Title

Agreement in measurement of optic cup-to-disc ratio with stereo bio-microscope funduscopy and digital image analysis.

Results from The Nigeria National Blindness and Visual Impairment Survey

Authors

1. Fatima Kyari[1,2] FWACS, MSc
2. Clare Gilbert[1] FRCOphth, MD

On behalf of the Nigeria National Blindness and Visual Impairment Study Group[3]

1. International Centre for Eye Health (ICEH), London School of Hygiene and Tropical Medicine (LSHTM), London, United Kingdom.
2. Department of Ophthalmology, College of Health Sciences, University of Abuja, Nigeria.
3. The Nigeria National Blindness and Visual Impairment Study Group also consisted of: Abdullahi Imam, Adenike Abiose, Abubakar Tafida, Christian Ezelum, Gabriel Entekume, Hannah Faal, Mansur M Rabiu, Mohammed M Abdull, Olufunmilayo O Bankole, and Pak Sang Lee.

Corresponding author

Fatima Kyari

Financial Support

The Nigeria National Blindness and Visual Impairment Study was supported by Sightsavers International, CBM, and Velux Stiftung.

The data analysis and writing was supported by the Fred Hollows Foundation as part of a research degree study for FK.

The funding organizations had no role in the design and conduct of the research.

Conflict of Interest - No conflicting relationship exists for any author.

Running Head – Agreement in measurement of optic cup-to-disc ratio

Key words – VCDR, glaucoma, Nigeria, Bland-Altman plot, limits of agreement

Correspondence address

Fatima Kyari, International Centre for Eye Health, Department of Clinical Research, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom. Email: Fatima.Kyari@lshtm.ac.uk

Acknowledgement

Tunde Peto, Ferenc Sallo and Nisha Parmar of the MEHRC for grading the optic disc images.

Previous presentation

The findings in this paper have been previously presented, in part, at the 23rd ISGEO Congress, London, United Kingdom. September, 2014.

This submission has not been published anywhere previously and it is not simultaneously being considered for any other publication.

Abstract

Purpose:

To determine agreement of estimations of vertical cup-to-disc ratios (VCDR) between clinical stereo-biomicroscopic funduscopy and digital fundus images analysis.

Methods:

Systematic sampling of 1-in-7 from a sample of 13,591 participants aged ≥40 years gave a subsample who were examined in detail. VCDR was estimated clinically by 60D aspheric lens biomicroscopic funduscopy (c-VCDR) and by fundus images (i-VCDR) graded at Moorfields Eye Hospital Reading Centre. Spearman's correlation coefficient, paired t-test and Bland-Altman method to assess limits of agreement (LOA) between the two methods were applied.

Results:

Of 1759 participants in the subsample, 848 participants (48%) who had normal FDT visual fields and data for i-VCDR and c-VCDR in both eyes (n=1696 eyes) were analysed. By absolute difference of VCDR values for each eye, between the two methods, 94% eyes (n=1585) differed by ≤0.2. Mean i-VCDR was 0.381, standard deviation (SD) 0.156; and mean c-VCDR 0.321, SD 0.145. i-VCDRs were significantly larger by a mean difference of 0.061 SD 0.121 (95% confidence interval [95%CI] 0.055-0.066; p<0.001). The 95% LOA assessed by the Bland-Altman method were lower limit -0.182 (95%CI -0.192; -0.172) and upper limit 0.303 (95%CI 0.293; 0.313). The interval of the 95% LOA narrowed with higher VCDRs.

Conclusion:

Digital image analysis and clinical assessment are two distinct methods of measurement for VCDR; with larger i-VCDRs in this survey. Applying i-VCDR cut-off values to c-VCDR measurements in the Nigeria Blindness Survey might have underestimated glaucoma prevalence. It is recommended that all participants in glaucoma surveys have VCDR by digital image measurement.

