Quality assurance in diabetic retinal screening in South Africa

S Cook,[1] MB BCh, FCS (Ophth) (SA); R **T Staff**,[2,3] PhD; **K A Goatman**,[4] PhD; **J A Olson**,[4,5] MD, FRCPE, FRCOphth; and the Scottish Diabetic Retinopathy Screening collaborative

[1] *The Eye Centre, East London, Eastern Cape, South Africa, and Department of Ophthalmology, Faculty of Health Sciences, Walter Sisulu University, Umtata, Eastern Cape, South Africa*
[2] *Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen, Scotland, UK*
[3] *Aberdeen Royal Infirmary, NHS Grampian, Scotland, UK*
[4] *School of Medicine and Dentistry, University of Aberdeen, Scotland, UK*
[5] *Retinal Screening, Aberdeen Royal Infirmary, NHS Grampian, Scotland, UK*

Corresponding author: S Cook (dr@eyecentre.co.za)

Background. Diabetic retinopathy (DR) is an important biomarker for microvascular disease and blindness. Digital fundus photography is a cost-effective way of screening for DR. Access to DR screening is difficult for many South Africans with diabetes.

Objective. To perform external quality assurance (EQA) on graders registered in the Ophthalmological Society of South Africa DR screening programme.

Methods. Graders registered on the South African (SA) Diabetic Register website were invited to participate in the study. The Scottish EQA software system was used to enable on-line grading of 100 retinal photographs. Expert National Health Service graders provided the consensus expert grading for the image set.

Results. Two hundred and sixty-one participants completed the EQA process, including nine ophthalmologists, 243 optometrists, and nine other graders. A wide range of outcomes were demonstrated, with a mean sensitivity of 0.905 (range 0.286 - 1.000) and mean specificity of 0.507 (0.000 - 0.935). The mean diagnostic odds ratio was calculated to be 12.3 (range 0.147 - 148.2).

Conclusions. This is the first quality assurance study conducted with SA healthcare professionals. The outcomes are of interest to all stakeholders dealing with the diabetes epidemic. The disparity in grader performance indicates room for improvement. The results demonstrate a high referral rate to ophthalmology, suggesting that on average graders are performing safely, but with a high number of inappropriate referrals.

The incidence of diabetes is increasing globally. Africa is no exception: the number of adults with diabetes is expected to almost double by 2030 to 23.9 million.[1] Diabetes is clearly becoming a pressing public health problem for Africa, for which effective interventions are required in the near future to avert the anticipated health burden. One such intervention is the introduction of a diabetic retinal screening programme for the early detection of diabetic retinopathy (DR) and the prevention of blindness.

Studies estimating the impact of screening programmes in Europe have concluded that they achieved worthwhile health gains and substantial reductions in the incidence of new blindness due to DR.[2,3] The introduction of a European-style screening programme in the African context is unlikely in the near future. However, raising current screening expertise and fully exploiting the available technology has the potential for significant health gains.

DR screening in South Africa (SA) is done on an *ad hoc* and opportunistic basis by a spectrum of healthcare providers. These providers have invested in fundus camera technology capable of producing high-resolution retinal images. This technology is key to established screening programmes, and is core technology for national screening approaches elsewhere. The ability to acquire and record high-quality retinal images is, however, only the first step: effective interpretation is essential to any screening process. How that screening expertise is disseminated and maintained is a challenge for national and professional bodies that wish to address the increasing health burden.

The Ophthalmological Society of South Africa (OSSA) has developed a framework for a national Diabetic Retinopathy Screening Programme that can accommodate the many sectors in the healthcare environment. This includes traditional teaching methods such as lecturing and publication. A key part of this framework is a quality assurance process that will encourage ongoing learning and the monitoring of performance.

