Validity of a New Optic Disc Grading Software for use in Clinical and Epidemiological Research

Journal:	*Clinical and Experimental Ophthalmology*
Manuscript ID:	CEO-12-12-0939.R1
Manuscript Category:	Original Article - Clinical Science
Date Submitted by the Author:	07-Mar-2013
Complete List of Authors:	Tham, Yih-Chung; Singapore Eye Research Institute, ; Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Cheung, Carol; Singapore Eye Research Institute, ; Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, ; Centre for Quantitative Medicine, Office of Clinical Sciences, Duke-NUS Graduate Medical School, Wong, Tien Yin; Singapore Eye Research Institute, ; Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Baskaran, Mani; Singapore Eye Research Institute, ; Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Liu, Jiang; Institute for Infocomm Research, Agency for Science, Technology and Research, Lee, Beng-Hai; Institute for Infocomm Research, Agency for Science, Technology and Research, Wang, Jie Jin; University of Sydney, Westmead Millennium Institute, Department of Ophthalmology, Centre for Vision Research; Mitchell, Paul; University of Sydney, Department of Ophthalmology, Centre for Vision Research Aung, Tin; Singapore Eye Research Institute, ; Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Cheng, Ching-Yu; Singapore Eye Research Institute, ; Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, ; Centre for Quantitative Medicine, Office of Clinical Sciences, Duke-NUS Graduate Medical School,
Keywords:	Optic nerve evaluation, Optic disc, Epidemiology

Original Article – Clinical Science

Validity of a New Optic Disc Grading Software for use in Clinical and Epidemiological Research

Yih-Chung Tham BSc (Hons) (1,2), Carol Y Cheung PhD (1,2,5), Tien Y. Wong FRANZCO PhD (1,2), Mani Baskaran MBBS (1,2), Jiang Liu PhD (6), Beng-Hai Lee MSc (6), Fengshou Yin BSc (6), Damon W.K Wong PhD (6), Jie Jin Wang PhD (4), Paul Mitchell FRANZCO PhD (4), Tin Aung FRCOphth PhD (1,2), Ching-Yu Cheng MD PhD (1,2,3,5)

(1) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore;
(2) Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore;
(3) Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore
(4) Department of Ophthalmology (Centre for Vision Research), Westmead Hospital, University of Sydney, Australia;
(5) Centre for Quantitative Medicine, Office of Clinical Sciences, Duke-NUS Graduate Medical School, Singapore;
(6) Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore.

Corresponding author:

Dr Ching-Yu Cheng

T: +65 6772 3897 F: +65 6777 7161

Email: ching-yu_cheng@nuhs.edu.sg

Postal Address: Department of Ophthalmology, National University Health System,

1E Kent Ridge Road, NUHS Tower Block Level 7, Singapore 119228

Short title: New Optic Disc Grading Software

Competing/ Conflict of Interest: No stated conflict of Interest.

Funding Sources: Grants from the National Medical Research Council and the Biomedical Research Council, Singapore [Grant No. 0796/2003 & 08/1/35/19/550].

Abstract

Background: To determine the reliability and agreement of a new optic disc grading software program for use in clinical, epidemiological research.

Design: Reliability and agreement study.

Samples: 328 monoscopic and 85 stereoscopic optic disc images.

Methods: Optic disc parameters including vertical cup-to-disc ratio (VCDR) were measured using a new optic disc grading software (Singapore Optic Disc Assessment, SODA) which is based on polynomial curve-fitting algorithm for demarcation of cup, disc margins. Two graders independently graded 328 monoscopic images to determine inter-grader reliability. One grader re-graded the images after 1 month to determine intra-grader reliability. In addition, 85 stereo optic disc images were separately selected and VCDRs were measured using both SODA and standardized Wisconsin manual stereo-grading method by the same grader 1-month apart. Reliability and agreement analyses were evaluated using intraclass correlation coefficient (ICC) and Bland-Altman plot analyses.

Main Outcome Measures: Optic disc parameters

Results: The intra- and inter-grader reliability for optic disc measurements using SODA was high (ICC ranging from 0.82 to 0.94). The mean differences (95% limits of agreement [LOA]) for intra- and inter-grader VCDR measurements were 0.00 (-0.12 to 0.13) and 0.03 (-0.15 to 0.09) respectively. The VCDR agreement between SODA and Wisconsin grading method was extremely close (ICC= 0.94). The mean difference (95% LOA) of VCDR measurement between SODA and Wisconsin grading methods was 0.03 (-0.09 to 0.16).

Conclusions: Intra- and inter-grader reliability using SODA was excellent. SODA measurements were highly comparable with standardized manual stereo-grading method. SODA is useful for grading digital optic disc images in clinical, population-based studies.

Introduction

Glaucoma is the leading cause of global irreversible blindness and affects upwards of 60 million people worldwide[1-3]. Objective, reliable evaluation of optic disc morphology is essential for clinical research on glaucoma[3-5]. This is particularly important as structural optic nerve head damage often precedes detectable loss in visual function[6,7]. In addition, evaluation of optic disc morphology in large-scale, population-based studies may contribute to the establishment of a comprehensive database and provide greater insights on the variations of optic disc morphology across populations.

Standardized optic disc grading in larger population-based studies is even more challenging due to the large number of subjects, optic disc images and resources involved. As a result, many epidemiological studies have relied on clinical assessment of optic disc characteristics to define glaucomatous optic neuropathy[8-12]. However, clinical assessment is prone to substantial intra- and inter-observer variability[13-15]. While various optic disc grading software programs are available, these have shortcomings such as being camera-specific, have limited optic disc marking points, use artificial reference plane, and have imprecise optic nerve head shaping algorithm, which limit the broader application of these software programs in optic disc measurements for clinical and epidemiological research[16-19].

Currently, subjective optic disc evaluation by observing the photograph pair with stereo viewer remains the gold standard to assess structural glaucomatous appearance[20,21]. Klein et al[22] established a manual stereo optic disc grading method (referred to here as the Wisconsin grading method) to quantify optic disc cupping from stereoscopic fundus photographs. The Wisconsin grading method was employed and reported in the Beaver Dam Eye Study[23] and the Blue Mountains Eye Study (BMES)[24], with optimal intra- and inter-grader agreement[22,25]. Nonetheless, this manual grading method is resource intensive, requires significant training, is time consuming (approximately 8 to 10 minutes per eye) and could not be easily applied to digital non-stereoscopic images. Hence, despite its optimal reproducibility, this method is not widely used and may not be ideal for digital optic disc grading in clinical and epidemiological studies.

In view of the lack of an efficient, time-saving and reliable optic disc grading tool for digital images in clinical and population-based research, we have developed a new semi-automated optic disc grading tool to measure optic nerve head parameters from retinal fundus photographs. The purpose of this study was to assess the reliability and agreement of optic nerve head measurement using this tool for use in clinic-based and population-based studies.

Methods

Study Population

We evaluated the optic disc grading software program (Singapore Optic Disc Assessment, SODA) using images from the Singapore Epidemiology of Eye Disease (SEED) program. SEED program consists of a series of population-based cross-sectional studies which evaluate prevalence of eye diseases in Malay, Indian and Chinese adults aged between 40 and 80 years old, residing in Singapore. The objectives and methodology of these population-based studies have been reported in detail elsewhere.[26,27] For all SEED studies, optic discs were examined and clinical cup-to-disc ratio (CDR) was evaluated by study ophthalmologists using slit lamp indirect ophthalmoscopy. The studies adhered to the Declaration of Helsinki, and ethics committee approval was obtained from the Institutional Review Board of Singapore Eye Research Institute (SERI). Written informed consent was obtained from all participants.

In this report, a multi-ethnic subsample of 328 optic disc images was randomly selected from SEED studies, comprising of 103 Malays, 112 Indians and 113 Chinese subjects. The selection was made to obtain subjects with a wide range of clinically measured CDR (as described previously; referred to here as clinical CDR) and without significant image opacities: 104 subjects with clinical CDR between 0.1 and 0.39, 144 subjects with clinical CDR between 0.4 and 0.59 and 79 subjects with clinical CDR 0.6 or larger. Optic disc images were randomly selected from either eye of each subject.

In addition, a random subsample of 85 stereoscopic optic disc photographs from the Blue Mountains Eye Study (BMES) was used to assess the agreement of vertical CDR (VCDR) measurement between SODA and the standard Wisconsin grading method[22]. The BMES is a population-based cohort study of vision and common eye diseases in an adult population aged 49 years or older. The methods and procedures have also been described elsewhere[28]. The study was approved by the Western Sydney Area Health Service Human Ethics Committee and written informed consent was obtained from all participants. As with the SEED sample, the selection was made to obtain subjects with a wide range of clinically measured CDR and without significant image opacities: 19 subjects with clinical CDR between 0.1 and 0.39, 29 subjects with clinical CDR between 0.4 and 0.59 and 37 subjects with clinical CDR 0.6 or larger. Optic disc images were randomly selected from either eye of each subject.