Introduction

Glaucoma is the leading cause of irreversible blindness. It is projected to affect up to 80 million people by the year 2020[1] and 111.8 million in 2040.[2] In population-based surveys, evidence of structural optic disc damage is an essential element in identifying individuals who may have glaucoma. The optic vertical cup-to-disc ratio (VCDR) is one of the ways of determining optic disc structural damage.

For the Nigeria National Blindness and Visual Impairment Survey (hereafter referred to as the Nigeria Blindness Survey)[3] glaucoma classification was according to the ISGEO criteria.[4] The glaucoma-defining VCDR values for the 97.5th (0.7) and 99.5th (0.75) percentiles in the study population[5] were applied, reporting a glaucoma prevalence of 5.02% (95%CI 4.60-5.47) among adults aged 40-years old and above.[6] The i-VCDR grading by Moorfields Eye Hospital Reading Centre (MEHRC) was considered the gold standard; it was objective, quantified with a scale and adjudicated. According to the survey protocol, participants who had good visual acuity of 6/9 or better in both eyes would not have fundus photography except if they were among the 1-in-7 subsample or if they had disc abnormalities suggestive of glaucoma detected by direct ophthalmoscopy. Also, fundus photography was not obtained for all participants. Thus, these i-VCDR cut-off values were also applied to clinically graded VCDR of participants who did not have digital fundus photography with optic disc imaging.

In this report, data from the Nigeria Blindness Survey were analysed to determine the agreement in measurement of i-VCDR and c-VCDR among adults aged ≥40 years in a subset of participants who had both methods of assessment i.e. the 1-in-7 subsample. Determining the agreement between the two VCDR measurement methods will potentially inform VCDR measurement in subsequent glaucoma prevalence surveys. It will also enable better interpretation of the results obtained by applying the i-VCDR defining percentile values to the whole dataset for glaucoma classification in the National Blindness Survey.

Methods

Details of all the methods used in the Nigeria Blindness Survey,[3] a report on the defining values for glaucoma in prevalence surveys in Nigeria[5] as well as the prevalence and types of glaucoma in Nigeria have been published.[6]

Study design, data collection and clinical assessment

The sample size calculation and sampling strategy for the Nigeria Blindness Survey gave a nationally representative sample of 15,375 persons aged 40 years and above in 310 clusters across the country. Multi-stage sampling using probability-proportional-to-size methods were used to select the study population. A further systematic sampling of 1-in-7 participants registered at the examination centre was done. All participants were invited to a temporary clinic-type set up for examination. Data were collected by two teams, each comprising

of two ophthalmologists, one optometrist and two ophthalmic nurses. The ophthalmologists received further training in survey protocols and standardising VCDR measurement.

Systematic sampling of 1-in-7 from a sample of 13,591 participants aged ≥40 years gave a subsample who were examined in detail, including visual field assessment with a Humphrey FDT visual field analyzer (Carl Zeiss Meditec AG Jena Germany).

The first ophthalmologist performed undilated direct funduscopy. Detailed eye examination performed by the second ophthalmologist included slit-lamp biomicroscopy (Zeiss SL 115 Classic Slit Lamp, Carl Zeiss Meditec AG Jena Germany) and dilated retinal examination and optic disc assessment using 60D aspheric condensing lens (Volk). VCDR was estimated clinically (c-VCDR) by determining the rim of the optic disc and estimating the cup size in the vertical meridian and calculating the spatial ratio between the optic cup and the optic disc.

Participants also had digital retinal photography (Zeiss Visucam Lite Desk Top Fundus Camera, Carl Zeiss Meditec AG Jena Germany) through a dilated pupil focused mid-point between the macular and optic nerve head displaying a field of 45 degrees showing both the macula and the optic disc in the observed field. Images were graded independently by the Moorfields Eye Hospital Reading Centre (MEHRC). Images were viewed "full screen" on either a 24-inch Eizo S2433W monitor calibrated using a Datacolor Spyder2 calibrator or on a 24-inch widescreen Dell 2407WFP LCD monitor calibrated using a Gretag Macbeth Eye-One Display 2 calibrator. After determining image quality and clarity, the scleral rim was identified and the boundaries of the disc and the cup were identified.

