VISUAL ACUITY SCORED BY THE LETTER-BY-LETTER OR PROBIT METHODS HAS LOWER RETEST VARIABILITY THAN THE LINE ASSIGNMENT METHOD

MICHAEL E. VANDEN BOSCH[1] and MICHAEL WALL

SUMMARY

Purpose: **The optimal method for scoring visual acuity measures is unknown. Our goal was to determine, in a clinical setting, the method of scoring visual acuity with the lowest test–retest variability.**
Methods: **We investigated the effect of three different scoring methods using the Early Treatment Diabetic Retinopathy Study (ETDRS) visual acuity chart comparing 32 patients with macular disease and 38 age-matched normal subjects. All subjects completed six repetitions of ETDRS charts. Three scoring methods were then used (line assignment, ETDRS or letter-by-letter and probit), the results were converted to log MAR values and the test–retest variabilities analysed.**
Results: **We found significant differences in variability among the three scoring methods ($p<0.0001$). The variability was greatest with the line assignment method and less with the ETDRS and probit methods. The ETDRS and probit methods had similar variabilities. The difference in variability between normals and patients was not statistically significant. There were no differences in the calculated visual acuities among the three methods, only the variabilities. Using the ETDRS or probit methods, the within-test standard deviation was about 0.04 log MAR units (two letters).**
Conclusion: **Test–retest variability of visual acuity measurements is lower using the ETDRS or probit methods than the traditional line assignment method.**

Snellen introduced the visual acuity chart in 1862.[1] Variations of the chart have evolved and various optotypes and scoring methods have been proposed.[2–18] For screening purposes, almost any chart that can assess a full range of visual angles is sufficient. Evaluating the progression of a disease process or response to treatment, however, requires that the test be accurate and precise. The test–retest variability of visual acuity measurements comparing methods of scoring has had little study.

Several articles mention that visual acuity improves with the second test;[5,8] however, these studies were not designed to evaluate learning effect or variability. Gibson and Sanderson[19] invited ophthalmology outpatients with lens opacities back for a repeat eye examination. Sixty-four of 300 accepted. Visual acuity examinations were performed twice (once each by a different nurse). They found only one-third had the same visual acuity on the two occasions, with 13% differing by two lines or more.

Elliot and Sheridan[20] had 21 subjects read an Early Treatment Diabetic Retinopathy Study (ETDRS) chart, three times on two occasions, 1 week apart. No significant difference was found between the two readings of the best corrected visual acuity. Bailey and colleagues[3] showed that test–retest variability was related to the coarseness of the scoring scale; that is the line assignment method had a higher variability than the letter-by-letter method because its scale was coarser.

Arditi and Cagenello[2] measured the test–retest variability of a standard acuity chart in highly trained normal subjects in a controlled laboratory setting, thereby minimising the variability. They also concluded the line assignment method had a higher variability than letter-by-letter scoring. In addition to these two methods, it has been suggested that the best way to score visual acuity is by using a probit analysis that defines visual acuity as the estimate of the letter size seen 50% of the time by the probit curve fitting function.[6,17]

Although the line assignment method has a higher variability than the letter-by-letter method and the

From: Departments of [1]Ophthalmology and [2]Neurology, Veterans Administration Medical Center, Iowa City, and University of Iowa, College of Medicine, Iowa City, Iowa, USA.
Correspondence to: Michael Wall, MD, University of Iowa, College of Medicine, Department of Neurology, 200 Hawkins Drive, #2007 RCP, Iowa City, IA 52242-1053, USA. Tel: +1 (319)-356-8758. Fax: +1 (319)-356-4505.

probit method appears to have a low variability, these three methods have not been directly compared. Furthermore, the effect of the method of scoring on variability in patients with damage to the sensory visual system has not been investigated for any of these comparisons. Therefore, to compare these three methods of visual acuity scoring, we tested 32 patients with macular pathology and 38 age-matched, normal subjects with six repetitions of the ETDRS chart. We then scored the test using three methods (line assignment, letter-by-letter, and probit) to determine which *method* of scoring has the lowest variability.

