

A comparison of stereoscopic and monoscopic evaluation of optic disc topography using a digital optic disc stereo camera

Ben Parkin, Garry Shuttleworth, Mark Costen, Charles Davison

Abstract

Aims—To compare stereophotographic and monophotographic optic disc assessments made using a digital optic disc stereo camera.

Methods—Stereo digital optic disc photographs of 150 selected patients who had presented to a glaucoma clinic were assessed by two masked observers on separate occasions using (1) the stereophotographs and a stereoviewer, (2) a single image from the same stereopair. Results were analysed for both right and left eyes separately. 95% tolerance limits for change (TC) and intraclass correlation coefficients (ICC) were calculated and a multivariate analysis using a general linear model for repeated measures was performed.

Results—A total of 201 optic disc images of 150 patients (84 females, 108 left eyes) were analysed. Mean age of patients was 64 years. The results for right eyes are as follows (similar results were obtained for left eyes). Intraobserver (stereoscopic compared to monoscopic) measurements of: horizontal cup:disc ratios (CDR), ICC = 0.5995 and 0.7269, TC = 34% and 27%; vertical CDR, ICC = 0.8298 and 0.817, TC = 25% and 27%; area CDR, ICC = 0.7757 and 0.8259, TC = 28% and 25%; circumference CDR, ICC = 0.7618 and 0.8103, TC = 28% and 25%. Interobserver measurements of: horizontal CDR, ICC stereoscopic (SS) = 0.7287; monoscopic (MS) = 0.5030; TC SS = 30%; MS = 32%; vertical CDR, ICC SS = 0.8439; MS = 0.7106; TC SS = 25%; MS = 31%; area CDR, ICC SS = 0.8392; MS = 0.6276; TC SS = 26%; MS = 32%; circumference CDR, ICC SS = 0.8433; MS = 0.6438, TC SS = 26%; MS = 31%. Systematic bias between observers and between methods was within acceptable limits.

Conclusions—This study using a digital stereo camera indicates that there may be little benefit of stereoscopic imaging over monoscopic imaging despite demonstrating small but inconsistent differences between both observers and methods.

(Br J Ophthalmol 2001;85:1347-1351)

The detection of early or progressive glaucomatous optic disc changes is essential in the optimum management of glaucoma patients. Despite the plethora of new technologies the accepted "gold standard" for optic disc assessment remains stereoscopic optic disc

photography.¹⁻⁴ However, measurements rely on a subjective evaluation by the clinician and are subject to poor interobserver agreement.⁵⁻¹⁰

The Discam optic disc camera (Marcher Enterprises Ltd, Hereford) obtains digital monochromatic sequential stereophotographic images of the optic disc and the allied software allows the computation of a number of optic disc parameters. In a previous study, the intraobserver and interobserver repeatability of optic disc measurements was shown to compare favourably with ophthalmoscopic and stereophotographic methods of assessment of the optic disc.¹¹ However, stereoscopic versus monoscopic assessments of the images were not performed.

The aims of this study were to determine whether there were significant differences in intraobserver and interobserver repeatability of the computed optic disc parameters using monophotographic and stereophotographic images to determine the margins of the optic disc and optic cup.

Materials and methods

Stereophotographic digital images of patients presenting to a glaucoma clinic were acquired by two experienced operators following pharmacological mydriasis. The detailed method has been previously described.¹¹ The quality of the images were assessed by observer 2 (MC), on the basis of contrast, clarity, stereoscopic appearance, illumination, and excessive movement between the image pair. As this study did not aim to determine the clinical usefulness of the Discam, but purely to compare two methods of assessment, only high quality images were used for analysis in order to maximise any differences. Exclusion criteria were poor illumination of the disc, poor quality image, lens opacity, poor stereoimage (large shift on the screen between images of stereo pair) and optic discs of anomalous configuration (those which were tilted, colobomatous, or myopic). A total of 201 images (from a total of 300) were considered suitable for inclusion in the study. Two masked observers (BP and MC), ophthalmologists experienced in the clinical use of the Discam, assessed the selected disc images on two separate occasions within 1 month, first making a stereoscopic assessment of the discs and on another occasion a monoscopic assessment. The stereoscopic assessment included an initial "flickering" of the image pairs which were subsequently displayed simultaneously on a high resolution computer screen for detailed examination using a hand held stereoviewer

Southampton Eye Unit, Southampton General Hospital, Tremona Road, Southampton SO16 6YD, UK
B Parkin