Retinal Photography

In SEED studies, 45º digital fundus photographs were taken using a telecentric monoscopic fundus camera (Canon CR-DGi with 10D/ 20D/ 40D SLR back; Canon, Tokyo, Japan) after pupil dilation. In each photograph, the optic disc was well positioned at the centre of the photograph.

In BMES, 30° stereoscopic retinal photographs were taken using a Zeiss FF3 fundus camera (Carl Zeiss, Oberkochen, Germany) after pupil dilation. Likewise, in each photograph, the optic disc was positioned at the centre of the photograph. These stereoscopic pair photographs were processed and developed as 35mm slide colour transparencies[24]. The two transparencies of pair photographs were mounted on clear plastic sheets and placed side by side during stereoscopic evaluation. In addition, these 35 mm colour transparencies were also digitized using a CanoScan FS2710 (Canon,Tokyo, Japan) scanner. Images were converted to an 8-bit grayscale digital format with resolution 2720 dpi and stored in Tagged Image File Format (TIFF). The scanner was driven by a PC via SCSI interface[15].

Grading Protocol

Digital optic disc images from the SEED subsample were analyzed using the SODA, a new, customized software program jointly developed by Institute for Infocomm Research (I^2R) and SERI. Quantitative optic disc grading was performed by carefully demarcating the optic disc and cup margin following the Wisconsin grading protocol[22]. The disc margin was defined as the inner margins of the peripapillary white scleral ring. Demarcation of cup contour was judged based on maximal inflection of vessels at the inner edge rather than the colour appearance on the optic disc surface. Vessels were considered to be part of the cup if there were no underlying rim tissues.

These optic disc images were measured using SODA on a computer workstation with 13-inch display LCD monitor. Images were displayed at a resolution of 1280 X 1024 pixels. During measurement, the optic disc and optic cup margins were first plotted; the plotted points were then connected by the program's polynominal curve fitting algorithm to segment the optic disc and optic cup regions. The software automatically calculated a range of parameters, including VCDR, vertical cup diameter, vertical disc diameter, cup-to-disc area ratio (CDAR), cup area and disc area based on the segmented regions. The raw measured values from SODA were expressed in pixels. Absolute value conversions (to mm and mm^2) were done by taking into account the scale factors (microns/ pixel) of respective camera magnifications used in this study. The detailed method to calculate the scale factors has been described elsewhere[24,29]. **Figures 1A-B** illustrates examples of the user interface of the SODA software program.

Stereoscopic retinal photograph transparencies from the BMES subset were graded using the Wisconsin grading method[22]. Details of this manual measurement method have been described in detailed elsewhere[22]. In brief, stereo pair transparencies were viewed using a Donaldson Stereoviewer against a light box background. Measurement of optic cup and optic disc margins were performed using a Pickett circle template. In addition, digitized images from BMES subset were also analyzed monoscopically using SODA.

An experienced optometrist (grader A) and glaucoma specialist (grader B), masked to subject characteristics and clinical diagnosis, independently measured the 328 SEED optic disc images using SODA to assess inter-grader reliability. In addition, grader A repeated the measurement using SODA after 1 month to assess intra-grader reliability. Grader A also measured VCDR in the 85 BMES images using both the stereoscopic Wisconsin grading method and SODA monoscopically within 1-month, to assess the agreement between the two methods. This was to validate the SODA software program for VCDR measurement. Both graders assessed the same sets of monoscopic and stereoscopic training photographs before commencing the grading task. **Figure 2** shows a flow diagram which summarizes the study design in this study.

Statistical Analysis

Data normality was assessed using the Shapiro-Wilk test. Intra-, inter-grader reliability and agreement between the 2 methods for assessing optic disc parameters were evaluated using the absolute agreement model of the intraclass correlation coefficient (ICC)[30] and Bland Altman plot analyses[31,32].

ICC values between 0.81 and 1.00 indicate almost perfect agreement, values between 0.61 and 0.80 indicate good agreement and values between 0.41 and 0.60 indicate moderate agreement. Values less than 0.40 indicate poor to fair agreement[33,30]. In Bland Altman plot analyses, the 95% limits of agreement (LOA) were defined as mean difference ± 1.96 x standard deviations. In the Bland Altman plots, the difference between 2 measurements was plotted against the average of 2 measurements. Where a trend in the plot was identified, the slope of the least squares regression line was tested to see if it significantly differed from zero to investigate the presence of any proportional bias. This was tested by Pearson's correlation coefficient[13,34,35]. On the other hand, the mean difference value was compared to the zero value of difference using one-sample test to investigate the presence of any systemic (fixed) bias. Presence of proportional bias would indicate that the discrepancies between the 2 measurements were not constant throughout the range of measurements. Conversely, systemic bias would indicate that the discrepancies were constant/ fixed throughout the range of measurements. In the intra- and inter-grader reliability analyses, using a 95% confidence interval, a sample size of 328 would yield a margin of error of 5.41%. In the agreement analysis between SODA and Winconsin grading method, using a 95% confidence interval, a sample size of 85 would yield a margin of error of 10.63%. Statistical analyses were performed using SPSS version 17.0 (SPSS Inc, Chicago, IL) and MedCalc version 12 (MedCalc Software bvba, Mariakerke, Belgium).

Results

Characteristics of the subjects are shown in Table 1. The average clinically measured VCDR using slit lamp indirect ophthalmoscopy was 0.48 ± 0.16 for Malays, 0.46 ± 0.13 for Indians, 0.46 ± 0.15 for Chinese and 0.57 ± 0.21 for Whites. In this study, the normality test (Shapiro-Wilk test) for various optic disc parameter measurements showed that the distribution of these measurements followed a normal distribution. Optic disc parameters measured by the two graders based on digital optic disc images sampled from the SEED studies are summarized in Table 2. The mean VCDR measured by grader A and B was 0.56 ± 0.12 and 0.59 ± 0.12, respectively.

Intra- and inter-grader reliability using SODA

Intra-grader measurements of respective optic disc parameters showed ICC values of 0.87 to 0.94, indicating high reliability (Table 3). Bland Altman plot analyses for VCDR and CDAR intra-grader reliability are shown in **Figures 3A and 3B**, respectively. The Bland Altman plot of VCDR shows 95% LOA of -0.12 to 0.13 with a mean difference of zero. The Bland Altman plot of CDAR shows 95% LOA of -0.11 to 0.12 with a mean difference of 0.01. No significant systemic and proportional bias was detected in intra-grader VCDR and CDAR measurement comparisons.

Inter-grader measurements showed almost perfect agreement with ICC values ranging from 0.82 to 0.94 for respective parameters. The Bland Altman analysis of VCDR shows 95% LOA of -0.15 to 0.09 with a mean difference of -0.03 (**Figure 4A**). The Bland Altman analysis of CDAR shows 95% LOA of -0.14 to 0.14 with a mean difference of zero (**Figure 4B**). No significant proportional bias was detected in the inter-grader VCDR and CDAR measurement comparison. Nonetheless, significant systemic bias was found in the inter-grader VCDR measurement comparison ($P<0.001$). Similarly, in the respective subgroups of clinical CDR ≥ 0.60, beta-type peripapillary atrophy and tilted disc, good intra- and inter-grader reliability was found with ICC values ranging from 0.70 to 0.95 (data not shown in Table 3).

Agreement between the SODA and the Wisconsin grading methods performed by the same grader

As shown in table 3, agreement between the SODA and the Wisconsin grading methods in measuring VCDR was very strong with an ICC value of 0.94 (95% CI, 0.88 to 0.97). The

Bland Altman plot analysis shows that SODA measured VCDR slightly larger as compared to the Wisconsin grading method with a mean difference of 0.03 and a 95% LOA of -0.09 to 0.16 **(Figure 5).** Significant systemic bias ($P < 0.001$) was found in the VCDR measurement between SODA and Wisconsin grading methods. In addition, the differences between SODA and Wisconsin grading methods measurements correlated significantly with the average of the 2 measurements ($r = -0.331$, $p = 0.002$). Similarly, in the respective subgroups of clinical CDR ≥ 0.60, beta-type peripapillary atrophy and tilted disc, strong agreement between the SODA and the Wisconsin grading methods was found with ICC values ranging from 0.87 to 0.94 (data not shown in Table 3).

Discussion

Evaluation of optic disc morphology is essential for clinic-based and population-based research in glaucoma. A simple, objective and reproducible optic disc grading method will enhance our ability to study the ethnic variations in optic disc morphology and the risk factors of glaucoma. In this study, we showed that SODA software program allows precise plotting and segmentation of the optic cup and optic disc borders. This optic disc grading method and protocol produce excellent intra- and inter-grader reproducibility. In addition, VCDR measurement using SODA is highly comparable to the reference standard method, the Wisconsin manual stereo optic disc grading method. These findings demonstrate that SODA software program has promising qualities for mass optic disc grading involving digital images.