METHODS

Subjects

Thirty-two subjects with ophthalmoscopic evidence of macular pathology evaluated in the retina clinic at University of Iowa, Department of Ophthalmology gave informed consent prior to inclusion in the study. The tenets of the 1964 Declaration of Helsinki were followed. Sixteen patients (50%) had maculopathy due to diabetes mellitus and 10 (31%) had age-related macular degeneration or presumed ocular histoplasmosis, with or without choroidal neovascularisation. The remaining patients had a variety of diagnoses. To be included, subjects had to read the top line of the ETDRS chart correctly at 4 m (20/200).

Thirty-eight normal subjects, matched for age within 5 years with an equal distribution within an age group (see below), were recruited from clinic staff and relatives of ophthalmology clinic patients. Normals had no history of eye trauma, surgery, glaucoma, or any ophthalmic diagnosis other than refractive error. In both groups at least 10 subjects were studied in each of three 20-year age categories (20–39; 40–59; 60–79+).

Equipment

Visual acuity was tested using the Lighthouse ETDRS charts mounted on a light box, with the subject seated 4 m from the chart. Room lighting was at office levels (about 50 foot-candles).

Visual Acuity Testing

All subjects were refracted using the method of the ETDRS study.[21] This study used the 'ETDRS' chart, which has 14 lines with five letters per line. One eye was tested for each subject. Subjects were then asked to read either the ETDRS chart 1 or 2. They were requested to start at the top left and work down to the smallest letters at the bottom right. They were instructed to give one reading for each letter, guessing if they were not sure. They were encouraged to guess even when they reached lines where they could not see any of the letters clearly. There

was no time limit. They were then asked to read the other chart (2 or 1, whichever they had not started with) in a similar fashion. Subjects were allowed to begin several rows from the top on subsequent charts if they easily read the top lines on the first reading, but always began reading at least three lines above the line where they first missed letters.

For the third chart reading, subjects were asked to read the first chart again, reading by row from right to left to decrease the chance of memorisation (although this decrease has not been demonstrated). The series of three chart readings was then repeated. All six readings were done at one sitting. Subject motivation was graded on a scale of 1–10, with 10 being most motivated to do well. This grading was a subjective impression based on the subject's effort to be correct and attitude towards guessing near the limit of their acuity.

Scoring and Statistics

Three scoring methods were used and the results from each method were converted to the log minimum angle of resolution (log MAR) values:

1. *Line assignment method.* The Snellen equivalent fraction was determined for each chart reading using the last line where three out of five letters were read correctly. The result was recorded in decimal form and then converted to the log MAR value (e.g. 20/40 = 0.5 = 0.3.)

2. *Letter-by-letter or ETDRS method.* This was scored by totalling the number of letters on one chart that were answered correctly. The transformation to log MAR units was done using the formula $(1.1 - 0.2\, T_c)$ where T_c is the total number of letters read correctly on the chart.

3. *Probit analysis.*[6,17] This was used to determine a 50% frequency of seeing threshold MAR with software provided by L. Frisén. The base 10 logarithm of this value was used to convert the values to log MAR units.

Visual acuity charts all truncate visual acuity (a 'ceiling effect') at the small visual angle end of the chart. Although this affected only a small number of our subjects, it was a factor in our probit analysis. We did not have a large tail at the small visual angle end of the visual acuity measurements to form the S-shaped cumulative gaussian function. Therefore, for these cases we made the assumption that the letters on lines corresponding to 20/8 and 20/6 would be missed by our subjects. To test this assumption, four subjects who could correctly identify at least one letter on the 20/10 line at 4 m were asked to read the 20/10 line from 5.1 m and 6.3 m. None was able to identify any of the letters better than that of chance guessing.

