6.5-m testing distance was then rejected, because with good visual acuity from reading all or part of the bottom line on the chart. It is desirable to prevent this, because the resultant truncation or ceiling effects may artificially reduce the TRV of those with good acuity. A longer testing distance was avoided, because for a testing distance of 6.3 meters, there would be three rows of letters above the +0.50 logMAR letter size, which is the expected threshold for 1.00 D of optical defocus.24,25 The subject’s responses were recorded on specially designed data pro formas.

Scoring and End Points

Each chart was attempted using a forced-choice testing paradigm. The end point for each measurement was taken as a full row of errors. Such a termination rule is desirable to limit TRV with Bailey-Lovie type logMAR charts.26 Each chart was scored using an interpolated scoring method, whereby credit is given for each letter correctly named.11,27 As mentioned, this method tends to limit TRV through the use of a small-scale increment. For each subject, at each level of optical defocus, the difference in the two acuity measurements was calculated by subtracting the score of the backward measurement from that of the forward measurement.

Statistical Methods

For each individual, the difference in the two acuity measurements at a given degree of defocus was calculated. The SD of these differences across all 40 individuals was estimated. TRV was quantified in terms of its 95% range. This 95% range was estimated in two ways: directly from the observed distribution and using the approach of Bland and Altman. The chart was scored using an interpolated scoring method, whereby credit is given for each letter correctly named. As mentioned, this method tends to limit TRV through the use of a small-scale increment. For each subject, at each level of optical defocus, the difference in the two acuity measurements was calculated by subtracting the score of the backward measurement from that of the forward measurement.

Results

Forty subjects were recruited. Age at last birthday ranged from 21 to 50 years (mean, 33 years). Refractive error ranged from +0.50 to –8.75 D (median spherical equivalent, –1.25 D). Astigmatism ranged from 0 to 1.25 D (median, 0 D). Acuity ranged from –0.32 to +0.12 logMAR with a median of –0.14 logMAR (Snellen equivalents: range 6/3–6/8, median 6/4.3). Three individuals had acuity worse than 0.00 logMAR (Snellen 6/6), one of which was worse than +0.10 logMAR (Snellen 6/7.5). The mean difference at each degree of defocus was close to 0 (Table 2), indicating that the chart was no more difficult when attempted backward. There was very strong evidence that the standard deviation of the differences varied with the degree of defocus (P = 0.0002), with the standard deviation

### Table 1. Published Estimates of Test–Retest Variability Using Charts Based on Bailey and Lovie’s x

<table>
<thead>
<tr>
<th>Author</th>
<th>95% TRV Range (mm)</th>
<th>Test Sessions (n)</th>
<th>Subjects</th>
<th>Refractive Correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elliott and Sheridan</td>
<td>±0.07</td>
<td>2</td>
<td>Normal</td>
<td>Full</td>
</tr>
<tr>
<td>van den Brom et al.</td>
<td>±0.08</td>
<td>1</td>
<td>Cataract</td>
<td>Full</td>
</tr>
<tr>
<td>Elliott and Sheridan</td>
<td>±0.09</td>
<td>2</td>
<td>Cataract</td>
<td>Full</td>
</tr>
<tr>
<td>Arditi and Cagengello</td>
<td>±0.09</td>
<td>2</td>
<td>Normal</td>
<td>Full</td>
</tr>
<tr>
<td>Bailey et al.</td>
<td>±0.10</td>
<td>1</td>
<td>Normal</td>
<td>Full</td>
</tr>
<tr>
<td>Rosser et al.</td>
<td>±0.11</td>
<td>1</td>
<td>Normal</td>
<td>Full</td>
</tr>
<tr>
<td>Lovie-Kitchen</td>
<td>±0.16</td>
<td>1</td>
<td>Normal</td>
<td>Unaided</td>
</tr>
<tr>
<td>Rosser et al.</td>
<td>±0.18</td>
<td>1</td>
<td>Cataract</td>
<td>Habitual</td>
</tr>
<tr>
<td>Reeves et al.</td>
<td>±0.19</td>
<td>2</td>
<td>Mixed</td>
<td>Full</td>
</tr>
</tbody>
</table>

* Estimated using the approach of Bland and Altman.16
(and hence the TRV) increasing as the level of optical defocus increased. At 0 D defocus the observed 95% TRV range and the 95% TRV range based on the assumption of normality were identical for practical purposes, with no evidence of non-normality (P = 0.38). At +1.00 D defocus the observed 95% range was slightly wider than that predicted by normal theory. However, the distribution of the observed data was consistent with normality (P = 0.38), and the confidence intervals for the upper and lower bounds derived from the observed distribution both included the bounds predicted by normal theory. At +0.50 D defocus, the observed 95% TRV range showed a degree of asymmetry and there was weak evidence against the hypothesis of normality (P = 0.04), largely due to one observation with an outlying value of +0.30 logMAR. Excluding this observation from the analysis resulted in an observed 95% range of −0.20 to +0.10 logMAR and a predicted 95% range based on normal theory of ±0.15 logMAR. Exclusion of this observation from the comparison of standard deviations across the three degrees of defocus did not compromise the strength of the evidence against the null hypothesis of equal standard deviations (P = 0.00005).

**DISCUSSION**

The results of this study suggest that optical defocus strongly influences the TRV of visual acuity data. This finding is consistent with the work of Carkeet et al.\(^6\) who demonstrated a significant flattening of the frequency-of-seeing curve (as described using Probit size\(^6\)) with optical defocus of 1.00 or 2.00 D. The data suggest that the increase in TRV may be considerable, even with degrees of optical defocus as small as 0.50 D, causing a blur that is consistent with an acuity of approximate, even with degrees of optical defocus as small as 0.50 D, largely due to one observation with an outlying value of +0.30 logMAR. Excluding this observation from the analysis resulted in an observed 95% range of −0.20 to +0.10 logMAR and a predicted 95% range based on normal theory of ±0.15 logMAR. Exclusion of this observation from the comparison of standard deviations across the three degrees of defocus did not compromise the strength of the evidence against the null hypothesis of equal standard deviations (P = 0.00005).

**REFERENCES**


**Table 2. Mean Differences between Paired Acuity Measurements and their 95% Ranges by Degree of Optical Defocus**

<table>
<thead>
<tr>
<th>Degree of Optical Defocus (D)</th>
<th>Mean Difference (logMAR)</th>
<th>SD</th>
<th>Observed 95% TRV Range</th>
<th>Predicted 95% TRV Range* (Assuming Normality)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.002</td>
<td>0.051</td>
<td>−0.100 to +0.100</td>
<td>±0.10</td>
</tr>
<tr>
<td>+0.50</td>
<td>−0.013</td>
<td>0.095</td>
<td>−0.200 to +0.295</td>
<td>±0.18</td>
</tr>
<tr>
<td>+1.00</td>
<td>−0.004</td>
<td>0.124</td>
<td>−0.279 to +0.259</td>
<td>±0.24</td>
</tr>
</tbody>
</table>

* ±1.96 SD.