Bristol Eye Hospital, Lower Maudlin Street, Bristol BS1 2LX, UK
G Shuttleworth

Bournemouth Eye Unit, Royal Bournemouth Hospital, Castle Lane East, Bournemouth BH7 7DW, UK
M Costen
C Davison

Correspondence to:
Mr Ben Parkin
benparkin@hotmail.com

Accepted for publication
25 April 2001

Table 1 Stereoscopic-monoscopic intermethod (intraobserver) variability statistics for right eyes

	Observer 1							Observer 2						
	Mean difference between measures	SD of difference	t test p value	Range of CDR	95% TC (% of range)	ICC	>0.2 CDR difference	Mean difference between measures	SD of difference	t test p value	Range of CDR	95% TC (% of range)	ICC	>0.2 CDR difference
H-CDR	-0.020	0.068	0.006	0.39	34	0.60	1 [1%]	0.014	0.063	0.040	0.46	27	0.73	1 [1%]
V-CDR	-0.029	0.058	0.000	0.46	25	0.83	1 [1%]	0.011	0.065	0.105	0.47	27	0.82	1 [1%]
A-CDR	-0.035	0.077	0.000	0.53	28	0.78	NA	0.018	0.071	0.015	0.57	25	0.83	NA
P-CDR	-0.026	0.055	0.000	0.38	28	0.76	NA	0.011	0.054	0.056	0.42	25	0.81	NA

TC = tolerance for change; V-CDR = vertical cup-disc ratio; H-CDR = horizontal cup-disc ratio; A-CDR = area cup-disc ratio; P-CDR = perimeter cup-disc ratio; ICC = intraclass correlation coefficient.

(Screen-Vu Stereoscope, USA). The inner edge of the scleral rim (the optic disc) and the point at which the neuroretina first deviated posteriorly (the optic cup) were traced out on the screen using a computer mouse. On a separate occasion, the optic disc and optic cup were drawn without the use of either the stereoviewer or previous "flickering" of the image pairs. The computer software (Discam version 3.0.0.1) then calculated the vertical and horizontal cup:disc ratios (CDR), together with cup area to disc area and cup circumference to disc circumference ratios.

A statistical analysis of the variability of both intraobserver and interobserver measures was then performed. The images acquired from right and left eyes were analysed separately to avoid bias due to intereye correlation. The distribution of the data was examined to demonstrate normality. Scatter plots of measurement differences plotted against measurement averages (Bland-Altman plots¹²) were produced. To analyse measurement variability, the 95% tolerance limits for change (TC) and the intraclass correlation coefficient (ICC) were calculated. The intraclass correlation coefficient (ICC) coefficient is equivalent to a quadratic weighted kappa statistic, which is a chance corrected measure of agreement, weighting degrees of discrepancies according to the square of the difference between the (paired)

measurements.¹³ The ICC only indicates perfect agreement if the two assessments are numerically equal—that is, if a plot of the two measurements has zero intercept and a slope of unity. However, the (crude) ICC is affected specifically by any systematic differences between the observations within the pairs. In other words, even perfect "agreement" in the context of such systematic differences will result in an ICC less than 1. In addition, a general linear model for repeated measures analysis was performed and disagreements of ≥ 0.2 units for horizontal and vertical CDRs were also determined.

Results

A total of 201 images (108 left and 93 right) of the optic discs of 150 patients (66 males and 84 females), mean age 64 (median 66; range 22–87), of normal, suspicious, and glaucomatous eyes were examined. All images were acquired between 9 September 1998 and 16 June 2000.

Intraobserver (stereoscopic/monoscopic = intermethod) and interobserver variability results for right and left eyes are shown in Tables 1–4. The distributions of all data and the difference between observations approximated to the normal distribution and scatter plots of repeated measurements showed close agreement.

Table 2 Interobserver variability statistics for right eyes

	Stereoscopic interobserver measures							Monoscopic interobserver measures						
	Mean difference between measures	SD of difference	t test p value	Range of CDR	95% TC (% of range)	ICC	>0.2 CDR difference	Mean difference between measures	SD of difference	t test p value	Range of CDR	95% TC (% of range)	ICC	>0.2 CDR difference
H-CDR	0.010	0.062	0.116	0.41	30	0.73	0	0.044	0.074	0.000	0.46	32	0.50	1 [1%]
V-CDR	-0.003	0.061	0.311	0.47	25	0.84	1 [1%]	0.033	0.077	0.000	0.5	30	0.71	3 [3%]
A-CDR	0.004	0.072	0.628	0.55	26	0.84	NA	0.057	0.092	0.000	0.57	32	0.63	NA
P-CDR	0.003	0.051	0.597	0.39	26	0.84	NA	0.034	0.066	0.000	0.42	31	0.64	NA

TC = tolerance for change; V-CDR = vertical cup-disc ratio; H-CDR = horizontal cup-disc ratio; A-CDR = area cup-disc ratio; P-CDR = perimeter cup-disc ratio; ICC = intraclass correlation coefficient.