Data were imported into Systat (Systat Intelligent Software, Evanston, IL) for further analysis. Differ-
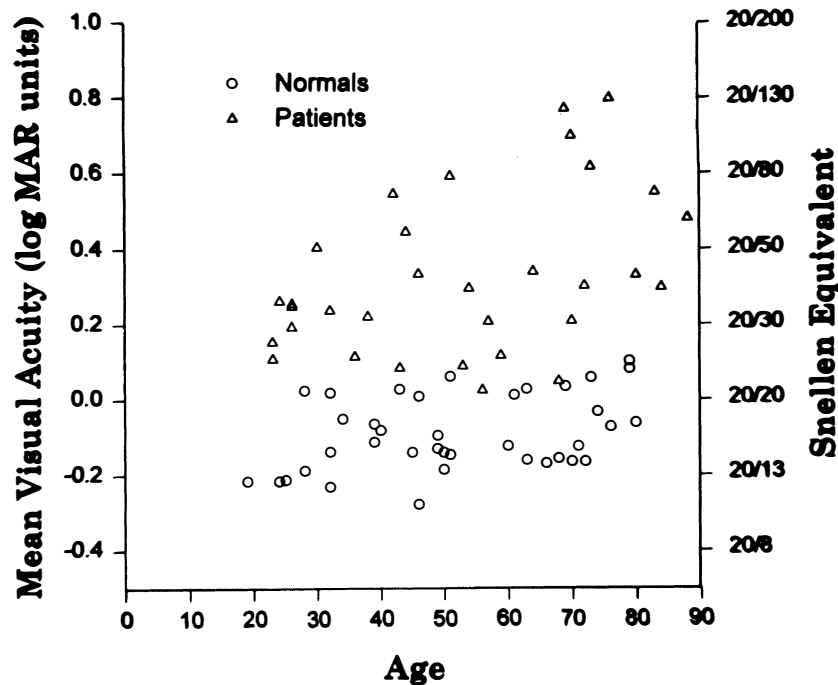
Fig. 1. *Mean visual acuities using the letter-by-letter scoring method plotted by age in normals and pa...nts. One line of visual acuity is equal to 0.1 log MAR unit.*

ences in variability between the two groups (patients and normals) as well as differences within groups between the different scoring methods were evaluated using an ANOVA on the standard deviations of the six repetitions. Group and method were modelled as factors and age and motivation as covariates. To determine whether there were differences in the visual acuity values among the groups a repeated measures ANOVA was performed on the log MAR transformed visual acuity values. Tukey post-hoc multiple comparisons were performed if significant differences were present. To determine the relationship between variability and age, the standard deviations of the letter-by-letter method were linearly regressed against age. Differences between groups were considered significant if $p<0.05$.

RESULTS

The visual acuities of the normal subjects ranged from 20/10 to 20/20; those of the patients ranged from 20/15 to 20/200 (Fig. 1). To be sure the difference in methods did not change the visual acuities of the groups, a repeated measures ANOVA

was performed on the log MAR transformed visual acuity values. As expected, there were differences between the groups (patients with worse visual acuities, $p<0.0001$) but there were no differences in the visual acuities among the methods.

The data for each scoring method by group are given in Table I and Fig. 2. Note the higher variability as shown by the mean of the standard deviations with the line assignment scoring method compared with the other two methods ($p<0.0001$; Tables II and III, Fig. 2). Post-hoc comparison of the probit and letter-by-letter methods showed no differences between these two methods. There were no significant differences between the standard deviations of the normals and the patients and there was no significant group-method interaction. With the letter-by-letter scoring method, the mean number of letters correct was 59.5 ± 1.7 for normals and 38.8 ± 1.9 for patients.