One successful measurement was performed per eye, along the vertical meridian, in Adobe Photoshop (version 7) using the measurement tool, resulting in a cup and a disc diameter value in pixels, along the same plane, the division of the two values producing the i-VCDR which was recorded to the nearest 0.05. Primary grading was performed by the 1st reader (FS), and a 2nd reader (NP) and inconclusive cases were adjudicated by a 3rd reader (TP).

Inter-observer agreement assessments were conducted for ophthalmologists on c-VCDR measurement during the training sessions and at intervals during fieldwork. There was one clinical ophthalmologist in each of the two teams throughout the survey. For i-VCDR grading, inter-observer agreement between the 1st and 2nd readers was assessed. Kappa statistics for inter-observer error were calculated.

Ethics

Ethical approval was obtained from the Ethics Committee of the London School of Hygiene and Tropical Medicine (LSHTM), UK and the Nigeria National Health Research and Ethics Committee (NHREC). Oral informed consent was obtained from participants. The study adhered to the tenets of the declaration of Helsinki. Persons with medical or eye conditions needing further assessment and treatment were referred to the nearest healthcare facility.

Data analysis

Statistical analysis was performed using Stata/IC 14.0 (Stata Corp, College Station, TX). Included in the analysis were both eyes of the 848 participants in whom both eyes had VCDR grading by the two methods (slit-lamp biomicroscopic funduscopy with 60D aspheric lens and digital fundus photography image analysis), normal FDT visual fields and no detected ocular pathology.

For kappa analysis, ophthalmologists' clinical measurement of VCDR within 0.1 in one session was assessed; and image grading within 0.2 obtained by the two primary readers was assessed.

Frequency distribution of the absolute difference between the VCDR values in each eye was determined. The frequency distributions of c-VCDR and i-VCDR were determined and compared; and the Shapiro-Wilk test of normality was applied. The association between the two methods of measurement was calculated and expressed as the Spearman's rank-order correlation coefficient. Paired t-test was applied for comparison of means to investigate the presence of any systematic (fixed) bias.

Bland-Altman method to assess 95% limits of agreement (LOA) between the two methods was applied.[7-9] To assess agreement on the Bland-Altman plot, the y-axis was the difference between the two measurement methods (i-VCDR minus c-VCDR), i.e. the amount of disagreement, plotted against the x-axis, the mean of the two measurements. The LOA were the mean differences $\pm$ 2 standard deviation of the differences. The 95% confidence intervals for the upper and lower LOA were calculated. The difference between the two measurement methods was regressed on the average of the two measurements and the slope of least-squares regression with the regression-based 95% LOA were

determined. The slope of the least squares regression line was tested if it significantly differed from zero, to investigate the presence of any proportional bias. The adjusted R-squared was calculated by linear regression analysis.

The Bland-Altman method was also applied to the data where c-VCDR was $\geq$0.6 and a quadratic regression line was determined.

Results

The kappa for inter-observer agreement for ophthalmologists' clinical measurement of VCDR within 0.1 was κ=0.86 (almost perfect agreement), calculated for one assessment session for 85 eyes. For the image VCDR grading at MEHRC, the overall inter-observer agreement between two graders for 847 eyes was 99.7% and the kappa for $\leq$0.2 difference between graders was κ=0.50 (moderate agreement).

Both eyes of 848 participants (n=1696) were included in the analysis. All eyes had data for VCDR obtained by the two methods of measurement. The clinical estimates recorded a zero (0) VCDR in 63 eyes (3.7%); and were recorded in 0.1 difference steps. The image grading was recorded in 0.05 steps in 103 eyes (6.1%) (Figure 1).