Table 3 Stereoscopic-monoscopic intermethod (intraobserver) variability statistics for left eyes

	Observer 1							Observer 2						
	Mean difference between measures	SD of difference	t test p value	Range of CDR	95% TC (% of range)	ICC	>0.2 CDR difference	Mean difference between measures	SD of difference	t test p value	Range of CDR	95% TC (% of range)	ICC	>0.2 CDR difference
H-CDR	-0.022	0.057	0.000	0.46	24	0.75	1 [1%]	-0.019	0.082	0.022	0.58	28	0.71	3 [3%]
V-CDR	-0.018	0.060	0.002	0.59	20	0.87	1 [1%]	-0.015	0.064	0.014	0.57	22	0.86	1 [1%]
A-CDR	-0.030	0.063	0.000	0.62	20	0.85	NA	-0.022	0.081	0.006	0.68	23	0.82	NA
P-CDR	-0.021	0.051	0.000	0.48	21	0.82	NA	-0.014	0.061	0.023	0.56	21	0.83	NA

TC = tolerance for change; V-CDR = vertical cup-disc ratio; H-CDR = horizontal cup-disc ratio; A-CDR = area cup-disc ratio; P-CDR = perimeter cup-disc ratio; ICC = intraclass correlation coefficient.

Table 4 Interobserver variability statistics for left eyes

	Stereoscopic interobserver measures							Monoscopic interobserver measures						
	Mean difference between measures	SD of difference	t test p value	Range of CDR	95% TC (% of range)	ICC	>0.2 CDR difference	Mean difference between measures	SD of difference	t test p value	Range of CDR	95% TC (% of range)	ICC	>0.2 CDR difference
H-CDR	0.043	0.088	0.000	0.58	30	0.58	5 [5%]	0.047	0.076	0.000	0.51	29	0.60	5 [5%]
V-CDR	-0.0003	0.074	0.969	0.61	24	0.84	1 [1%]	0.0023	0.055	0.662	0.56	19	0.88	0
A-CDR	0.023	0.087	0.008	0.67	26	0.79	NA	0.031	0.078	0.000	0.67	23	0.80	NA
P-CDR	0.010	0.066	0.108	0.55	23	0.80	NA	0.018	0.053	0.001	0.5	21	0.82	NA

TC = tolerance for change; V-CDR = vertical cup-disc ratio; H-CDR = horizontal cup-disc ratio; A-CDR = area cup-disc ratio; P-CDR = perimeter cup-disc ratio; ICC = intraclass correlation coefficient.

INTRAOBSERVER (STEREOSCOPIC COMPARED TO MONOSCOPIC) ANALYSES

Bland-Altman scatter plots showed that the magnitude of intraobserver differences were unrelated to the size of the measurement. However, for observer 2 there is a suggestion that H-CDR and V-CDR differences between stereoscopic and monoscopic assessments (SM) were smaller in more severely cupped discs (indicated by a funnel shape to the scatter plot in Fig 1). There was also a suggestion in

left eyes that V-CDR monoscopic measurements were larger than stereoscopic measurements for less cupped discs (that is, the calculated differences tended to be negative for smaller CDR as shown in Fig 1).

Mean monoscopic assessments made by observer 1 were slightly larger than the corresponding stereoscopic assessments, reaching statistical significance for all measurement parameters (H-CDR, V-CDR, A-CDR, and P-CDR) for both right and left eyes. Observer 2 also made slightly larger mean monoscopic assessments of all measurement parameters for left eyes, which again reached statistical significance, but for right eyes observer 2 made slightly smaller mean monoscopic assessments. However, overall the systematic errors were very small and best illustrated graphically (Fig 2).

For both H-CDR and V-CDR, agreements between monoscopic and stereoscopic assessments were within 0.2 CDR units in at least 97% of eyes. ICC showed very significant agreements between all assessments but was lowest for H-CDR assessments. Tolerance limits for change, which take into account the range of measurements, varied from 20% to 34% and were generally smaller for left eyes.