Age was a significant covariate of the retest variability ($p<0.001$). Fig. 3 shows the results of the linear regression analysis of the standard deviations against age for normals and patients for the ETDRS

Table I. Means and standard deviations for the three methods of scoring the six repeated visual acuity measures

Group	Line assignment	ETDRS	Probit
Normals			
Mean	−0.111 ± 0.11	−0.091 ± 0.10	0.145 ± 0.10
Standard deviation	0.049 ± 0.02	0.034 ± 0.01	0.035 ± 0.01
Patients			
Mean	0.302 ± 0.21	0.323 ± 0.21	0.275 ± 0.21
Standard deviation	0.049 ± 0.02	0.038 ± 0.01	0.038 ± 0.02

ETDRS refers to the letter-by-letter scoring method.
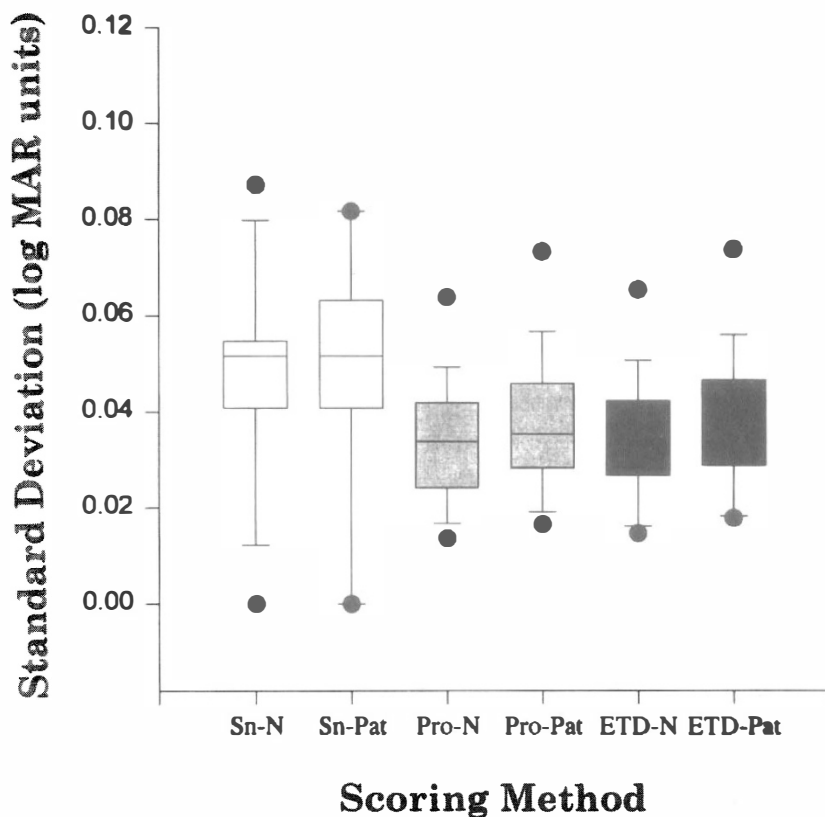
Fig. 2. *Tukey box plot of the average standard deviations for the six repetitions for each scoring method. Each box represents the 75th, 50th and 25th percentiles. Error bars show 90th and 10th percentiles, and filled circles the 95th and 5th percentiles. Sn, Snellen equivalent; Pro, probit analysis; ETD, ETDRS or lettery-by-letter scoring method; N, normals; Pat, patients.*

scored data (which was representative of the other methods). The relationship was: $E_{sd} = 0.025 + (0.0002 \times Age)$, where E_{sd} is the standard deviation using the letter-by-letter (ETDRS) scoring method.

Subject motivation was generally very good. Motivation on the 1–10 scale for the normals was 9.2 ± 0.88; the motivation for the patients was 8.8 ± 0.93. Motivation was not a statistically significant covariate.

DISCUSSION

The method by which visual acuity is scored may be as important a factor as the test's lighting or optotypes used.[2,4,6] This study is the first to compare three different scoring methods. With regard to comparison of two of the methods – letter-by-letter with line assignment – our results are similar to others: the variability of the line assignment method is significantly higher.[2,4,6,20] We also found the probit method had a similar variability to the letter-by-letter

method. Although the line assignment method is easier to score, its variability is considerably higher.