Frequency distribution of the absolute difference of VCDR values for each eye, between the two methods showed that 75% eyes (n=1278) differed by $\leq$0.1, and 94% eyes (n=1585) differed by $\leq$0.2.

The Shapiro-Wilk test for normality of data showed that the i-VCDR had a high value of W (0.994) but $p<0.001$. Thus the distribution of values for i-VCDR in this sample did not follow a normal distribution. The distribution of values for c-VCDR followed a normal distribution (W=998, p=0.10).

The Spearman's rank-order correlation coefficient showed a strong positive correlation between i-VCDR and c-VCDR which was statistically significant ($r_s=0.67$, $p<0.001$).

The mean VCDR differed significantly between the i-VCDR, 0.381 standard deviation (SD) 0.156 and the c-VCDR, 0.321 SD 0.145. The difference in the means was 0.061 (95%CI 0.055 – 0.066), SD 0.121, $p<0.001$ and suggestive of a systematic (fixed) bias. Where the c-VCDR was $\geq$0.6 (n=98) the difference in the two means was 0.022 (95%CI 0.002-0.042), SD 0.101; p=0.03.

The Bland-Altman plot in Figure 2 shows agreement between the two methods of estimating VCDR where the difference between the two measurement methods was plotted against the average of the two measurements, which is assumed to be the best estimate of the true value. The line of no difference (a; solid green line) indicates where there is no difference between the two methods of measurement at this level. Most of the points are above this line, indicating that measurements of i-VCDR were higher than c-VCDR with an average discrepancy bias of 0.061, indicated as a solid red line (b). The 95% LOA are indicated by the two solid horizontal black lines (c) which demarcate the upper LOA 0.303 (95%CI 0.293 to 0.313) and the lower LOA -0.182 (95%CI -0.192 to -0.172). Figure 2 also shows the least squares regression line indicated by the red dash line (d) with regression-based 95% LOA (e; dash-3dot lines). The trend in the plot showed that the interval between the upper and lower LOA narrowed with

higher VCDR values showing fewer data points and indicating that the differences between the two measurements became smaller for higher average VCDR values. The least squares linear regression line (d; red dash line) significantly differed from zero ($p<0.001$), indicating the existence of proportional bias. The existence of proportional bias implies that the two methods of VCDR measurement did not agree equally through the range of measurements.

Further, in eyes with c-VCDR$\geq$0.6 (n=98) the quadratic model (Figure 3), appeared to fit the data better than a linear model with the quadratic regression line (d; red dash line) trending towards the line of no difference (a; solid green line) as the average VCDRs increase.

Discussion

In investigating the difference between the two methods of measurement of VCDR by clinical slit-lamp biomicroscopic funduscopy with 60D aspheric lens and by digital fundus photography image grading, the use of Bland-Altman method of assessment[7-9] determines how closely the two methods agreed. The discrepancy between the two methods was clinically substantial and showed that they were two distinct methods of measurement of VCDR with a statistically significant average discrepancy of 0.061. With higher c-VCDRs, the average discrepancy was less (0.022). This may mean that when the disc changes are obvious, the detection and measure would be easier and more similar for both

methods. Digital image analysis gave larger VCDRs than clinical assessment. We acknowledge this difference, thus the application of image VCDRs glaucoma-defining percentile values to the whole Nigeria Blindness Survey data might have underestimated the prevalence of glaucoma in the population.

The advantages of using digital image grading of the VCDR were that the disc images, under the survey conditions, could be captured and kept as records, which could be reviewed objectively and quantified with a scale. The images could also be used for follow-up. However, the disadvantages were that the fundus camera did not take stereoscopic images so monocular clues were used to determine cup and disc boundaries; and the angle of projection might affect the spatial measurement of disc parameters. Whereas with the 60D aspheric lens, stereoscopic images were viewed and assessed but a measurement graticule was not used; and the disc assessments were not documented with hand-drawings on a template, thus it would be difficult to review afterwards or adjudicate.