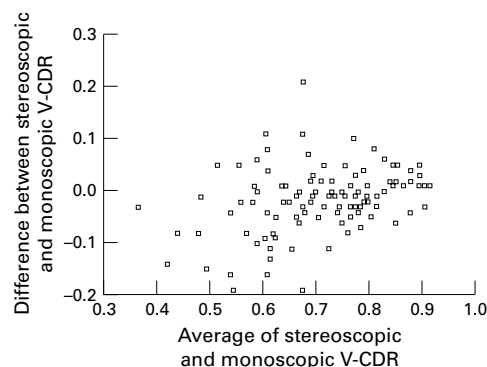


Figure 1 Observer 2 stereoscopic versus monoscopic assessment for left eyes, vertical cup:disc ratios (V-CDR).

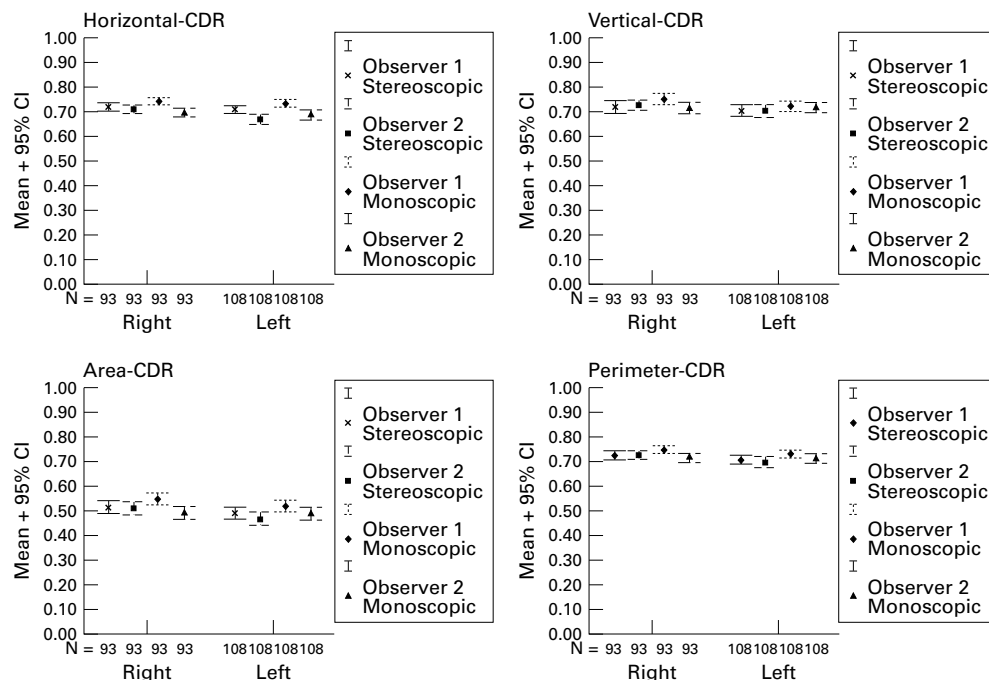


Figure 2 Mean and 95% confidence intervals for each of the optic disc variables measured, with right and left eyes analysed separately.

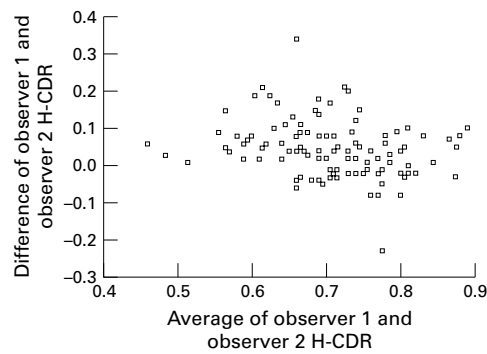


Figure 3 Observer 1 versus 2, monoscopic assessment, horizontal cup:disc ratios (H-CDR) for left eyes.

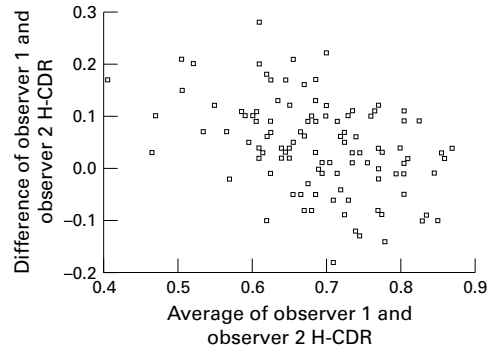


Figure 4 Observer 1 versus 2, stereoscopic assessment, horizontal cup:disc ratios (H-CDR) for left eyes.