Frisén and Frisén[6] suggested that probit analysis to determine a 50% frequency of seeing value would be the best way to score visual acuity. Our results confirm this is a method with relatively low variability. However, it may be less practical in a clinical setting than the letter-by-letter method. To use 10 or more letters per line, as they suggest, might improve the variability of this method. However, the added time for more letters and need for software to perform the probit least squares fit also limits this method for use in a clinical setting.

Ophthalmologists have always accepted some variability with visual acuity testing; it has been standard clinical practice to consider a change of at least two lines on the Snellen chart as a cutoff value for clinically meaningful change in visual acuity.[3] Many factors contribute to this high value, including differences in charts, their illumination and test

Table 2. Post-hoc Tukey pairwise comparison probabilities between the methods combining the groups

Post-hoc comparison	Mean difference in standard deviations	p value
Line assignment with ETDRS	0.013	<0.0001
Line assignment with probit	0.012	<0.0001
ETDRS with probit	0.001	NS

NS, not significant; ETDRS refers to the letter-by-letter scoring method.

Table III. Repeated measures analysis of variance of the standard deviation scores of the three scoring methods in the two groups (patients and normals)

Source	Sum-of-squares	d.f.	Mean-square	F-ratio	p
Group	0.0004	1	0.0005	1.38	0.242
Method	0.007	2	0.004	11.04	<0.0001
Group × method	0.00008	2	0.00004	0.13	0.880
Age	0.004	1	0.004	12.34	0.001
Motivation	0.0004	1	0.0004	1.34	0.248
Error	0.067	202	0.0003		

instructions between examiners. Elliot and Sheridan[20] found 95% confidence limits for retest variability to be 3.5 letters for letter-by-letter and 10.5 letters for line assignment. Bailey and Lovie[4] concluded for a letter-by-letter scoring method the 95% confidence interval was five letters or one line. For the line assignment method the 95% confidence interval was 10 letters or two rows. However, Arditi and Cagenello,[2] using five trained observers, concluded 4.5 letters was the 95% confidence limit for change using letter-by-letter scoring and seven letters for the line assignment method. For comparison, if we calculate a 95% confidence limit by multiplying our average standard deviations by 1.96, our results are similar at 3.5 letters for letter-by-letter but not as high as these studies (five letters) for the line assignment method. Our values represent the intratest variability while other studies include the intertest variability. This is likely the reason our values are lower.

Arditi and Cagenello state 'poor best corrected visual acuities are associated with significant visual impairment which is widely believed to be associated with increased variability in visual acuity measurements'. Although variability is higher in conventional automated perimetry in patients with optic nerve damage than normals,[22,23] this is not true for automated perimetry that determines threshold by stimuli that vary size instead of luminance.[24,25] Since visual acuity testing uses a similar method – finding threshold by varying size – this may explain why we did not find differences in variability in patients compared with normals.

