DA Rosser<sup>1,2</sup>, IE Murdoch<sup>1</sup>, FW Fitzke<sup>3</sup> and DAH Laidlaw<sup>4</sup>

npg

# Improving on **ETDRS** acuities: design and results for a computerised thresholding device

#### Abstract

Aims All visual acuity data are subject to test-retest variability (TRV). This measurement error obscures true clinical change and reduces the statistical power of clinical trials using acuity as a primary outcome measure. This study was designed to assess whether a computerised system can reduce TRV by taking repeated acuity measurements and averaging them. A computerised system (PC-test) was developed for this purpose and compared in terms of TRV with the current Gold Standard ETDRS logMAR chart.

Methods A total of 19 subjects with a mean acuity of +0.16 logMAR (range +0.49 to -0.10 logMAR) were recruited. The performance of two computerised tests (one averaging 10 repeats and one five) was compared with that of the ETDRS logMAR chart in terms of TRV and agreement of acuity data.

Results The 10 and five repeat computerised tests (PC-tests) produced a TRV of  $\pm 0.11$  and  $\pm$ 0.10 logMAR, respectively, compared with  $\pm$ 0.18 logMAR for the ETDRS chart. No significant bias was observed between PC-test and ETDRS acuities.

Conclusions A computerised system that takes repeated acuity measurements and averages them is subject to less TRV than a single ETDRS acuity measurement. A reduced TRV of visual acuity data allows earlier detection of true clinical change in individual patients. It also allows smaller differences between groups to be detected in clinical trials for a given degree of statistical confidence and power.

Eye (2003) 17, 701-706. doi:10.1038/ sj.eye.6700496

Keywords: logMAR; visual acuity; ETDRS

#### Introduction

The Bailey–Lovie logMAR visual acuity chart,<sup>1</sup> along with the version produced by Ferris et al<sup>2</sup> (known as the 'ETDRS' chart), have become the gold standard tool for the measurement of visual acuity in prospective clinical research.<sup>3,4</sup> Nevertheless, even in the absence of any clinical change, consecutive visual acuity measurements on a given subject using such charts are subject to a degree of variability. This variability can be thought of as a form of measurement noise, and will be referred to throughout as test-retest variability (TRV). The ability to detect true visual change decreases as the TRV of the acuity data increases. Increased TRV may therefore compromise the management of ophthalmic patients over time. The published data suggest that changes in acuity of around two lines of letters may be required to achieve significance at the 5% level.5-8 Increased TRV also has implications for the design of any clinical trial which uses acuity as a primary outcome measure.9 For such a trial, any increase in the TRV of acuity will necessitate an increased sample size to demonstrate a given clinical change with a given degree of statistical power.

The repetition and averaging of measurements to reduce measurement noise is a commonly used experimental technique, which can be readily applied to visual acuity measurement via the use of a personal computer. Use of a computer allows the presentation of stimuli in a random order (to avoid memorisation effects) as well as automated processing and statistical analysis of subject responses.

Our aim was to assess whether a computerised acuity test can produce lower levels of TRV (and hence allow earlier detection of change) than the Gold Standard ETDRS logMAR chart through repeating and averaging

<sup>1</sup>Department of Epidemiology and International Eye Health Institute of Ophthalmology University College London London, UK

<sup>2</sup>Optometry Department Moorfields NHS Trust City Road, London, UK

<sup>3</sup>Department of Visual Science Institute of Ophthalmology University College London Bath Street, London, UK

<sup>4</sup>Vitreo-Retinal Unit St Thomas' Hospital London, UK

Correspondence: DA Rosser Department of Epidemiology and International Eye Health Institute of Ophthalmology University College London Bath Street London EC1V 9EL, UK Tel: +44 20 7608 6896 Fax: +44 20 7250 3207 E-mail: danrosser@ blueyonder.co.uk

Received: 22 July 2002 Accepted in revised form: 20 December 2002

The authors have no proprietary interests

a series of acuity threshold measurements. This computerised form of test will be referred to throughout as a 'PC-test'. TRV was defined after Bland and Altman<sup>10</sup> as the 95% confidence limits of agreement ( $\pm$ 1.96 SDs of the distribution of differences between paired acuity measurements).

# Materials and methods

## Overview

702

A series of clinically stable subjects had their acuity measured using each of three acuity tests:

- 1. The ETDRS logMAR chart (reference standard),
- 2. PC10-test (computerised test using 10 averaged measurements), and
- 3. PC5-test (computerised test using five averaged measurements).