Stereo-photography has always been regarded as the gold standard in optic nerve head evaluation[19,36,37]. Previous studies reported that stereo-photography produced better inter-observer agreement than monoscopic photographs in CDR measurements[15,18,38]. Using stereo-photography, Varma et al[38] studied the inter-observer agreement in VCDR measurement; they found good agreement between 6 glaucoma experts (Kappa Value = 0.67). Similarly, Abrams et al[39] evaluated VCDR using stereoscopic optic disc photography and found good inter-observer agreement between 6 ophthalmologists (Kappa value = 0.68). Likewise, Sung et al[13] reported good inter-observer agreement in CDAR measurements using digital sequential stereoscopic optic disc images (ICC = 0.79). Compared to these previous studies, we found greater VCDR and CDAR inter-observer reliability (ICC = 0.89 to 0.92) despite the usage of digital monoscopic images in our study. This can be partly explained by the fact that previous studies involved more graders and standardized training was not given, both of which might have resulted in poorer inter-observer agreement. Furthermore, our findings further indicate that it is feasible to achieve reliable measurements from digital monoscopic images by adhering to a rigorous protocol.

In previous studies, it has been consistently shown that inter-grader measurements have higher variability than intra-grader measurements in optic disc parameter measurements particularly in optic cup measurements[38-40]. In contrast, Shuttleworth et al[19] showed almost perfect agreement for both intra-grader (ICC 0.92 to 0.95) and inter-grader (ICC 0.89 to 0.92) optic disc parameter measurements (VCDR, CDAR) using stereo-photography. In this study, we also found excellent intra- and inter-grader agreement for respective optic disc parameters despite employing digital monoscopic images. In addition, the range of LOA for respective intra-grader optic disc parameter measurements was comparable with the range of LOA for respective inter-grader optic disc parameter measurements (Table 3). There was no systemic and proportional bias in the intra-grader measurements for VCDR and CDAR. In comparison, a statistically significant systemic bias (mean difference, -0.03) was found in the inter-grader VCDR measurement. Nevertheless, this systemic bias is minimal and clinically insignificant[19,38].

Harper et al[40] previously examined inter-grader agreement in VCDR measurement between 3 optometrists and 2 ophthalmologists using stereo optic disc photographs. They found fair to moderate agreement with Kappa value varying between 0.23 and 0.64. In comparison, almost perfect inter-observer agreement was obtained between the optometrist (grader A) and glaucoma specialist (grader B) in this study. This finding can be explained by the standardization of optic cup and optic disc demarcation definitions and strict adherence to a rigorous grading protocol in our study. In addition, both graders in our study underwent the same training set for standardization purpose before embarking on the actual grading task. Importance of standardization was also highlighted in previous studies where observer differences were attributed to difference in the definition of the cup and disc borders[16,40]. In view of the comparable measurements between trained optometrist and glaucoma specialist in this study, it is conceivable to allocate optic disc grading task to trained graders other than ophthalmologists to maximize productivity and efficiency in grading projects.

The present study suggested excellent agreement between SODA and the Wisconsin grading method based on ICC values. This could be explained in part by the fact that both methods adhered to the same optic disc and cup demarcation protocol (Wisconsin protocol). In addition, we also found that SODA measured VCDR slightly larger (a mean difference of 0.03, 95% LOA -0.09 to 0.16) as compared to the Wisconsin grading method. However, this discrepancy is within clinically acceptable range[19,38]. It should also be noted that the measurement difference between SODA and Wisconsin grading methods was not consistent throughout the range of VCDR measurements (presence of proportional bias). Specifically, such overestimation was slightly more prominent in the smaller VCDR range (<0.80) as shown by the regression line of the Bland Altman plot (Figure 5). Similarly, Parkin et al[20] also reported that measurements on monoscopic digital images yielded slightly larger VCDR as compared to stereoscopic images in smaller VCDR range. This may be explained in part by previous observations that there was greater variation in the determination of cup margins for discs of smaller CDRs[19,25,40,41], while smaller variation in discs of larger CDRs[14]. Moreover, such overestimation may be also due to less acute vessel bending at optic cup margins and shallower optic cupping in discs with smaller CDR as compared to discs with large CDR[37]. On the contrary, Morgan et al[18] and Hanson et al[42] found that stereoscopic optic disc assessment provided greater measurement of CDR as compared to monoscopic assessment. Such inconsistency may be due to different stereoscopic viewing methods and image sources employed in previous studies. Notably, Morgan et al[18] employed a software program that demarcated the inner border of neuroretinal rim based on the scleral rim position at the same depth level. In addition, stereo-images in their study were compressed and had poorer image quality.

Our findings indicate that SODA potentially permits a cost-effective, time-efficient tool for assessing optic nerve head from digital optic disc images. The grading procedures and protocol are relatively inexpensive to implement and the measurements require little time to perform. On average, measurement of an optic disc image requires approximately 2 minutes. Therefore, SODA is highly comparable to the standard Wisconsin grading method but with significantly better time efficiency. In addition, we also demonstrated that SODA is useful in grading images which were captured from different camera settings. With these features, SODA may have great potential for use in multi-centre and population-based studies which involve mass load of images from different study sites. Furthermore, with the advent of telemedicine technology, this software when equipped with a digital fundus camera; may also have the potential to be applied as a useful on-site glaucoma screening software in primary health care settings. Future research is required to substantiate the clinical value of SODA as a glaucoma screening tool in communities.

Subjects included in this study have a wide range of clinically measured VCDR and all image measurements were performed according to a standardized grading protocol. The high agreement reported in this study is unlikely to be biased by other ocular characteristics of that favor higher inter-observer agreement such as glaucomatous discs[14]. Therefore, our findings may be conceivably generalized to a daily practice of grading. In the agreement analysis between the SODA and Wisconsin grading methods, a single grader was employed; thus eliminating potential measurement errors caused by inter-grader variability. This study nevertheless has a few limitations. First, the grading of tilted discs using SODA may be more challenging than the stereo grading method due to the utilization of monoscopic images in this software[43,44]. Second, optic disc size was not controlled for in this study. This factor should be taken into consideration in future studies as disc size may potentially introduce bias in the judgment of CDR[45,46].

In conclusion, we described a new semi-automated software program (SODA) that gives excellent intra- and inter-grader reliability for optic disc parameter measurements. SODA compares favorably with the standard manual stereo grading method. This software therefore has great potential usage for assessing optic nerve head from digital retinal fundus photographs in clinical and population based studies.

References:

1. Wong TY, Loon SC, Saw SM. The epidemiology of age related eye diseases in Asia. *Br J Ophthalmol* 2006; **90**: 506-11.
2. Kingman S. Glaucoma is second leading cause of blindness globally. *Bull World Health Organ* 2004; **82**: 887-8.
3. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol* 2006; **90**: 262-7.
4. Girkin CA, McGwin G, Long C, DeLeon-Ortega J, Graf CM, Everett AW. Subjective and Objective Optic Nerve Assessment in African Americans and Whites. *Investigative Ophthalmology & Visual Science* 2004; **45**: 2272-8.
5. Greaney MJ, Hoffman DC, Garway-Heath DF, Nakla M, Coleman AL, Caprioli J. Comparison of Optic Nerve Imaging Methods to Distinguish Normal Eyes from Those with Glaucoma. *Investigative Ophthalmology & Visual Science* 2002; **43**: 140-5.
6. Kerrigan-Baumrind LA, Quigley HA, Pease ME, Kerrigan DF, Mitchell RS. Number of ganglion cells in glaucoma eyes compared with threshold visual field tests in the same persons. *Invest Ophthalmol Vis Sci* 2000; **41**: 741-8.
7. Xin D, Greenstein VC, Ritch R, Liebmann JM, De Moraes CG, Hood DC. A comparison of functional and structural measures for identifying progression of glaucoma. *Invest Ophthalmol Vis Sci* 2011; **52**: 519-26.
8. Foster PJ, Oen FT, Machin D, Ng TP, Devereux JG, Johnson GJ, Khaw PT, Seah SK. The prevalence of glaucoma in Chinese residents of Singapore: a cross-sectional population survey of the Tanjong Pagar district. *Arch Ophthalmol* 2000; **118**: 1105-11.
9. Raychaudhuri A, Lahiri SK, Bandyopadhyay M, Foster PJ, Reeves BC, Johnson GJ. A population based survey of the prevalence and types of glaucoma in rural West Bengal: the West Bengal Glaucoma Study. *Br J Ophthalmol* 2005; **89**: 1559-64.
10. Shen SY, Wong TY, Foster PJ, Loo JL, Rosman M, Loon SC, Wong WL, Saw SM, Aung T. The prevalence and types of glaucoma in malay people: the Singapore Malay eye study. *Invest Ophthalmol Vis Sci* 2008; **49**: 3846-51.
11. Casson RJ, Newland HS, Muecke J, McGovern S, Abraham L, Shein WK, Selva D, Aung T. Prevalence of glaucoma in rural Myanmar: the Meiktila Eye Study. *Br J Ophthalmol* 2007; **91**: 710-4.
12. Palimkar A, Khandekar R, Venkataraman V. Prevalence and distribution of glaucoma in central India (Glaucoma Survey 2001). *Indian J Ophthalmol* 2008; **56**: 57-62.
13. Sung VC, Bhan A, Vernon SA. Agreement in assessing optic discs with a digital stereoscopic optic disc camera (Discam) and Heidelberg retina tomograph. *Br J Ophthalmol* 2002; **86**: 196-202.
14. Tielsch JM, Katz J, Quigley HA, Miller NR, Sommer A. Intraobserver and interobserver agreement in measurement of optic disc characteristics. *Ophthalmology* 1988; **95**: 350-6.
15. Lichter PR. Variability of expert observers in evaluating the optic disc. *Trans Am Ophthalmol Soc* 1976; **74**: 532-72.
16. Garway-Heath DF, Poinoosawmy D, Wollstein G, Viswanathan A, Kamal D, Fontana L, Hitchings RA. Inter- and intraobserver variation in the analysis of optic disc images: comparison of the Heidelberg retina tomograph and computer assisted planimetry. *Br J Ophthalmol* 1999; **83**: 664-9.
17. Garway-Heath DF, Hitchings RA. Quantitative evaluation of the optic nerve head in early glaucoma. *Br J Ophthalmol* 1998; **82**: 352-61.