We also found an increased variability with age. Although it is well known that visual acuity decreases with age,[6,26] this is not the explanation for the overall variability in our study, since the differences in the methods were significant ($p<0.006$) even when age was entered into the model as a covariate.
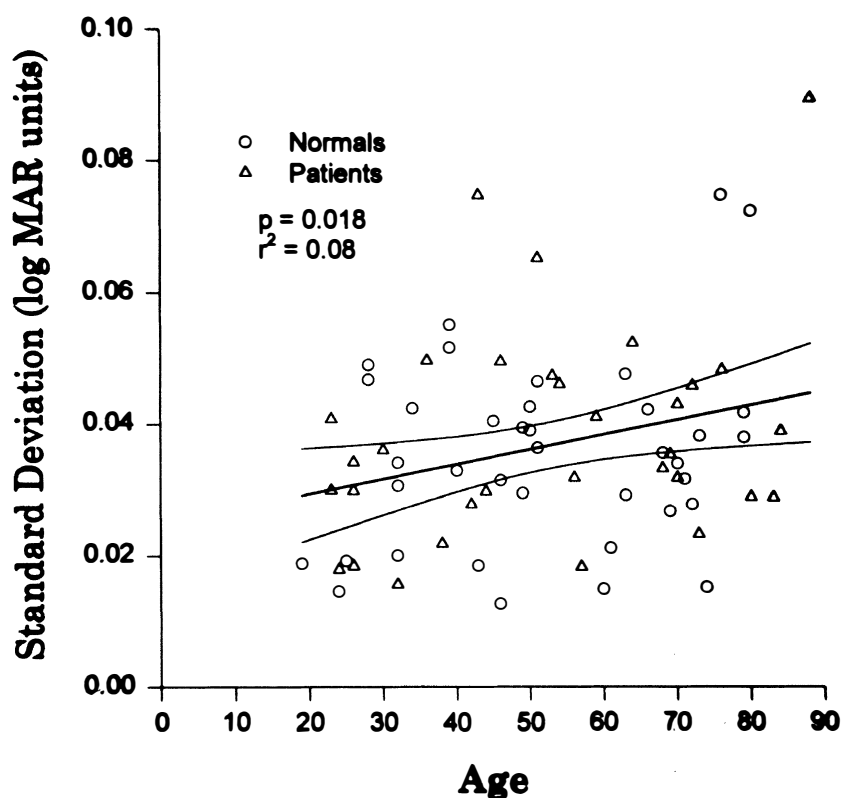
Fig. 3. *Average standard deviation of the six repetitions using the ETDRS or letter-by-letter scoring method plotted by age in normals and patients. The regression line and its 95% confidence interval are shown.*

Guessing can also alter the results of visual acuity testing. Arditi and Cagenello[2] point out that when guessing is allowed, the size of the pool of possible answers can have a significant effect on the results. For the Landolt C, 'tumbling' or illiterate E, or HOTV charts the chance of guessing correctly is 25%. With letter charts, it is 1/26 (in theory, but after reading several lines subjects may realise there is a limited pool of letters) although for the ETDRS charts since only 10 letters are used, it is 1/10 (if subjects realise the pool of letters is only 10). Likewise, assuming 1/10 leads to error if the subject is guessing based on 26 letter options.

Because subjects often stop when letters are difficult to see, yet can sometimes read one or more additional lines without errors when encouraged, they should be urged to guess in an effort to get the best possible visual acuity. We limited inter- and intra-examiner variation in motivating subjects to guess by using the same examiner for all test subjects and attempting to instruct and encourage equally. Use of one examiner probably contributes to our low variability compared with other studies.

Guessing probably has different effects on the three scoring methods. A one letter difference would have little effect on letter-by-letter scoring but could make the difference in being credited with a whole line better visual acuity with line-by-line scoring (Snellen). Guessing would probably have an intermediate effect on the probit scoring method.

The effect of chart memorisation can also affect variability. We tried to limit this possibility by reading two different but methodologically identical charts[5] and also by reading one of them backwards. This allowed six consecutive chart readings without using a chart more than twice. Even so, re-use of charts may have contributed to our low variability.

In conclusion, the test–retest variability of the ETDRS chart varies with the method of scoring the test and is high for the line assignment method. For monitoring a disease process or treatment regimen, using the letter-by-letter scoring or probit analysis will have a lower variability. Using the letter-by-letter scoring method our results, along with the results of others, suggest a change of five letters or more on an ETDRS chart is significant. Use of this method of scoring rather than line assignment should allow earlier detection of change in visual acuity.