The Scottish External Quality Assurance (EQA) process is an internet-based assessment tool for the Scottish Diabetic Retinopathy Screening collaborative. The programme has been in place since 2007. It includes a compulsory twice-yearly procedure that involves all screeners working for the National Health Service in Scotland. Once completed and assessed, feature-based feedback is given to individual screeners and the lead person responsible at each screening site. The process involves each screener grading 100 retinal images. These grades are compared with a reference standard generated from the consensus of the expert graders, who are experienced ophthalmologists with a high level of proficiency in assessing retinal images. Longitudinal analysis of screener performance has shown that it significantly improves performance for all grades of screener.[4] It has enabled the service to assess individual screeners and sites, to identify strengths and weaknesses, and to improve quality.

Using this software, the current study assessed the performance of volunteer screeners in SA, by comparison with a group of expert graders from the Scottish service. We aimed to assess the feasibility of the EQA process in SA; to compare the performance of

screening groups; and to identify potential improvements in the screening practice of individuals with a view to future personal and group feedback and training.

Methods

Screener recruitment

The Diabetic Retinopathy Screening Programme was advertised and promoted by a variety of means: conference presentations, personal contact, posters, and publications in the *SAMJ* and the Diabetic Register website (www.diabeticregister. co.za). The Diabetic Register is a closed user group site for healthcare professionals. Individuals who expressed an interest were invited to participate in the EQA process. The minimum requirements to take part in the study were access to the internet, an e-mail account, and an occupation with the requirement or opportunity for retinal screening. Those who indicated interest were sent a link to a web-based survey to establish their professional characteristics, e.g. post type, experience and workload (www.surveymonkey.com/s/B6B3BNC). Once the survey had been completed, a further link was sent explaining how to access the EQA site and carry out the EQA process. The site was open for data collection between November 2013 and January 2014. Technical support in grading was provided by a dedicated employee of the Eye Centre in East London, Eastern Cape, SA. Queries that could not be resolved locally were dealt with by the EQA experts in Scotland.

Running in parallel to this process, expert graders from the Scottish diabetic retinal screening service were asked to participate so as to provide an external consensus reference standard. These graders have all been part of the screening service for more than 5 years, have been tested regularly via the EQA process, have significant screening workloads, and have been shown to be high and consistent performers.

Image selection

The retinal images used in this study were selected by one of the authors (SC). Anonymised images were selected from the Eye Centre fundus photo database. There was no pre-selection for quality or level of retinopathy.

The EQA process

The EQA web-based software closely matches the feature-based grading system used in Scotland. The interface (Fig. 1) is compatible with all popular web browsers and consists of an image display on the left, together with controls for contrast, brightness, zoom and red-free colour display, a ruler for measuring the size of features, and a feature grading panel on the right-hand side. The number of times these controls were used was recorded for each screener, along with the time taken for grading all the images. Each grader was presented with the images in a different random order. Screeners were also allowed to use a 'sandbox' feature that enabled them
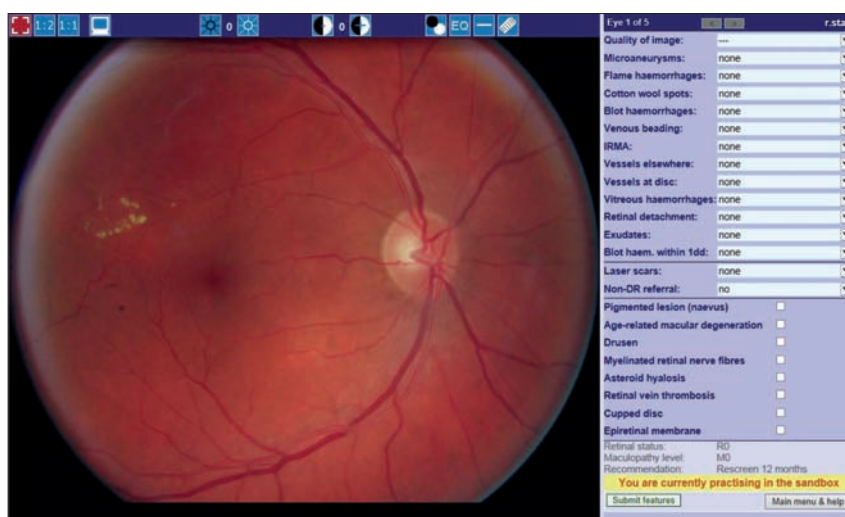

Fig. 1. The External Quality Assurance software interface.