The smallest tolerance limits overall were for the V-CDR and the widest for H-CDR.

INTEROBSERVER ANALYSES

Bland-Altman scatter plots showed that the magnitude of interobserver differences was again essentially unrelated to the size of the measurement. However, for observer 1 the H-CDR monoscopic assessments were generally larger than those made by observer 2 (that is, calculated differences tended to be greater than zero as shown in Fig 3). In addition, observer 1 H-CDR stereoscopic assessments were generally larger than those made by observer 2, particularly in the less severely cupped discs (indicated by a downward slope to the scatter plot in Fig 4).

Mean stereoscopic assessment differences were generally smaller than the mean monoscopic assessment differences and also the mean intraobserver monoscopic/stereoscopic differences. However, the tolerance limits for change were broadly similar. The mean stereoscopic differences between observers only reached statistical significance for left eye

H-CDR and A-CDR, whereas mean monoscopic differences reached statistical significance for all measurement parameters except left eye V-CDR. Overall, observer 1 mean assessments were larger than those of observer 2; however, these only reached statistical significance consistently for monoscopic assessments. Once again all of the systematic errors were very small (Fig 2).

For both H-CDR and V-CDR, agreements between monoscopic and stereoscopic assessments were within 0.2 CDR units in at least 95% of eyes. ICC showed very significant agreements between all assessments but was again generally lowest for H-CDR assessments. Tolerance limits for change varied from 19% to 32% and were similar for both right and left eyes. They were, however, generally wider than for the intraobserver monoscopic/stereoscopic comparisons. Overall the narrowest tolerance limits were again those of V-CDR and the widest those of H-CDR.

MULTIVARIATE ANALYSIS

The general linear model for repeated measures analysis revealed that for both right and left eyes, the effect of the observer on the measurements is statistically significant with the exception of V-CDR for left eyes (Table 5). The analysis also shows that for left but not right eyes, the effect of method of assessment on repeatability (stereoscopic *v* monoscopic) is statistically significant. It is not surprising therefore that observer/method interactions are not apparent in left eyes whereas they are significant for right eyes. This implies that for right eyes at least there is some inconsistency in the errors made by each observer when using the different methods.

Discussion

It is of paramount importance that any topographical measurement of the optic disc is highly repeatable in order to facilitate both the reliable detection of the abnormalities and the temporal changes that occur in chronic optic neuropathies such as glaucoma.

In a previous study the intraobserver and interobserver repeatability of stereoscopic optic disc assessments using the Discam optic disc camera was shown to be as reliable as conventional stereophotographic assessments¹¹ and better than clinical optic disc assessment.¹⁰ For this reason, it was not felt necessary to include intraobserver repeatability measurements in this study. Although stereoscopic methods of photographic optic disc assessment are considered to be better than monoscopic methods³ the effect of stereoscopic versus monoscopic delineation of the optic disc and cup margins was not addressed by the investigators. Whereas stereoscopic assessments require a stereoviewer, careful alignment, stereopsis, and concentration the monoscopic assessments require only the latter. This study was therefore designed to determine the benefits, if any, of stereoscopic optic nerve head assessment using the Discam camera.

The results of the study reveal a tendency for monoscopic assessments to yield slightly larger

Table 5 General linear model for repeated measures analysis

	<i>p</i> Values			
	H-CDR	V-CDR	A-CDR	P-CDR
Right eyes (n=93)				
Observer	0.000	0.024	0.000	0.002
Method	0.566	0.065	0.191	0.091
Observer/method interaction	0.000	0.000	0.000	0.000
Left eyes (n=108)				
Observer	0.000	0.838	0.000	0.003
Method	0.000	0.000	0.000	0.000
Observer/method interaction	0.696	0.731	0.406	0.284

optic disc parameters than stereoscopic assessments. This effect was more marked for observer 1 whose monoscopic assessments were also slightly larger than those made by observer 2. Intraobserver versus interobserver repeatability comparisons usually show that the former is more reliable than the latter and a similar trend appeared in this study. Although any comparison is not strictly valid, as methods of image assessment were compared with interobserver variation the results suggest that the effect of the observer may be more important than the effect of the method. Interestingly, and as expected, stereoscopic interobserver differences appeared smaller than the corresponding monoscopic interobserver differences. However, the effects were very small.