The measurements were repeated not less than 2 weeks later such that the TRV of each test could be assessed, and any systematic bias evaluated.

# Subjects

Subjects were recruited from the outpatient clinics of Moorfields Eye Hospital who fulfilled the following inclusion criteria:

- (a) Able to understand and comply with the testing protocol.
- (b) Stable visual acuity determined via assessment of current clinical diagnosis.

One eye of each subject was assessed. Where both eyes met the criteria, the eye with the poorer acuity was used as the study eye. All subjects wore their habitual spectacle correction and viewed the acuity tests from a distance of 4 m.

# Equipment

## ETDRS logMAR chart (Lighthouse International)

Display: The printed panel charts were back-lit in the standard Lighthouse box achieving a luminance of  $111 \text{ cd/m}^2$ .

Acuity stimuli: The ETDRS chart has five letters per row ranging in size from +1.0 to  $-0.30 \log$ MAR in 0.1 logMAR steps (see Figure 1). The chart has been described in detail by Ferris *et al.*<sup>2</sup> Versions 1 and 2 were used.

*Testing paradigm:* Subjects were required to attempt each letter on the chart until they responded to all the letters on a single row incorrectly,<sup>11</sup> at which point the test was terminated. In accordance with usual practice, the ETDRS test consisted of a single reading of the chart. *PC-tests (PC10-test and PC5-test)* 

*Display*: Stimuli were displayed using a 20-in CRT monitor (Ergovision 2040, Taxan Europe Ltd) using a resolution of  $1024 \times 768$  and a noninterlaced refresh rate of 85 Hz. The monitor was driven by a standard IBM compatible 150 MHz PC. After a 30 min warm-up period, luminance was measured at 31 cd/m<sup>2</sup>.

*Acuity stimuli:* Three horizontal rows of three letters were displayed, the centre row of the three containing the three test stimuli. The upper and lower rows of letters were included to simulate the contour interaction provided by the surrounding letters on the ETDRS chart (see Figure 1). The display format followed that of the ETDRS convention in terms of interletter spacing, interline spacing, use of the Sloan set of letters, and a 0.1 logMAR unit increment between lines. The letters were generated at random, the only restriction being no repeats on a given line. All letters were generated to ETDRS specifications.<sup>2</sup>

*Testing paradigm:* Each computerised test contains a series of individual acuity measurements. As used in this study, the test commences at a stimulus size of +0.8 logMAR (although poorer acuities could be measured by



Figure 1 Arrangement of letters on PC-test and ETDRS chart. (a) PC-test captured display. (b) ETDRS chart.

reducing the viewing distance). The examiner enters the subject's responses to the three stimuli at this size level. Providing at least one of the three responses is correct, the stimulus size decreases progressively by 0.1 logMAR. This represents a 'scrolling' of the chart down to the next line such that the dimensions of stimuli, surrounding optotypes and the spaces between them all decrease by 0.1 log units. Once the subject has made a full line of incorrect responses, the computer calculates and stores the acuity score before initiating the next in the series of measurements by increasing the stimulus size by 0.5 logMAR and repeating the process. Once the specified number of measurements within the series (in this case either 10 or 5) has been completed, the computer averages the individual measurement values to produce the final PC-test result. The 0.5 logMAR increase in stimulus size between consecutive measurements was incorporated to save time by not requiring subjects to read letters that are so far above threshold, that there is a negligible chance of them being misnamed. This magnitude of 0.5 logMAR was chosen based on unpublished data from this group.

## Scoring

All acuity scores were calculated using the interpolated method described by Ferris *et al*<sup>2</sup> such that credit is given for each letter correctly named. This method is known to produce less TRV than the line-assignment method commonly encountered in clinical practice.<sup>12,13</sup> For the ETDRS test, the score is derived from the total number of correctly named letters, whereas producing the final score for the PC-tests requires the additional step of averaging the series of individual measurements that make up each PC-test result.

## Investigations

Four acuity measurements were taken on one eye of each subject using ETDRS charts 1 and 2, as well as the PC10-test and PC5-test in random order. Following an interval of not less than 2 weeks, the subjects attended for a second visit during which they underwent repeat testing using the same four tests, again in random order.

## Outcome measures

The objectives were:

(a) To determine in terms of 95% confidence limits, the TRV of acuity data for the ETDRS chart, the PC10-test, and the PC5-test.