18. Morgan JE, Sheen NJ, North RV, Choong Y, Ansari E. Digital imaging of the optic nerve head: monoscopic and stereoscopic analysis. *Br J Ophthalmol* 2005; **89**: 879-84.
19. 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.
20. Parkin B, Shuttleworth G, Costen M, Davison C. A comparison of stereoscopic and monoscopic evaluation of optic disc topography using a digital optic disc stereo camera. *Br J Ophthalmol* 2001; **85**: 1347-51.
21. Stone RA, Ying GS, Pearson DJ, Bansal M, Puri M, Miller E, Alexander J, Piltz-Seymour J, Nyberg W, Maguire MG, Eledath J, Sawhney H. Utility of digital stereo images for optic disc evaluation. *Invest Ophthalmol Vis Sci* 2010; **51**: 5667-74.
22. Klein BE, Magli YL, Richie KA, Moss SE, Meuer SM, Klein R. Quantitation of optic disc cupping. *Ophthalmology* 1985; **92**: 1654-6.
23. Klein BE, Klein R, Lee KE, Hoyer CJ. Does the intraocular pressure effect on optic disc cupping differ by age? *Trans Am Ophthalmol Soc* 2006; **104**: 143-8.
24. Mitchell P, Smith W, Attebo K, Healey PR. Prevalence of open-angle glaucoma in Australia. The Blue Mountains Eye Study. *Ophthalmology* 1996; **103**: 1661-9.
25. Feuer WJ, Parrish RK, 2nd, Schiffman JC, Anderson DR, Budenz DL, Wells MC, Hess DJ, Kass MA, Gordon MO. The Ocular Hypertension Treatment Study: reproducibility of cup/disk ratio measurements over time at an optic disc reading center. *Am J Ophthalmol* 2002; **133**: 19-28.
26. Foong AW, Saw SM, Loo JL, Shen S, Loon SC, Rosman M, Aung T, Tan DT, Tai ES, Wong TY. Rationale and methodology for a population-based study of eye diseases in Malay people: The Singapore Malay eye study (SiMES). *Ophthalmic Epidemiol* 2007; **14**: 25-35.
27. Lavanya R, Jeganathan VS, Zheng Y, Raju P, Cheung N, Tai ES, Wang JJ, Lamoureux E, Mitchell P, Young TL, Cajucom-Uy H, Foster PJ, Aung T, Saw SM, Wong TY. Methodology of the Singapore Indian Chinese Cohort (SICC) eye study: quantifying ethnic variations in the epidemiology of eye diseases in Asians. *Ophthalmic Epidemiol* 2009; **16**: 325-36.
28. Mitchell P, Smith W, Attebo K, Wang JJ. Prevalence of age-related maculopathy in Australia. The Blue Mountains Eye Study. *Ophthalmology* 1995; **102**: 1450-60.
29. Pakter HM, Fuchs SC, Maestri MK, Moreira LB, Dei Ricardi LM, Pamplona VF, Oliveira MM, Fuchs FD. Computer-assisted methods to evaluate retinal vascular caliber: what are they measuring? *Invest Ophthalmol Vis Sci* 2011; **52**: 810-5.
30. Fleiss JL, Cohen J. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educational and Psychological Measurement* 1973; **33**: 613-9.
31. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307-10.
32. Bland JM, Altman DG. Agreed statistics: measurement method comparison. *Anesthesiology* 2012; **116**: 182-5.
33. Sherry LM, Wang JJ, Rochtchina E, Wong T, Klein R, Hubbard L, Mitchell P. Reliability of computer-assisted retinal vessel measurementin a population. *Clin Experiment Ophthalmol* 2002; **30**: 179-82.
34. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet* 1995; **346**: 1085-7.
35. Li EY, Mohamed S, Leung CK, Rao SK, Cheng AC, Cheung CY, Lam DS. Agreement among 3 methods to measure corneal thickness: ultrasound pachymetry, Orbscan II, and Visante anterior segment optical coherence tomography. *Ophthalmology* 2007; **114**: 1842-7.
36. Caprioli J. Clinical evaluation of the optic nerve in glaucoma. *Trans Am Ophthalmol Soc* 1994; **92**: 589-641.

37. Jonas JB, Budde WM, Panda-Jonas S. Ophthalmoscopic evaluation of the optic nerve head. *Surv Ophthalmol* 1999; **43**: 293-320.
38. Varma R, Steinmann WC, Scott IU. Expert agreement in evaluating the optic disc for glaucoma. *Ophthalmology* 1992; **99**: 215-21.
39. Abrams LS, Scott IU, Spaeth GL, Quigley HA, Varma R. Agreement among optometrists, ophthalmologists, and residents in evaluating the optic disc for glaucoma. *Ophthalmology* 1994; **101**: 1662-7.
40. Harper R, Reeves B, Smith G. Observer variability in optic disc assessment: implications for glaucoma shared care. *Ophthalmic Physiol Opt* 2000; **20**: 265-73.
41. Varma R, Spaeth GL, Steinmann WC, Katz LJ. Agreement between clinicians and an image analyzer in estimating cup-to-disc ratios. *Arch Ophthalmol* 1989; **107**: 526-9.
42. Hanson S, Krishnan SK, Phillips J. Observer experience and Cup:Disc ratio assessment. *Optom Vis Sci* 2001; **78**: 701-5.
43. Park CY, Kim YT, Kee C. Evaluation of the influence of tilt of optic disc on the measurement of optic disc variables obtained by optical coherence tomography and confocal scanning laser ophthalmoscopy. *J Glaucoma* 2005; **14**: 210-4.
44. Hyung SM, Kim DM, Hong C, Youn DH. Optic disc of the myopic eye: relationship between refractive errors and morphometric characteristics. *Korean J Ophthalmol* 1992; **6**: 32-5.
45. Crowston JG, Hopley CR, Healey PR, Lee A, Mitchell P. The effect of optic disc diameter on vertical cup to disc ratio percentiles in a population based cohort: the Blue Mountains Eye Study. *Br J Ophthalmol* 2004; **88**: 766-70.
46. Hoffmann EM, Zangwill LM, Crowston JG, Weinreb RN. Optic disk size and glaucoma. *Surv Ophthalmol* 2007; **52**: 32-49.

Figure Legends

Figure 1. Illustration of Singapore Optic Disc Assessment (SODA) Software Program: A) Concentric Enlargement of Optic Disc Cupping, B) Optic Disc with Inferior Rim Notching and Beta-type Peripapillary Atrophy.

Figure 2. Flow diagram of the study design.

Figure 3. Bland Altman plots of intra grader measurements (N = 328): A) Vertical cup-to-disc ratio, B) Cup-to-disc area ratio. Pink dashed line represents regression line of difference between measurements.

Figure 4. Bland Altman plots of inter-grader measurements (N = 328): A) Vertical cup-to-disc ratio, B) Cup-to-disc area ratio. Pink dashed line represents regression line of difference between measurements.

Figure 5. Bland Altman plots of vertical cup-to-disc ratio measurement between SODA and the standard Wisconsin grading method (N = 85). The difference was calculated by the SODA measurement *minus* the measurement from the standard Wisconsin grading method. Pink dashed line represents regression line of difference between measurements.