Table 1. A summary of the Scottish Diabetic Retinopathy Grading Scheme

Grading	Description	Outcome
R0 (no visible retinopathy)	No visible diabetic retinopathy anywhere	Re-screen in 12 months
R1 (mild)	At least one dot, blot or flame haemorrhage, microaneurysm, exudate or cotton-wool spot anywhere	Re-screen in 12 months
R2 (observable background)	Four or more blot haemorrhages (>AH standard photograph 2a) in one hemi-field only (where the inferior and superior hemi-fields are delineated by a line passing through the centres of the fovea and optic disc)	Re-screen in 6 months
R3 (referable background)	Any of the following features: (*i*) four or more blot haemorrhages (>AH standard photograph 2a) in the inferior and superior hemi-fields; (*ii*) venous beading (>AH standard photograph 6a); (*iii*) intraretinal microvascular anomalies (>AH standard photograph 8a)	Refer to ophthalmology
R4 (proliferative)	Any of the following features: (*i*) active new vessels; (*ii*) vitreous haemorrhage	Refer to ophthalmology
R6 (inadequate)	Insufficient clarity or field of view	Slit-lamp examination
M1 (observable maculopathy)	Exudate within a radius of >1 but <2 disc diameters of the centre of the fovea	Re-screen in 6 months
M2 (referable maculopathy)	Any blot haemorrhage or exudate within a radius of 1 disc diameter of the centre of the fovea	Refer to ophthalmology

AH = Airlie House standard photography scheme.[9]

to become familiar with the system by grading example images prior to grading the test images.

The software implements the Scottish grading scheme (http://www.ndrs-wp.scot.nhs.uk), summarised in Table 1. The retinopathy grades are derived automatically from the features the grader selects. There are eight possible grading outcomes: four

of these outcomes require referral (M2, R3, R4 and R6), and two indicate more frequent review with a 6-month interval (M1 and R2). The remaining two categories (R0 and R1) result in re-screening in 12 months.

Results

The promotion process generated interest from 398 individuals. These individuals

accessed the survey site, filled in the questionnaire and gave permission for their data to be used. Two hundred and sixty-one participants gave all the information requested, and went on to register on the EQA site and complete the process. The characteristics of this group are shown in Table 2.

The nine expert graders achieved a consensus on the grading of 90 out of the 100 images. The responses to these images by each participant were used to assess the participant's performance. According to the expert external screeners, the 90 images were classified as follows: R0 – 22, R1 – 38, M1 – 2, R2 – 0, R6 – 2, M2 – 13, R3 – 4, R4 – 9.

Fig. 2 shows the sensitivity and specificity of each participant in a receiver operating characteristic (ROC) diagram. Each circle represents a single participant. The squares represent the performance of the external expert graders. All graders were assessed using the consensus grades as the reference standard. The top left-hand corner indicates the best screener performance. The expert graders are shown as dark blue squares. The SA graders are represented as lighter blue circles. Note that the expert graders would be expected to perform better, as they each contributed to the standard.

There is a significant variation in the performance across all graders. For example, at one extreme, a grader detects just over 20% of cases, which would normally be referred to an eye clinic for ophthalmology assessment; while at the other, two graders have 0% specificity – i.e. they are sending everyone to the eye clinic. Table 3 shows the agreement (or lack thereof) between the participants and the external expert screeners. This table shows how images of each grade were graded: the grades along the top show the most serious retinopathy or maculopathy grade according to the standard. The leading diagonal (in bold italics) indicates exact agreement between the standard and the graders. Note that none of the images had a consensus grading of R2 (observable retinopathy) or R6 (technical failure, unassessable image). The bottom left-hand corner (in bold) indicates 'over-grading' according to the standard, and the top right-hand corner (in italics) indicates 'under-grading' of referable images. The numbers at the end of the rows and columns show the total number of grades in that row or column.