Table 1	Summary	of	acuity	tests	
---------	---------	----	--------	-------	--

Test	No. of thresholds per measurement
ETDRS	1
PC10	10
PC5	5

(b) To determine in terms of mean difference and 95% confidence intervals for the mean, the extent to which measurements taken using the PC10-test and the PC5-test agreed with those of the Gold Standard ETDRS chart.

The methods of Bland and Altman<sup>10</sup> were used to determine (a) and (b) above.

# Results

A total of 19 subjects were recruited. The range of acuities was from  $\pm 0.64$  to -0.20 (median  $\pm 0.12$ ) logMAR (as measured with the ETDRS chart). Table 2 shows TRV in terms of 95% confidence limits ( $\pm 1.96$  SDs of the differences between paired measurements). TRV as measured for the PC10-test and PC5-test was  $\pm 0.11$  and  $\pm 0.10$  logMAR, respectively (see Table 2). This represents reductions of 39 and 44% respectively on the level of measurement variability compared with the value of  $\pm 0.18$  logMAR achieved by the ETDRS chart.

Table 3 shows the mean of the differences between paired measurements with the 95% confidence intervals. For both comparisons (PC10-test *vs* ETDRS, and PC5-test *vs* ETDRS) the mean difference is greater than zero; however, in both cases the 95% confidence interval includes zero. Hence, there is no indication of notable systematic bias. The similarity between the mean differences suggests that increasing the number of repeats from five to 10 does not affect the absolute acuity value.

Figure 2a and b show Bland–Altman plots for the ETDRS and PC10-test and PC5-test data, respectively. These show the difference between paired measurements plotted against their mean. The greater spread of points

Table 2	Test-retest	variability	(TRV)
---------	-------------	-------------	-------

Measurement tool	No. of repeats	TRV (logMAR) <sup>a</sup>
PC10-test	10	$\pm 0.11 \log$ MAR
PC5-test	5	$\pm 0.10 \log$ MAR
ETDRS <sup>b</sup>	N/A	$\pm 0.18 \log$ MAR

 $^a95\%$  Confidence limits for TRV (±1.96 SDs of the paired differences).  $^bAverage$  for ETDRS charts 1 and 2.

Table 3 Agreement with ETDRS

Measurement tool	Mean difference <sup>a</sup> (logMAR)	95% Confidence interval <sup>b</sup> (logMAR)
PC-test (10 repeats)	+0.026	-0.015 - +0.067
PC-test (five repeats)	+0.016	-0.026 - +0.058

<sup>a</sup>The mean of the differences between paired PC-test and ETDRS acuities (ETDRS subtracted from PC-test).

<sup>b</sup>95% Confidence interval for the mean.



**Figure 2** Bland–Altman plot for (a) ETDRS chart and (b) PC-test.

in Figure 2a as compared with Figure 2b indicates the greater variability between paired measurements on the same individual subject for ETDRS as compared with the PC-tests.

#### Discussion

A previous study conducted by this group demonstrated levels of TRV for the ETDRS chart of  $\pm 0.18 \log MAR.^8$  Based on this finding, a subject's visual acuity measurement is required to deteriorate (or improve) by

at least 0.20 log units (equivalent to a 58% increase in letter size) before the change exceeds the test's measurement error and can therefore be deemed 'real'. In this method-comparison study, the average of a series of five acuity measurements as measured using the PC5test produced a TRV of  $\pm 0.10$  logMAR. TRV for the ETDRS chart was again measured at  $\pm 0.18$  logMAR. Hence, repeating and averaging appears to produce a considerable reduction in TRV with a commensurate improvement in the ability of the test to detect change. However, differences exist between the displays of the two tests (eg with respect to luminance), which may confound the effects of repeating and averaging. Accordingly, additional evidence for the main determinants of TRV in this study was gained by recalculating TRV for the PC5 and PC10-tests using data from only the first repeat of each test. This results in TRV increasing by more than twofold to  $\pm 0.28 \log$ MAR and  $\pm 0.25$  logMAR for the PC5 and PC10-tests, respectively. While not ruling out any influence of display type upon TRV, this increase suggests that, for small numbers of repeats, the level of TRV is strongly influenced by the number of repeats.

It should be noted that there is considerable variation in the published levels of TRV for the ETDRS chart. The level of TRV for ETDRS acuities in this study ( $\pm 0.18$  logMAR) is consistent with a number of published research papers.<sup>5–8</sup> However, some workers including Elliott and Sheridan<sup>14</sup> and Arditi and Cagenello<sup>15</sup> have achieved levels of TRV using the ETDRS chart, which are similar to those achieved using the PC5-test in this study.