Table 1: Characteristics of Study Subjects

	Malays (n=103)	Indians (n=112)	Chinese (n=113)	Whites (n=85)
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
Age, years	56 (9.3)	56 (8.9)	58 (8.8)	68 (10.4)
Female	52 (50.5%)	64 (57.1%)	65 (57.5%)	36 (42.4%)
Spherical Equivalent, dioptres	-0.02 (1.29)	0.33 (1.48)	-0.45 (2.38)	0.51 (2.39)
Clinical Ophthalmoscopy:	**N (%)**	**N (%)**	**N (%)**	**N (%)**
CDR 0.1 to 0.39	32 (31%)	38 (34%)	34 (30%)	19 (22%)
CDR 0.4 to 0.59	44 (43%)	49(44%)	52 (46%)	29 (34%)
CDR ≥ 0.6	27 (26%)	25 (22%)	27 (24%)	37 (44%)
Presence of Peripapillary Atrophy (Beta type):	**N (%)**	**N (%)**	**N (%)**	**N (%)**
	36 (35%)	32 (32%)	32 (32%)	33 (39%)
Presence of Tilted Disc:	**N (%)**	**N (%)**	**N (%)**	**N (%)**
	7 (7%)	5 (4%)	15 (13%)	15 (18%)

Data are expressed as number (percentage), except for age and spherical equivalent, which are expressed as mean (SD).

Table 2: Summary of Optic Disc Parameter Measurements using SODA (SEED studies sample)

	Grader A, 1st measurement Mean (SD)	Grader A, 2nd Measurement Mean (SD)	Grader B measurement Mean (SD)
Vertical Cup-to-Disc Ratio	0.56 (0.12)	0.56 (0.12)	0.59 (0.12)
Vertical Cup Diameter, mm	1.00 (0.28)	0.98 (0.28)	1.07 (0.28)
Vertical Disc Diameter, mm	1.76 (0.20)	1.73 (0.20)	1.78 (0.20)
Cup-to-Disc Area Ratio	0.37 (0.12)	0.38 (0.12)	0.38 (0.12)
Cup Area, mm^2	0.87 (0.40)	0.85 (0.38)	0.91 (0.40)
Disc Area, mm^2	2.27 (0.49)	2.17 (0.48)	2.35 (0.51)

Table 3: Summary of Intra-, Inter-grader Reliability and Agreement Analysis between SODA and Wisconsin Grading Methods

	Optic Disc Parameters	ICC (95% CI)	Mean difference (95% LOA)	P Value*	Pearson's correlation coefficient, r	P Value**
Intra-grader Reliability using SODA†	Vertical Cup-to-Disc Ratio	0.87 (0.84 to 0.89)	0.00 (-0.12 to 0.13)	0.802	-0.003	0.956
	Vertical Cup Diameter, mm	0.91 (0.89 to 0.93)	-0.02 (-0.25 to 0.21)	0.005	-0.039	0.495
	Vertical Disc Diameter, mm	0.93 (0.86 to 0.96)	-0.04 (-0.17 to 0.10)	<0.001	-0.043	0.443
	Cup-to-Disc Area Ratio	0.88 (0.85 to 0.90)	0.01 (-0.11 to 0.12)	0.082	-0.017	0.757
	Cup Area, mm^2	0.94 (0.92 to 0.95)	-0.03 (-0.30 to 0.24)	<0.001	-0.010	0.077
	Disc Area, mm^2	0.94 (0.81 to 0.97)	-0.10 (-0.37 to 0.17)	<0.001	-0.091	0.098
Inter-grader Reliability using SODA‡	Vertical Cup-to-Disc Ratio	0.84 (0.73 to 0.90)	-0.03 (-0.15 to 0.09)	<0.001	0.079	0.153
	Vertical Cup Diameter, mm	0.91 (0.77 to 0.95)	-0.07 (-0.27 to 0.14)	<0.001	0.012	0.827
	Vertical Disc Diameter, mm	0.93 (0.91 to 0.95)	-0.02 (-0.16 to 0.13)	<0.001	0.021	0.699
	Cup-to-Disc Area Ratio	0.82 (0.78 to 0.85)	0.00 (-0.14 to 0.14)	0.249	-0.007	0.757
	Cup Area, mm^2	0.93 (0.91 to 0.95)	-0.04 (-0.32 to 0.24)	<0.001	-0.046	0.408
	Disc Area, mm^2	0.94 (0.89 to 0.96)	-0.08 (-0.38 to 0.23)	<0.001	-0.082	0.139
Agreement between SODA and Wisconsin Grading Method performed	Vertical Cup-to-Disc Ratio	0.94 (0.88 to 0.97)	0.03 (-0.09 to 0.16)	<0.001	-0.331	0.002

by same grader

LOA, Limits of Agreement; ICC, Intraclass Correlation Coefficient; CI, Confidence Interval.
†Mean difference was determined from the 2nd time measurement minus the 1st time measurement.
‡Mean difference was determined from Grader A measurement minus Grader B measurement.
*P value of one sample t-tests (comparing between mean difference and zero value) to indicate presence of systemic bias
**P value of Pearson's correlation coefficients of regression line to indicate presence of proportional bias

Figure 1A

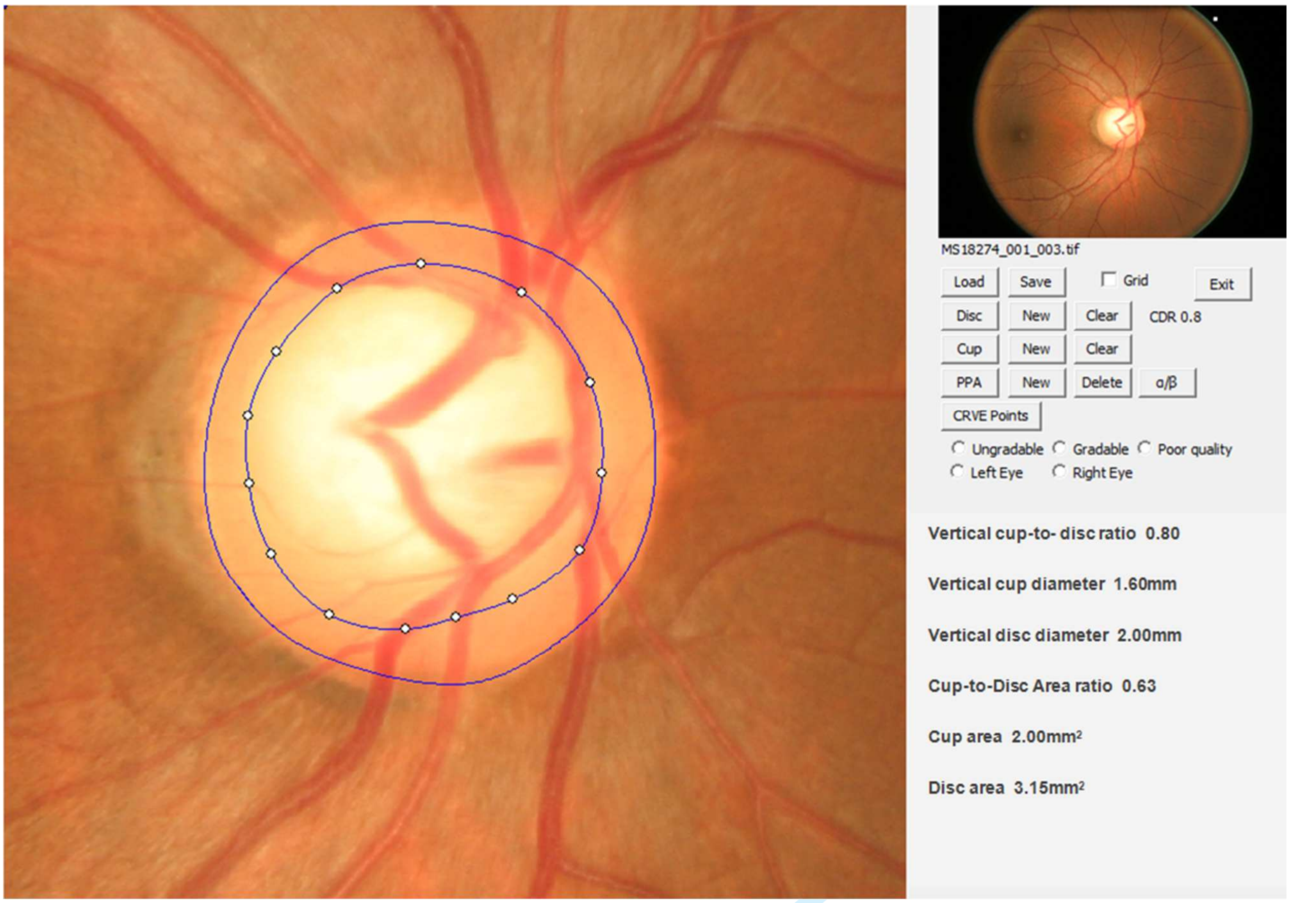

MS18274_001_003.tif
Load
Save
Grid
Exit
Disc
New
Clear
CDR 0.8
Cup
New
Clear
PPA
New
Delete
α/β
CRVE Points
Ungradable
Gradable
Poor quality
Left Eye
Right Eye
Vertical cup-to- disc ratio 0.80
Vertical cup diameter 1.60mm
Vertical disc diameter 2.00mm
Cup-to-Disc Area ratio 0.63
Cup area 2.00mm²
Disc area 3.15mm²