The Bland-Altman method to assess LOA has been applied to various methods of measuring VCDR and the results vary. Jayasundera[10] and Durmus,[11] in their respective studies, found poor agreement between stereoscopic photographs, clinical assessment, HRT and digital stereoscopic optic disc camera which was worse for small discs and smaller cups.[10] The Rotterdam study also showed that semi-automated VCDR measurements were larger than ophthalmoscopic estimates with a moderate correlation.[12] Perera compared clinical measurement of VCDR (mean 0.40$\pm$0.12) using an eyepiece graticule with HRT (mean

0.37±0.21) and with OCT (mean 0.50±0.14) and found lack of agreement ($p<0.001$), with OCT tending to overestimate VCDR and HRT tending to underestimate values.[13]

Limitations of this study are that it was a retrospective analysis and a graticule was not incorporated in the clinical VCDR measurement. A 0.01mm graticule might have increased the accuracy of clinical measurement and also improved inter-observer agreement.[14] Regarding image quality, lack of centration of the optic disc in the image could have contributed to differences due to magnification and positioning. An algorithm for measurement which takes into account the magnification factor and the actual size in micrometre of one pixel used directly on the images with the participant still available for re-examination has been found to be useful in both population-based measurement and clinical practice.[15]

Based on the differences between the two methods, the use of optic disc imaging/photography with digital image analysis for measurement of VCDR is recommended in glaucoma prevalence survey for all participants. This will provide uniformity and objective evaluation of VCDR and comparable glaucoma prevalence estimates.

References

1. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. Br J Ophthalmol. 2006:90(3);262-7.

2. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology. 2014:121(11);2081-90.

3. Dineen B, Gilbert CE, Rabiu M, Kyari F, Mahdi AM, Abubakar T et al. The Nigerian national blindness and visual impairment survey: Rationale, objectives and detailed methodology. BMC Ophthalmol. 2008:8;17.

4. Foster PJ, Buhrmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. Br J Ophthalmol. 2002:86(2);238-42.

5. Kyari F, Abdull MM, Sallo FB, Spry PG, Wormald R, Peto T et al. Nigeria normative data for defining glaucoma in prevalence surveys. Ophthalmic Epidemiol. 2015:22(2);98-108.

6. Kyari F, Entekume G, Rabiu M, Spry P, Wormald R, Nolan W et al. A Population-Based Survey of the Prevalence and Types of Glaucoma in

Nigeria. The Nigeria National Blindness and Visual Impairment Survey. BMC Ophthalmol. 2015:15(1);176.

7. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986:1(8476);307-10.

8. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet. 1995:346(8982);1085-7.

9. Bland JM, Altman DG. Agreed statistics: measurement method comparison. Anesthesiology. 2012:116(1);182-5.

10. Jayasundera T, Danesh-Meyer HV, Donaldson M, Gamble G. Agreement between stereoscopic photographs, clinical assessment, Heidelberg retina tomograph and digital stereoscopic optic disc camera in estimating vertical cup:disc ratio. Clin Experiment Ophthalmol. 2005:33(3);259-63.

11. Durmus M, Karadag R, Erdurmus M, Totan Y, Feyzi Hepsen I. Assessment of cup-to-disc ratio with slit-lamp funduscopy, Heidelberg Retina Tomography II, and stereoscopic photos. Eur J Ophthalmol. 2009:19(1);55-60.

12. Wolfs RC, Ramrattan RS, Hofman A, de Jong PT. Cup-to-disc ratio: ophthalmoscopy versus automated measurement in a general population: The Rotterdam Study. Ophthalmology. 1999:106(8);1597-601.

13. Perera SA, Foo LL, Cheung CY, Allen JC, Chua D, Tham YC et al. Cup-to-Disc Ratio From Heidelberg Retina Tomograph 3 and High-Definition Optical Coherence Tomography Agrees Poorly With Clinical Assessment. J Glaucoma. 2015.