Reeves et al<sup>16</sup> suggested 'ceiling effects' as a potential reason for the discrepancy between the level TRV measured in his study ( $\pm 0.19 \log$ MAR), and that published by Elliott and Sheridan<sup>14</sup> ( $\pm 0.07 \log$ MAR). He proposed that repeated acuity measurements on subjects with very good vision may produce an artificially low level of TRV because the measurements are truncated by the end of the scale. Although feasible, this effect appears not to explain the difference between Reeves' and Elliott's data as Elliott's own results for subjects with cataract produced TRV levels of  $\pm 0.09 \log$ MAR, despite these subjects having acuities that did not approach the bottom of the scale (maximum acuity  $+0.40 \log MAR$ ). Reeves et al16 also suggested uncorrected refractive error as a potential confounder for TRV, while admitting that this alone could not explain the difference in TRV between his study and that of Elliott and Sheridan. Indeed, the more recent data of Siderov and Tiu<sup>5</sup> suggest that the correcting refractive error has no effect on TRV. Having raised ocular pathology as another potential confounder, Reeves proceeded to discount this based on his own study,<sup>16</sup> in which he found no difference in TRV between subgroups of normal and abnormal eyes. This



finding also appears to be supported by Elliott's data in which TRV levels for normals and subjects with cataract were within  $\pm 0.02$  logMAR (one ETDRS letter) of one another. A final potential confounder suggested by Reeves *et al* was that of the time interval between test and retest. This also appears to be discounted by the available data, as this study, along with those of Elliott, Arditi, Reeves, and Siderov, shows varying levels of TRV despite employing similar time intervals between test and retest.

In spite of the fact that the discrepancies between published levels of TRV are difficult to account for, the fact that the levels of TRV produced by the PC5-test and the ETDRS test in this study ( $\pm 0.10$  and  $\pm 0.18$  logMAR, respectively) were measured on a single group of subjects under identical conditions remains compelling evidence for the effect of repeating and averaging on TRV.

The benefit of the approach of repeating and averaging appears to be limited to not more than five repeated measurements as we can see from the TRV figure of  $\pm 0.11$  logMAR produced by the PC10-test. This test differs from the PC5-test only in the number of individual measurements, which are taken and averaged to produce the final acuity result, and yet the PC10-test produced a slightly higher (equivalent to half an ETDRS letter) level of TRV than the PC5-test. If we consider the series of measurements within each PC-test as a sample from an infinite distribution of repeated measurements, we would expect diminishing returns with respect to improved (lower) TRV as the sample size increases. This is because the variability of the mean of the sample is inversely proportional to the square root of the sample size. The absence of any improvement in TRV with a larger number of averaged measurements might suggest that beyond five repeated and averaged measurements, a factor other than short-term measurement noise is the limiting factor preventing further reduction in TRV. An example of such a factor might be a type of long-term fluctuation in visual acuity analogous to that seen in perimetric defects in glaucoma. This suggestion is, however, speculative and would require further study. In fact a small increase in TRV was observed with the PC10-test as compared with the PC5-test. The fact that this is contrary to the expected finding and may be explained by increased fatigue associated with a longer test duration.

As is common in method-comparison studies, the prototype PC-tests were designed to produce acuities which agree well with those of the Gold Standard ETDRS test. The results suggest that there is no systematic bias between acuities measured with either PC-test and the Gold Standard ETDRS test, when used in this population. It is interesting that none of the departures from the ETDRS chart design have introduced measurement bias. Such departures include:

- 1. The use of three letters per line rather than five (because of finite display screen size),
- the use of letters chosen at random (although without adjacent repeats) rather than being grouped such that the total difficulty of all lines of letters was as similar as possible, and
- 3. lower background illumination.