Figure 1B

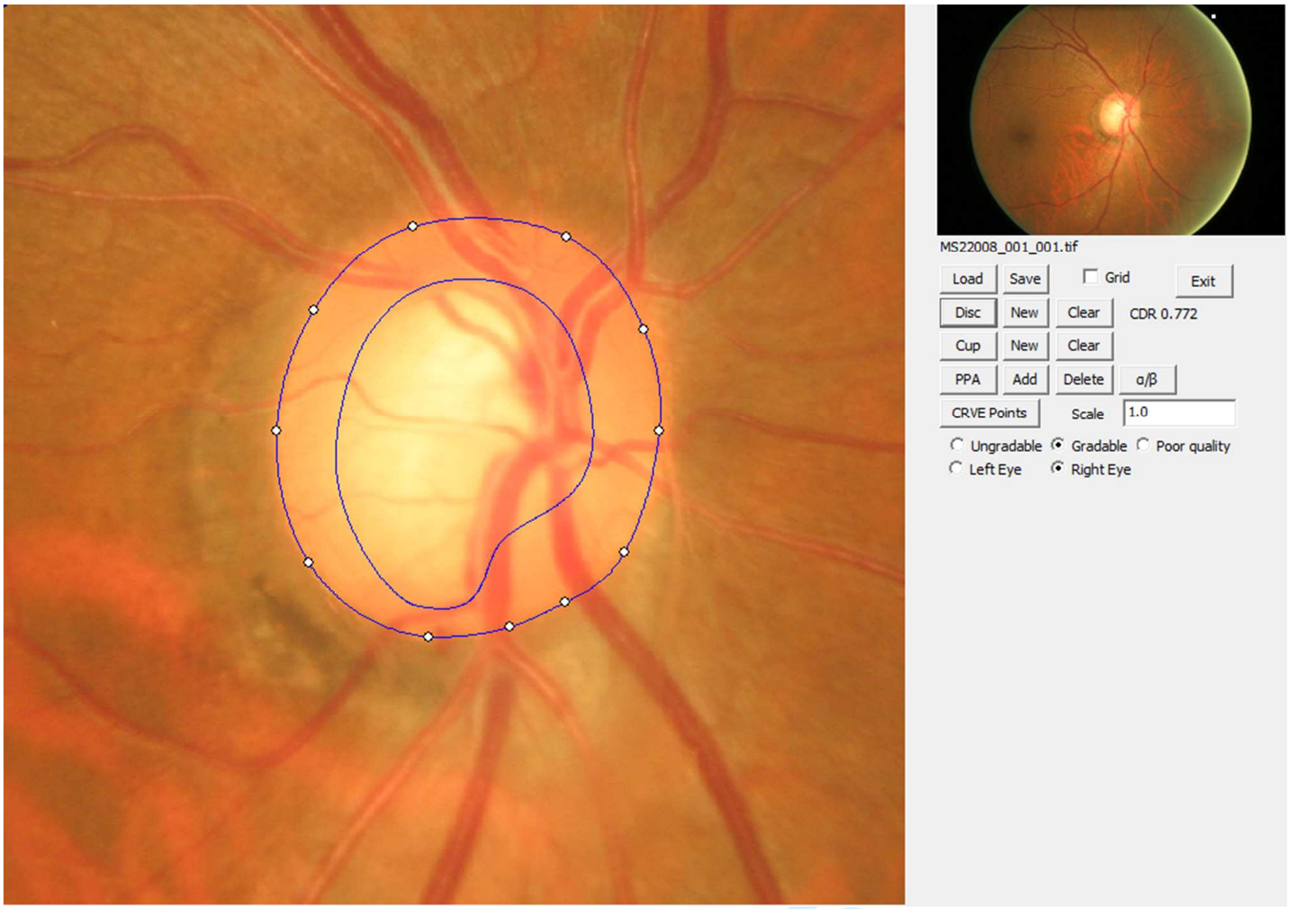

MS22008_001_001.tif
Load
Save
Grid
Exit
Disc
New
Clear
CDR 0.772
Cup
New
Clear
PPA
Add
Delete
α/β
CRVE Points
Scale
1.0
Ungradable
Gradable
Poor quality
Left Eye
Right Eye

Figure 2 :

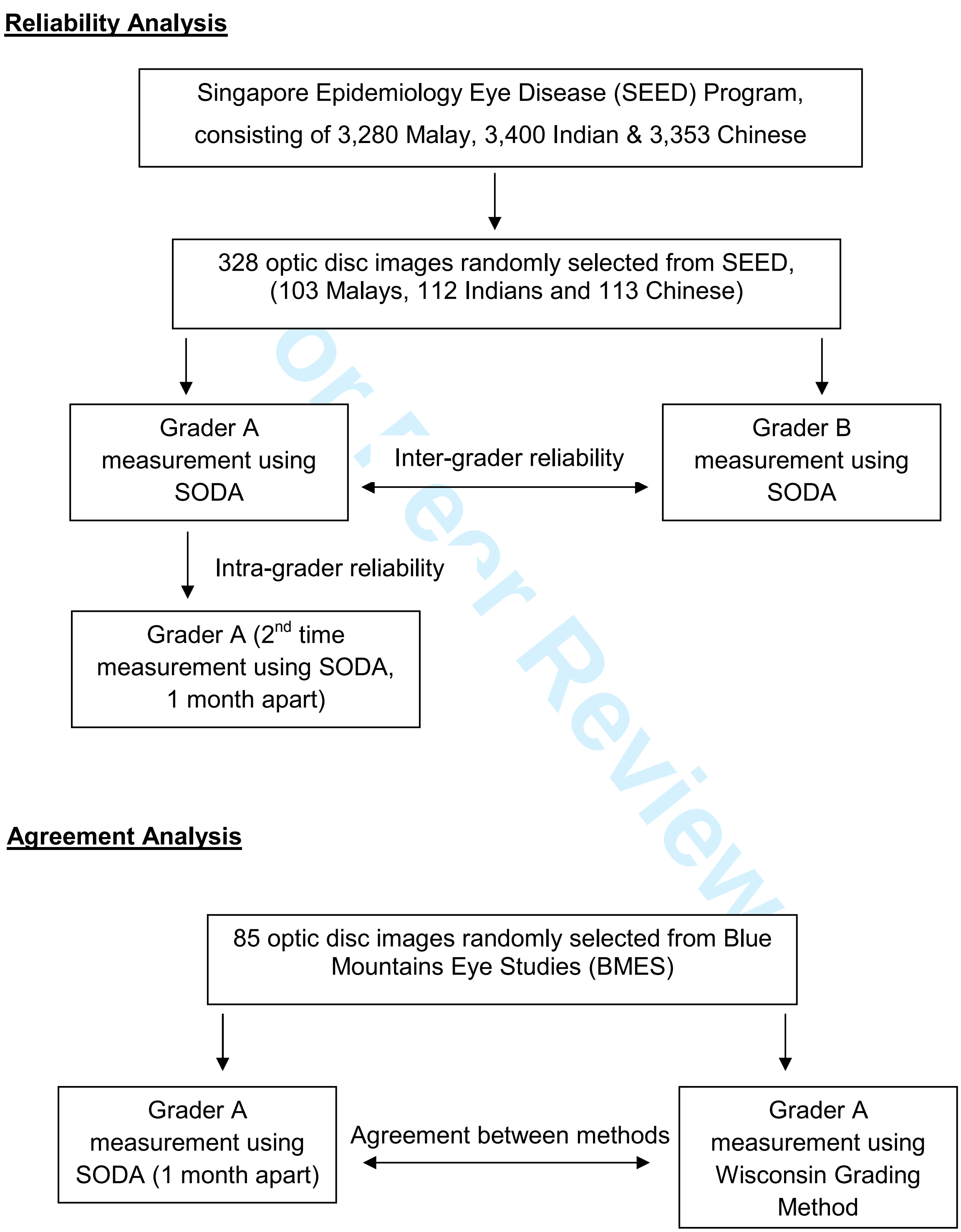

Reliability Analysis
Singapore Epidemiology Eye Disease (SEED) Program, consisting of 3,280 Malay, 3,400 Indian & 3,353 Chinese
328 optic disc images randomly selected from SEED, (103 Malays, 112 Indians and 113 Chinese)
Grader A measurement using SODA
Inter-grader reliability
Grader B measurement using SODA
Intra-grader reliability
Grader A (2nd time measurement using SODA, 1 month apart)
Agreement Analysis
85 optic disc images randomly selected from Blue Mountains Eye Studies (BMES)
Grader A measurement using SODA (1 month apart)
Agreement between methods
Grader A measurement using Wisconsin Grading Method