14. Haslett RS, Batterbury M, Cuypers M, Cooper RL. Inter-observer agreement in clinical optic disc measurement using a modified 60 D lens. Eye (Lond). 1997:11(Pt 5);692-7.

15. Bartling H, Wanger P, Martin L. Measurement of optic disc parameters on digital fundus photographs: algorithm development and evaluation. Acta Ophthalmol. 2008:86(8);837-41.

Figures legend

Figure 1. The frequency distribution of values for image VCDR and clinical VCDR

Figure 2. Bland-Altman plot comparing VCDR measured by clinical slit-lamp biomicroscopy funduscopy and by fundus photography image grading

Figure 3. Bland-Altman plot comparing VCDR measured by clinical slit-lamp biomicroscopy funduscopy and by fundus photography image grading if c-VCDR $\geq$0.6

Figure 1. The frequency distribution of values for image VCDR and clinical VCDR

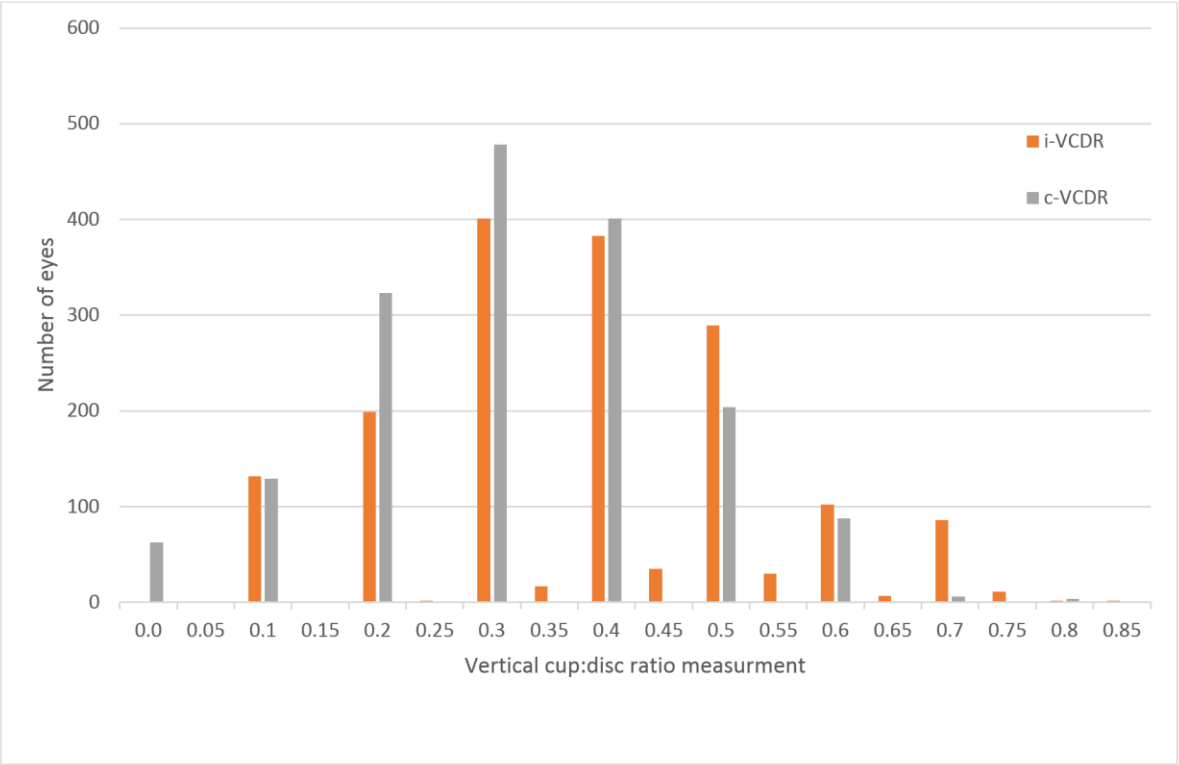
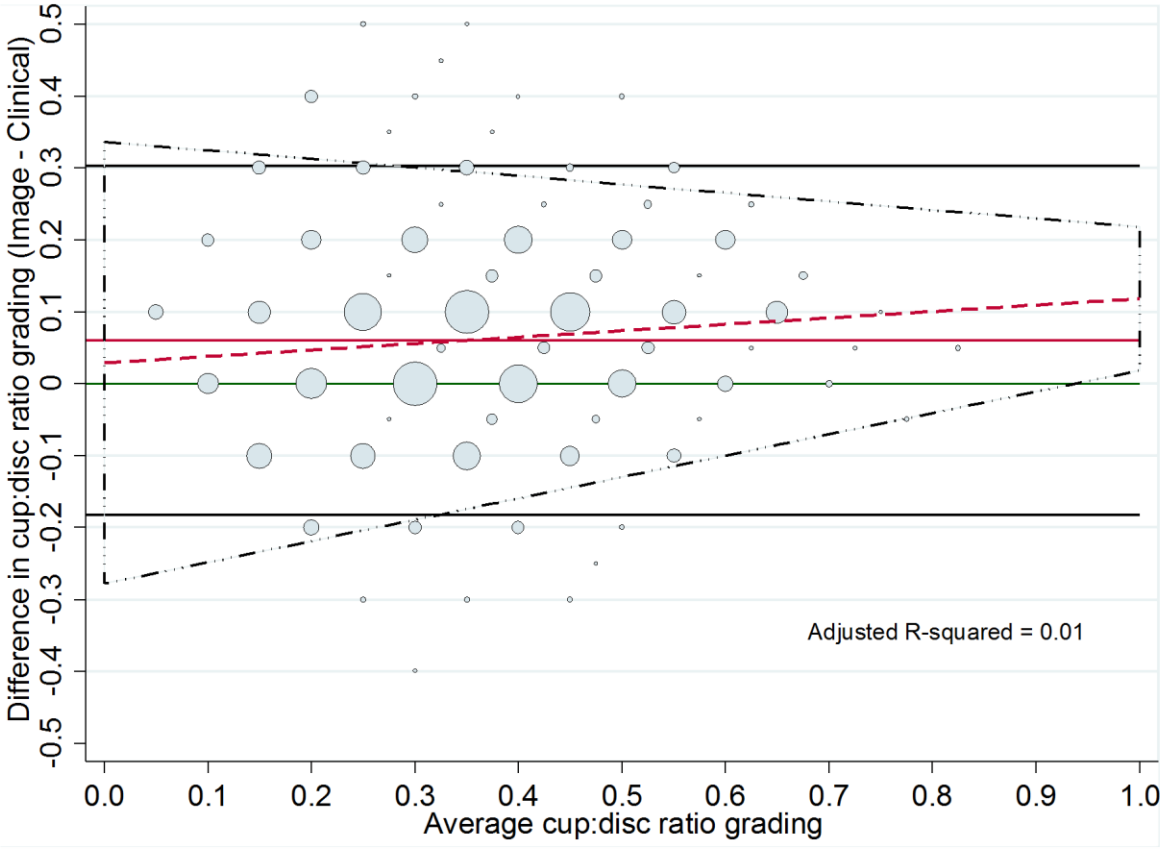


Figure 2. Bland-Altman plot comparing VCDR measured by clinical slit-lamp biomicroscopy funduscopy and by fundus photography image grading



Lines shown indicate:

a) _____ Solid green line is the line of no difference at zero.