Agreements were generally extremely good but analyses revealed that the V-CDR was consistently more repeatable than H-CDR having narrow tolerance limits for change, larger ICCs and fewer CDR differences > 0.2 units. However, ICCs were slightly but consistently lower than those previously published even for the directly comparable interobserver stereoscopic assessment.¹¹ Tolerance limits for change were similarly slightly wider; however, the standard deviations of the differences were of very similar magnitude. These wider tolerance limits for change (which is a function of the measurement interval) may in part be attributed to the narrower range of CDR parameters in the current study and serves to highlight a potential weakness of expressing data in this way, as outliers have disproportionate influence. The tendency for V-CDR measurement differences to be greater towards the lower end of the measurement range as reported in the previous Discam paper¹¹ was not evident in this study possibly because none of the optic discs imaged had very small optic cups.

Overall, the variations observed were very small as illustrated clearly in Figure 2. Although the assessment of the edge of the optic cup was subjectively facilitated by a stereoscopic view, it is doubtful whether the statistical systematic errors detected, which may reflect the large sample size, will be of clinical significance, particularly when they are compared to the tolerance limits for change. Given that the findings of this study relate to a group of selected optic disc images of good stereoquality, it is probable that both intermethod (intraobserver) and interobserver differences would be less marked or not apparent in unselected cases. It is not possible to extrapolate the findings to a "normal" population of optic discs outside the glaucoma clinic. However, the Discam is most likely to be used

clinically for the assessment of the optic discs of a similar population to that examined in this study.

It would, therefore, appear that the Discam images can be analysed on a high resolution screen nearly as effectively monoscopically as stereoscopically which would run contrary to intuition. There are clearly practical advantages to monophotographic Discam image analysis, which relate to the speed and convenience of assessment (nothing more than a standard computer and mouse are required). We stress however that our data are derived from assessment and analysis of screen images and cannot be extrapolated to other media—for example, printed Discam images.

In conclusion, the analysis of intermethod (intraobserver) and interobserver variability revealed systematic biases between observers, between the methods used to determine the margins of the optic disc and cup and also demonstrated an observer/method inconsistencies. However the effects were very small and are probably not clinically significant. Although there may be a subjective benefit in the use of stereoscopic rather than monoscopic images in the measurement of cup to disc ratios, the study did not demonstrate a clinically significant benefit.

- Rosenthal AR, Kottler MS, Donaldson DD, *et al.* Comparative reproducibility of the digital photogrammetric procedure utilizing three methods of stereophotography. *Invest Ophthalmol Vis Sci* 1977;16:54–60.
- Krohn MA, Keltner JL, Johnson CA. Comparison of photographic techniques and films used in stereophotogrammetry of the optic disk. *Am J Ophthalmol* 1979;88:859–63.
- Varma R, Steinmann WC, Scott IU. Expert agreement in evaluating the optic disc for glaucoma. *Ophthalmology* 1992;99:215–21.
- Boes DA, Spaeth GL, Mills RP, *et al.* Relative optic cup depth assessments using three stereo photograph viewing methods. *J Glaucoma* 1996;5:9–14.
- Kahn HA, Leibowitz H, Ganley JP, *et al.* Randomized controlled clinical trial. National Eye Institute workshop for ophthalmologists. Standardizing diagnostic procedures. *Am J Ophthalmol* 1975;79:768–75.
- Schwartz JT. Methodologic differences and measurement of cup-disc ratio: an epidemiologic assessment. *Arch Ophthalmol* 1976;94:1101–5.
- Lichter PR. Variability of expert observers in evaluating the optic disc. *Trans Am Ophthalmol Soc* 1977;74:532–72.
- Klein BE, Moss SE, Magli YL, *et al.* Optic disc cupping as clinically estimated from photographs. *Ophthalmology* 1987;94:1481–3.
- Abrams LS, Scott IU, Spaeth GL, *et al.* Agreement among optometrists, ophthalmologists, and residents in evaluating the optic disc for glaucoma. *Ophthalmology* 1994;101:1662–7.
- Spry PG, Spencer IC, Sparrow JM, *et al.* The Bristol Shared Care Glaucoma Study: reliability of community optometric and hospital eye service test measures. *Br J Ophthalmol* 1999;83:707–12.
- Shuttleworth GN, Khong CH, Diamond JP. A new digital optic disc stereo camera: intraobserver and interobserver repeatability of optic disc measurements. *Br J Ophthalmol* 2000;84:403–7.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
- Streiner DL, Norman GR. *Health measurement scales*. Oxford: Oxford Medical Publications, 1995:104–26 and 163–80.