Figure 3A

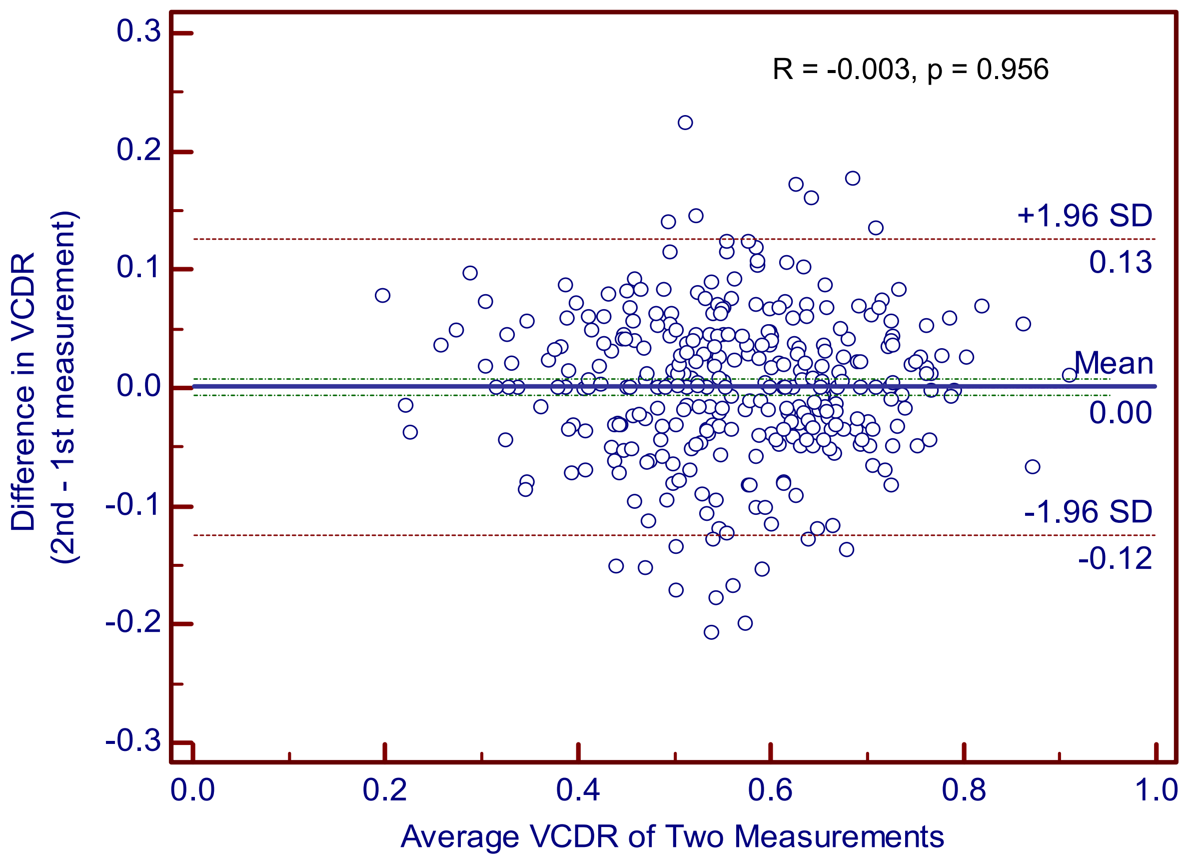

R = -0.003, p = 0.956
+1.96 SD
0.13
Mean
0.00
-1.96 SD
-0.12
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
Difference in VCDR
(2nd - 1st measurement)
0.0
0.2
0.4
0.6
0.8
1.0
Average VCDR of Two Measurements

Figure 3B

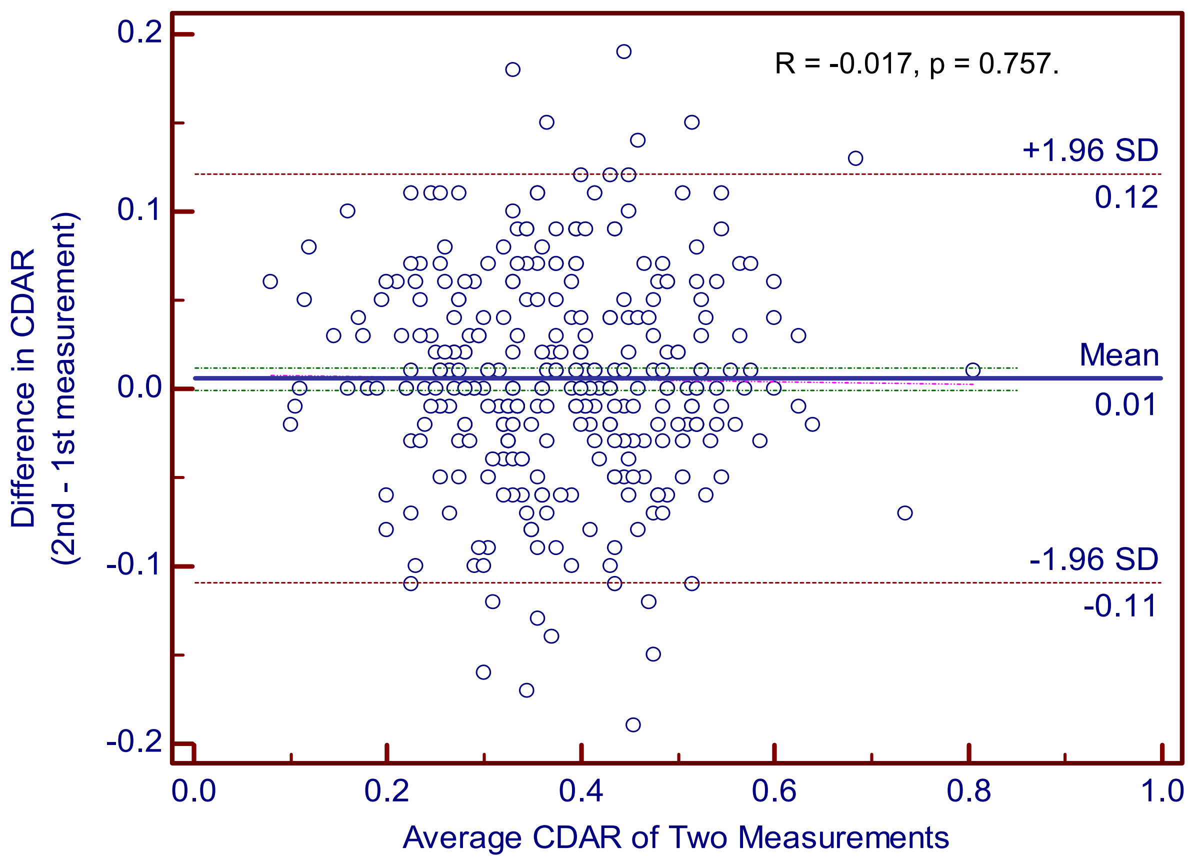

R = -0.017, p = 0.757.
+1.96 SD
0.12
Mean
0.01
-1.96 SD
-0.11
0.2
0.1
0.0
-0.1
-0.2
Difference in CDAR
(2nd - 1st measurement)
0.0
0.2
0.4
0.6
0.8
1.0
Average CDAR of Two Measurements