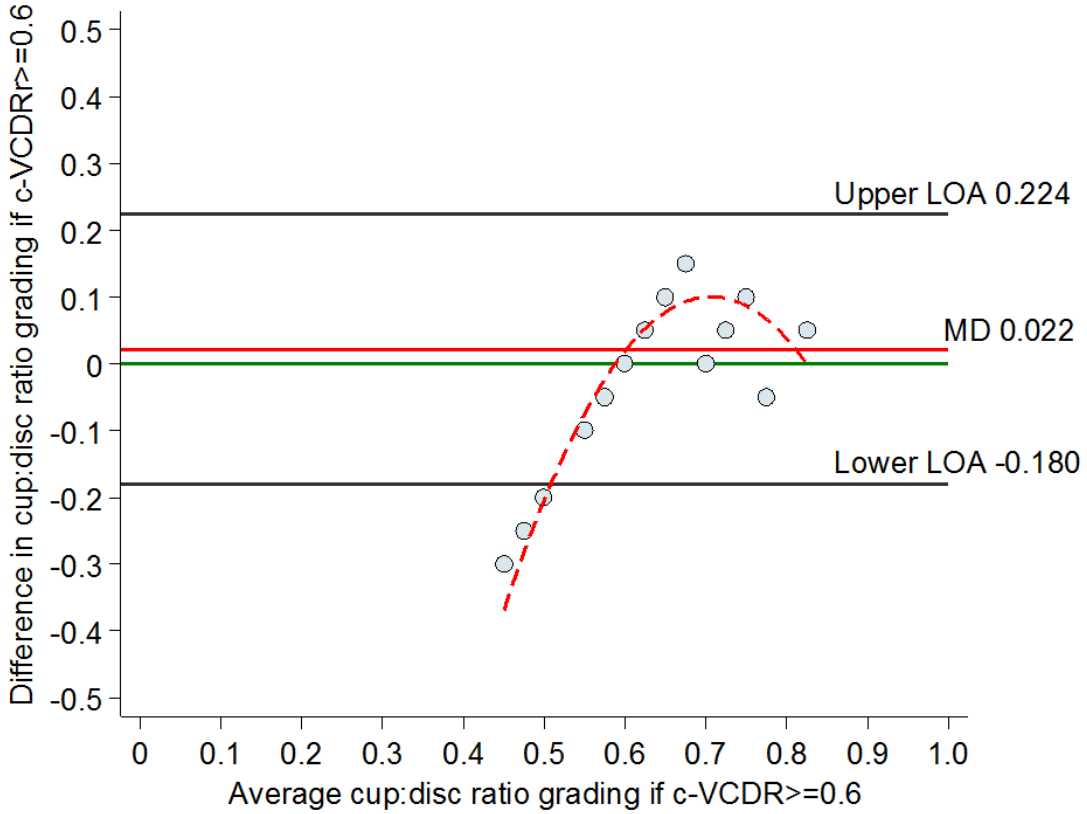
b) _____ Solid red line is the mean difference (0.061)

c) _____ Solid horizontal black lines demarcate the upper limit of agreement (Mean + 2SD) 0.303; and lower limit of agreement (Mean – 2SD) -.182

d) ---------- Red dash line is the slope of least-squares regression for the difference between the two measurements on the average of the 2 measurements.

e) _..._..._ Dash-3dot lines are the regression-based 95% limits of agreement

Figure 3. Bland-Altman plot comparing VCDR measured by clinical slit-lamp biomicroscopy funduscopy and by fundus photography image grading if c-VCDR ≥0.6



Lines shown indicate:

a) _____ Solid green line is the line of no difference at zero.
b) _____ Solid red line is the mean difference (0.022)
c) _____ Solid horizontal black lines demarcate the upper limit of agreement (Mean + 2SD) 0.224; and lower limit of agreement (Mean – 2SD) -0.180
d) ---------- Red dash line is the quadratic regression for the difference between the two measurements on the average of the two measurements.

c-VCDR = clinical vertical cup:disc ratio; LOA = limit of agreement; MD = mean difference; SD = standard deviation.