In general, these findings indicate a lack of specificity in those screeners who took part. The participating group was heterogeneous, with a range of experience, occupations and workloads. The following estimates

Table 2. Participant sample and performance by characteristic

	n	AUC (SE)	DOR (95% CI)
Post			
Other	9	0.922 (0.025)	25.81 (14.41 - 46.24)*
Ophthalmologist	9	0.842 (0.009)	22.48 (13.53 - 37.36)*
Optometrist	243	0.842 (0.009)	11.57 (10.14 - 13.21)
Experience (year/s)			
<1	110	0.825 (0.012)	11.12 (9.35 - 13.21)
2 - 5	60	0.833 (0.026)	11.64 (8.25 - 16.42)
>6	91	0.842 (0.009)	13.68 (11.25 - 16.63)
Competence			
Novice	115	0.815 (0.011)	10.02 (8.52 - 11.61)*
Competent	122	0.815 (0.011)	12.55 (10.14 - 15.53)*
Experienced	24	0.897 (0.021)	24.11 (16.78 - 34.64)
Workload (examinations/month)			
≤10	146	0.838 (0.011)	12.47 (10.69 - 14.53)
11 - 50	60	0.843 (0.019)	11.72 (8.62 - 15.93)
≥51	55	0.843 (0.019)	11.77 (8.73 - 15.87)

AUC = area under the receiver operating characteristic curve; SE = standard error; DOR = diagnostic odds ratio; CI = confidence interval.
*Mean statistically significantly different ($p<0.05$) from the group at the bottom of each section.
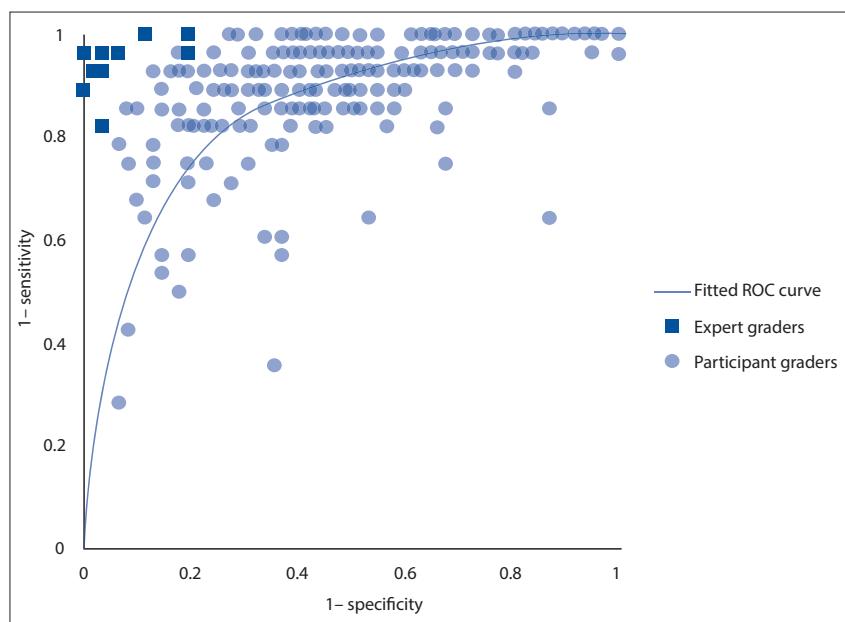

Fig. 2. A scatter plot and ROC curve of grader performance in terms of sensitivity and specificity. Each mark represents a single grader. Dark blue squares represent the external expert graders. Lighter blue circles represent the study participants. The ROC line was calculated from the study participants only. (ROC = receiver operating characteristic.)