Factors 1 and 2 might be expected to compromise the level of TRV; factor 1 because of the increased scale increment resulting from a smaller number of letters per line, and factor 2 because variation in average letter difficulty between lines is a potential confounder. It appears that any such factors tending to increase TRV were more than compensated for by the reduction in TRV produced by repeating and averaging. Factor 3, along with factor 1, might be expected in introduce systematic bias; factor 1 because the letters on a three-letter-per-line chart are less crowded, and factor 3 because visual acuity is known to be proportional to chart luminance. As no significant bias was encountered, either these factors were insufficient to result in bias, or that they produced opposing biases of approximately equal magnitude. A factor that might produce acuity-dependent bias between the PC-tests and the ETDRS chart is the one related to the pixelated nature of the PC-test display. Pixelisation is likely, if present, to have a more detrimental effect upon good acuities (for which stimulus sizes are small) than poor acuities, where the pixel size is small relative to letter size. Examination of the data showed no variation in the level of agreement between the PC5 and ETDRS tests with the underlying acuity, suggesting that pixelisation did not influence legibility of the smaller letters.

One benefit of reduced TRV not yet discussed relates to the ability to detect differences between groups in clinical research. The effect of reduced TRV upon a clinical trial, which uses visual acuity as a main outcome measure, is best illustrated by an example. Using the above results, a study using ETDRS acuity data to compare the outcome of two treatments would need to recruit a total of 136 subjects to show a difference in mean acuity outcome between the groups of 0.10 logMAR, at the 5% significance level and with a power of 90%. Using the PC5-test, the required number of subjects could be reduced to 42.

In summary, this study suggests that repeating and averaging acuity measurements using a computerised visual acuity test may produce lower levels of TRV than the ETDRS logMAR chart. Reduced TRV allows earlier detection of true visual change in individuals, and for clinical trials using visual acuity as a primary outcome measure enables differences between groups to be demonstrated with a smaller number of subjects. A larger study is required to confirm this finding as well as to further investigate whether the benefits of repeating and averaging are dependent upon the underlying level of acuity.

# Acknowledgement

This study was supported by a grant from the Special Trustees of Moorfields Eye Hospital NHS Trust (No. MURI 1011).

## References

- 1 Bailey L, Lovie E. New design principles for visual acuity letter charts. *Am J Optom Physiol Opt* 1976; **53**: 740–745.
- 2 Ferris FL, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. *Am J Ophthalmol* 1982; 94(1): 91–96.
- 3 National Acadamy of Sciences—National Research Council Committee on Vision. Report of working group 39: recommended standard procedures for the clinical measurement and specification of visual acuity. Adv Ophthalmol 1980; 41: 103.
- 4 Ferris FL, Bailey I. Standardizing the measurement of visual acuity for clinical research studies: guidelines from the Eye Care Technology Forum. *Ophthalmology* 1996; **103**(1): 181–182.
- 5 Siderov J, Tiu AL. Variability of measurements of visual acuity in a large eye clinic. *Acta Ophthalmol Scand* 1999; **77**(6): 673–676.

- 6 Lovie Kitchin JE. Validity and reliability of visual acuity measurements. *Ophthalmic Physiol Opt* 1988; 8(4): 363–370.
- 7 Reeves B, Wood J, Hill A. Vistech VCTS 6500 charts, withinand between-session reliability. *Optom Vis Sci* 1991; **68**(9): 728–737.
- 8 Rosser DA, Laidlaw DA, Murdoch IE. The development of a 'reduced logMAR' visual acuity chart for use in routine clinical practice. *Br J Ophthalmol* 2001; **85**(4): 432–436.
- 9 Gordon MO, Schechtman KB, Davis LJ, McMahon TT, Schornack J, Zadnik K. Visual acuity repeatability in keratoconus: impact on sample size. Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study Group. Optom Vis Sci 1998; 75(4): 249–257.
- 10 Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; i: 307–310.
- 11 Carkeet A. Modeling logMAR visual acuity scores: effects of termination rules and alternative forced-choice options. *Optom Vis Sci* 2001; **78**(7): 529–538.
- 12 Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. *Invest Ophthalmol Vis Sci* 1991; **32**(2): 422–432.
- 13 Vanden Bosch ME, Wall M. Visual acuity scored by the letter-by-letter or probit methods has lower retest variability than the line assignment method. *Eye* 1997; **11**: 411–417.
- 14 Elliott DB, Sheridan M. The use of accurate visual acuity measurements in clinical anti-cataract formulation trials. *Ophthalmic Physiol Opt* 1988; **8**(4): 397–401.
- 15 Arditi A, Cagenello R. On the statistical reliability of letterchart visual acuity measurements. *Invest Ophthalmol Vis Sci* 1993; 34(1): 120–129.
- Reeves BC, Wood JM, Hill AR. Reliability of high- and low-contrast letter charts. *Ophthalmic Physiol Opt* 1993; 13(1): 17–26.