REFERENCES

1. Snellen H. Probebuchstaben zur Bestimmung der Sehscharfe. Utrecht: PW van de Weijer, 1862.
2. Arditi A, Cagenello R. On the statistical reliability of letter-chart visual acuity measurements. Invest Ophthalmol Vis Sci 1993;34:120–9.
3. Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effect of scaling. Invest Ophthalmol Vis Sci 1991;32:422–32.
4. Bailey IL, Lovie JE. New design principles for visual acuity letter charts. Am J Optom Physiol Opt 1976; 53:740–5.
5. Ferris FL, III, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. Am J Ophthalmol 1982;94:91–6.
6. Frisén L, Frisén M. How good is normal visual acuity? A study of letter acuity thresholds as a function of age. Graefes Arch Klin Exp Ophthalmol 1981;215:149–57.
7. Hedin A, Olsson K. Letter legibility and the construction of a new visual acuity chart. Ophthalmologica 1984;189:147–56.
8. Klein R, Klein BE, Moss SE, DeMets D. Interobserver variation in refraction and visual acuity measurement using a standardized protocol. Ophthalmology 1983;90:1357–9.
9. Lovie-Kitchin JE. Validity and reliability of visual acuity measurements. Ophthalmic Physiol Opt 1988;8: 363–70.
10. Anonymous. Recommended standard procedures for the clinical measurement and specification of visual acuity. Report of working group 39. Committee on vision. Assembly of Behavioral and Social Sciences, National Research Council, National Academy of Sciences, Washington, DC. Adv Ophthalmol 1980;41: 103–48.
11. Parr JC. Clinical assessment of visual acuity. Trans Ophthalmol Soc NZ 1981;33:157–67.
12. Sakuma JT, da Silva JA, Velasco e Cruz A. Measurement of visual acuity with logarithmic charts: comparison between methods of threshold estimation. Optom Vis Sci 1991;68:666–9.
13. Sloan LL, Rowland WM, Altman A. Comparison of three types of test target for the measurement of visual acuity. Q Rev Ophthalmol 1952;8:4–15.
14. Sloan LL. Measurement of visual acuity: a critical review. Arch Ophthalmol 1951;45:704–25.
15. Sloan LL. New test charts for the measurement of visual acuity at far and near distances. Am J Ophthalmol 1959;48:807–13.
16. Strong G, Woo GC. A distance visual acuity chart incorporating some new design features. Arch Ophthalmol 1985;103:44–6.
17. Tinning S, Bentzon MW. A new method for exact measurements of visual acuity. Determination of threshold curves for the resolving power of the eye by computerized curve fitting. Acta Ophthalmol (Copenh) 1986;64:180-6.
18. Wong D, Kaye SB. Chart for visual acuity screening. Br J Ophthalmol 1989;73:457–60.
19. Gibson RA, Sanderson HF. Observer variation in ophthalmology. Br J Ophthalmol 1980;64:457–60.
20. Elliott DB, Sheridan M. The use of accurate visual acuity measurements in clinical anti-cataract formulation trials. Ophthalmic Physiol Opt 1988;8:397–401.
21. Early Treatment Diabetic Retinopathy Study. Manual of operations. Chapter 12, 1985.
22. Heijl A, Lindgren A, Lindgren G. Test–retest variability in glaucomatous visual fields. Am J Ophthalmol 1989;108:130–5.

23. Werner EB, Petrig B, Krupin T, Bishop KI. Variability of automated visual fields in clinically stable glaucoma patients. Invest Ophthalmol Vis Sci 1989;30: 1083–9.

24. Chauhan BC, House PH. Intratest variability in conventional and high-pass resolution perimetry. Ophthalmology 1991;98:79–83.

25. Wall M, Lefante J, Conway M. Variability of high-pass resolution perimetry in normals and patients with idiopathic intracranial hypertension. Invest Ophthalmol Vis Sci 1991;32:3091–5.

26. Reeves BC, Hill AR, Aspinall PA. The clinical significance of change. Ophthalmic Physiol Opt 1987; 7:441–6.