Table 3. Cross-agreement (%) between the external graders (standard) and the participants*

Participants	Standard								
	R0	R1	M1	R2	R6	M2	R3	R4	Total
R0	*41.1*	15.2	1.5		9.4	*1.7*	*0.0*	*1.1*	4 011
R1	30.1	*36.5*	18.6		*19.5*	*8.6*	*1.1*	*5.4*	5 979
M1	1.7	1.6	*33.3*		*3.4*	*4.0*	*1.0*	*1.3*	625
R2	0.3	1.1	1.7		*0.6*	*1.1*	*1.1*	*0.4*	194
R6	**1.6**	**0.5**	**0.4**		*35.4*	0.3	0.1	0.6	356
M2	**10.6**	**21.4**	**5.7**		9.4	*43.1*	8.2	8.3	4 554
R3	**7.4**	**15.7**	**25.1**		11.9	31.7	*56.8*	21.8	4 361
R4	**7.3**	**7.9**	**13.6**		10.3	9.3	31.8	*61.0*	3 410
Total	5 742	9 918	522	0	522	3 393	1 044	2 349	

*The values are the percentage of images that were graded by the participants for each standard grading category. The values at the bottom of each column and at the end of each row represent the total number of images graded into each category by the external graders and participants, respectively.
The leading diagonal (***bold italics***) indicates exact agreement between the standard and the graders, the bottom left-hand corner (**bold**) indicates 'over-grading' according to the standard, and the top right-hand corner (*italics*) indicates 'under-grading' of referable images.
The grading outcomes are defined in Table 1.

of the performance for each group were calculated: the area under the ROC curve (AUC) and diagnostic odds ratio (DOR) (Table 2). A comparison of the performance using the Metadisc software[5] found that the optometrist performance was inferior to the other two groups. Those who considered themselves experienced screeners performed better than the competent and the novice screeners. No difference was found to be associated with either the number of years of screening or workload.

In order to examine differences in screener behaviour, we ranked the participants in terms of performance (DOR) and divided them into three groups: low performers (DOR≤10.12, n=87), medium performers (10.12<DOR≤22.24, n=87) and high performers (22.24>DOR, n=87). Table 4 shows the time taken and the mean use of the red filter, the zoom and the ruler by each of these performance groups. The use of these tools by the external expert reference group is also included for comparison. Differences between the groups were tested using a one-way analysis of variance and a *post hoc* Tukey approach. It is clear from this table that the expert reference group took less time and used the tools more. In addition, within the tested groups, the more the tools were used the better the performance.

Discussion

DR has been identified as a valuable biomarker for systemic risk of microvascular complications of diabetes mellitus. With the publication of the Atherosclerosis Risk Study findings by Wong *et al.*[6] and the subsequent publication of the findings of the Japanese Diabetes Complications Study,[7] there has been a new appreciation of the importance of detecting any retinopathy. These studies demonstrated a significantly (approximately two times) increased risk of coronary heart disease and stroke in the study groups. This has changed the emphasis of screening for DR from a blindness prevention initiative (detection of advanced retinopathy) to a primary healthcare initiative (detection of any retinopathy). The importance of a biomarker for systemic complications at primary healthcare level cannot be overstated, particularly in a resource-poor setting. DR provides such characteristics.