Figure 4A

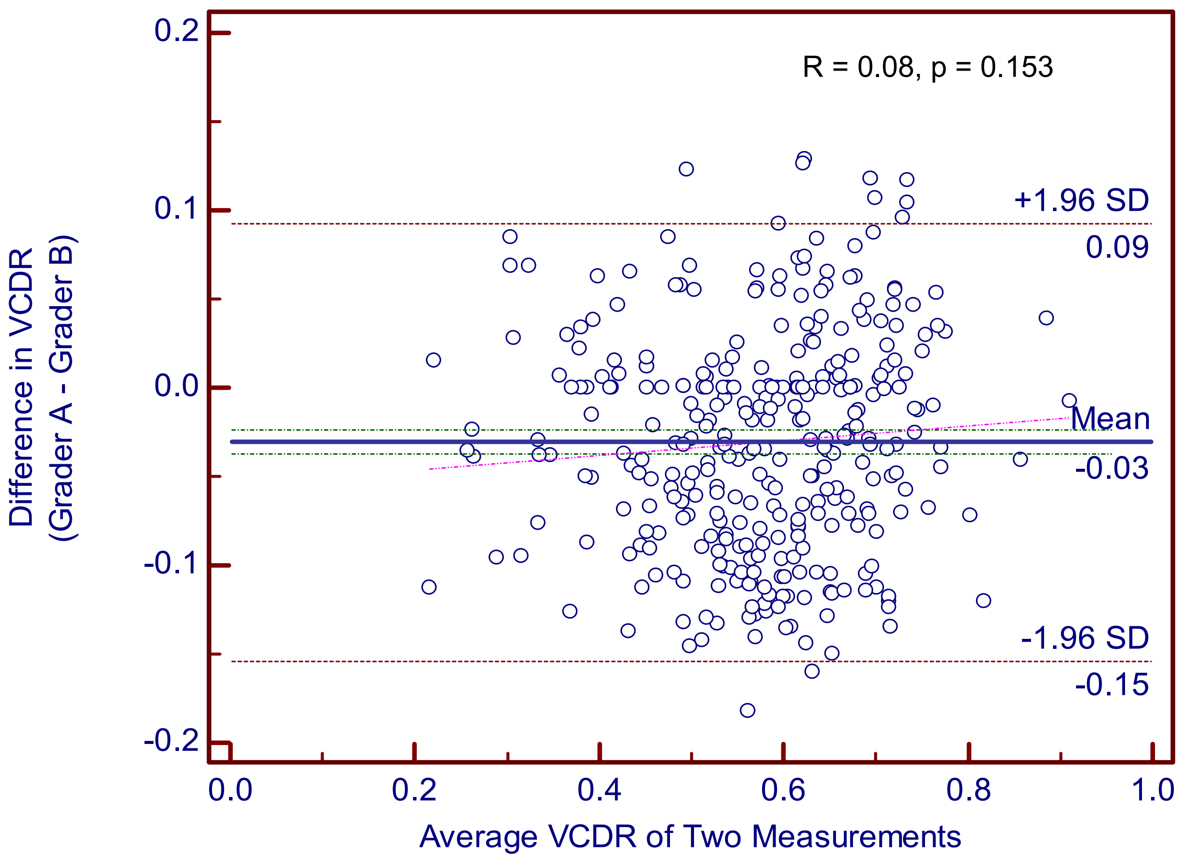

R = 0.08, p = 0.153
+1.96 SD
0.09
Mean
-0.03
-1.96 SD
-0.15
0.2
0.1
0.0
-0.1
-0.2
Difference in VCDR
(Grader A - Grader B)
0.0
0.2
0.4
0.6
0.8
1.0
Average VCDR of Two Measurements

Figure 4B

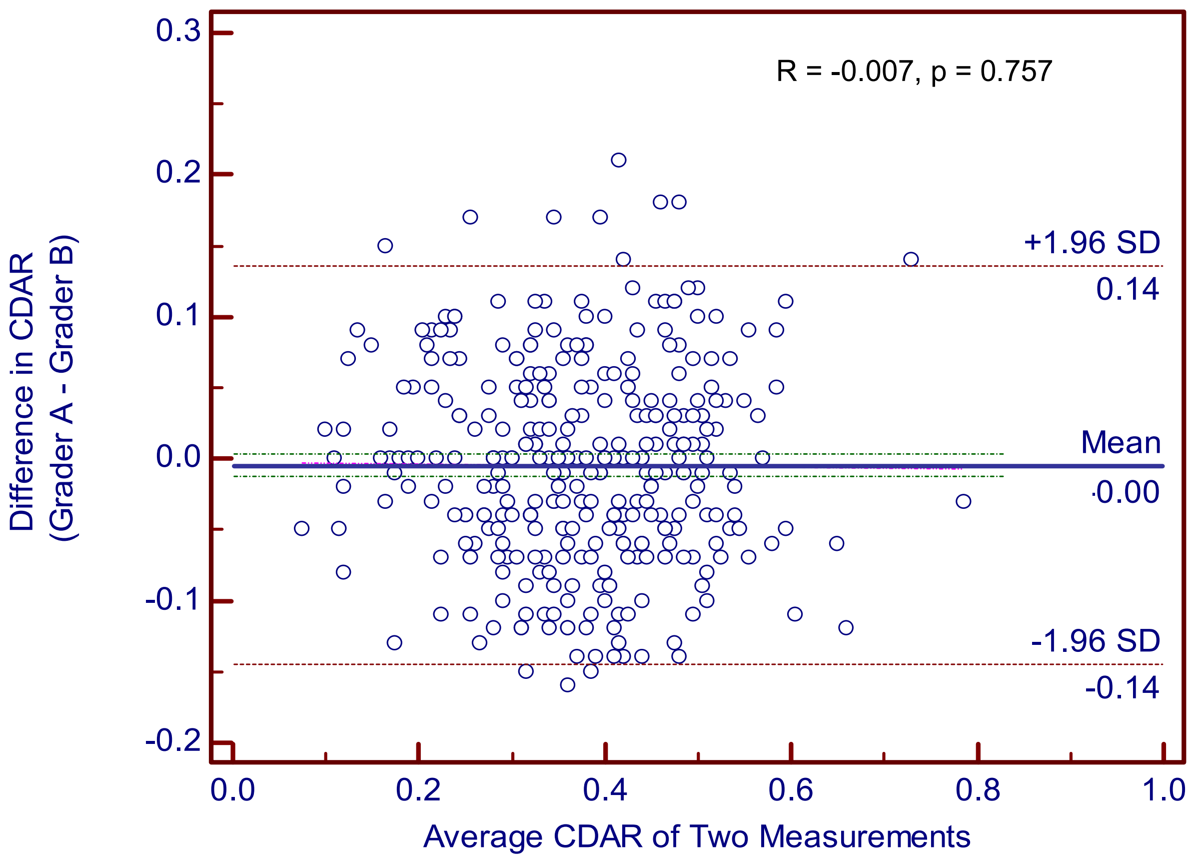
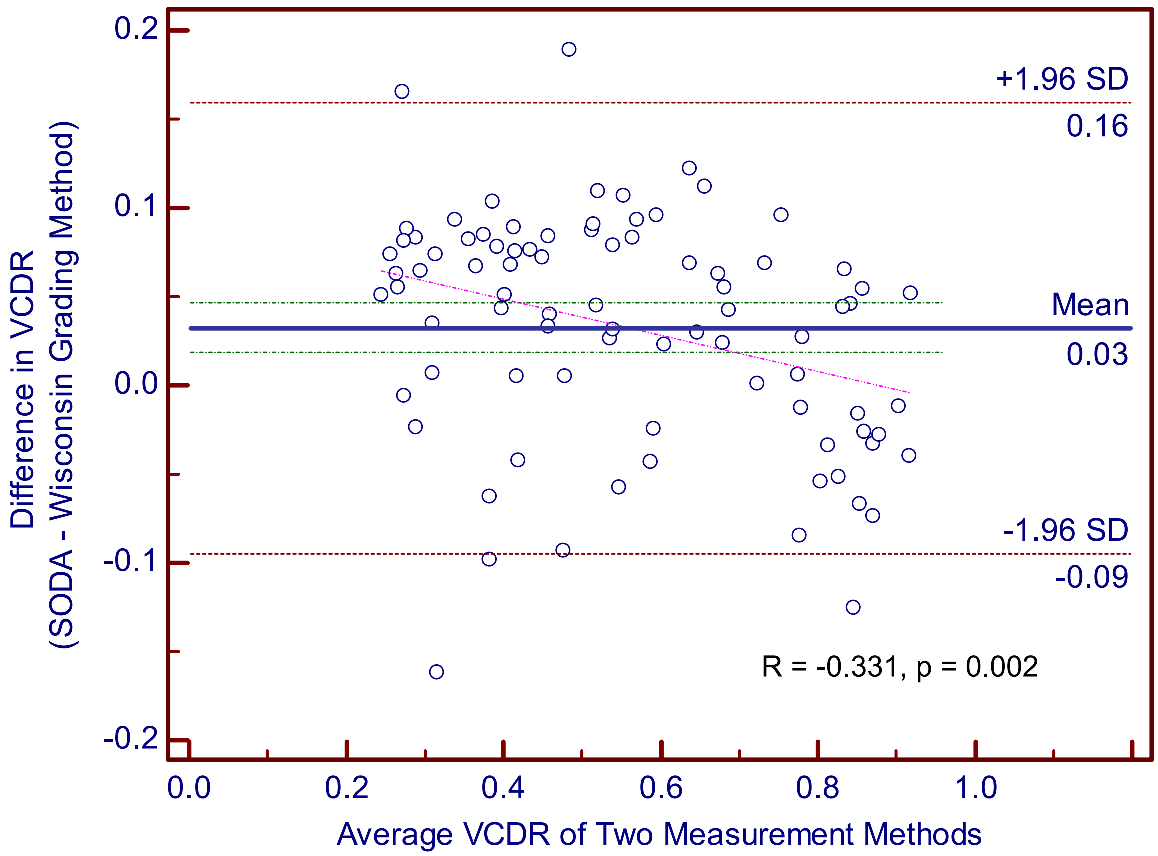

R = -0.007, p = 0.757
+1.96 SD
0.14
Mean
-0.00
-1.96 SD
-0.14
Difference in CDAR
(Grader A - Grader B)
Average CDAR of Two Measurements
0.3
0.2
0.1
0.0
-0.1
-0.2
0.0
0.2
0.4
0.6
0.8
1.0

Figure 5


+1.96 SD
0.16
Mean
0.03
-1.96 SD
-0.09
R = -0.331, p = 0.002
0.2
0.1
0.0
-0.1
-0.2
Difference in VCDR
(SODA - Wisconsin Grading Method)
0.0
0.2
0.4
0.6
0.8
1.0
Average VCDR of Two Measurement Methods