There has been a general decline in the provision of screening by GPs in SA. This is because GPs only have ready access to direct ophthalmoscopy, a technology that has low sensitivity,[8] even in the best hands, and is unpopular with patients because it requires the pupils to be dilated. In the ophthalmic world, fundus photography has transformed the ability to detect disease as well as creating a permanent digital record. Fundus photography rather than direct ophthalmoscopy will therefore probably become the standard of care for DR. Patient-friendly, good-quality fundus photography relies on image acquisition utilising non-

Table 4. Screener use of the image manipulation tools and time taken

Group	Time taken (hours) mean (SD)	Red filter (%) mean (SD)	Zoom (%) mean (SD)	Ruler (%) mean (SD)
Low	3.91 (0.24)	57.62 (4.11)*	10.71 (1.98)*	9.39 (1.82)
Medium	4.17 (0.22)*	63.61 (3.58)*	25.30 (3.46)*	11.44 (1.53)
High	4.70 (0.27)*	69.38 (3.16)*	36.21 (3.67)	14.46 (1.74)
Expert	2.02 (0.23)	98.12 (1.18)	57.23 (13.20)	22.44 (5.12)

*Groups in which there is a significant difference between the participant group and the expert group ($p < 0.05$).

mydriatic cameras. These are capital-intensive items that have not been available on a widespread basis in SA.

The OSSA DR programme has strived to lower the barriers to access to screening opportunities for people living with diabetes. OSSA has endorsed the use of non-ophthalmologist graders to try to cope with the burden of disease. There is a widespread appreciation of the value of non-medical personnel as graders. These graders need the backing of a robust, scientific-quality assurance system. This initiative, coupled with optometrist interest, has seen the introduction of many new cameras into the SA healthcare market. A key public health issue has been to establish a responsible way of implementing quality assurance, encouraging participation and ongoing learning. Fundus photo screening is a new discipline in SA. This means that there is a wide range of levels of competence. A system of accreditation was required. This needed to be on an ongoing basis rather than a once-off pass/fail scenario. The system should encourage improvement over time. The Scottish EQA system has demonstrated these characteristics in the Scottish Diabetic Retinopathy Screening collaborative.

The Scottish Retinopathy Grading System was chosen for implementation in SA because of its simple, hierarchical grading of retinopathy characteristics. Simplicity and clear cut-offs for referrals are vital for management of DR. The inclusion of R2 with more frequent review at screening (primary) level is important for the public sector eye clinics, which are already swamped with blindness prevention work. The Scottish system and algorithm for referral has been moderated for the SA scenario to increase safety. The modifications are that any maculopathy (M1 or M2) is to be referred to ophthalmology, and the concept of systemic risk has been incorporated into the algorithm. This is particularly important in SA, where levels of control of diabetes and hypertension are generally very poor. The risk calculator developed by Prof. Einar Steffanson (www.risk.is) has been introduced for use in Africa (www.riskafrica.co.za). This enables non-medical graders to calculate risk and modify the review period for poor control.

The outcomes were significantly better for the ophthalmologist group than for the optometrist group. More experienced graders had higher scores than those with less time since qualifying and a lower level of perceived experience. No significant difference was noted in outcome between the different daily workload groups. The wide range of performance across the groups was larger than that observed in the first Scottish EQA in 2008.[4] It is expected that this will be reduced and overall performance be improved over repeated EQAs, as happened in Scotland.

Safety of non-ophthalmologist graders was a concern. The results show that this group tended to over-refer rather than under-refer. This was reassuring from a safety point of view, but it does mean that more cases than necessary would have been referred. Specificity will be the key aspect of training initiatives in order to address this. Furthermore, 12.6% of cases that should have been referred were not (compared with 4.6% calculated from the Scottish 2010 EQA round).

The time taken to complete the task was similar between the different performance groups. Use of the tools increased with performance, with the high-performance group making the most use of the red-free filter, zoom and ruler controls. After the EQA process was completed, individual feedback and results were given anonymously to each participant. The individual was able to see his or her outcome relative to the peer group. Our plan is to repeat this process on an annual basis, encouraging new participants and monitoring performance. This system is thought to be ideal for the SA environment, where many graders possess their own cameras and would respond better to a peer group-driven incentive to improve rather than an absolute pass/fail outcome.

Study limitations

This study has a number of weaknesses that should be addressed in future work. Not all grading groups were present in our test set. When selected, the imaging mix was thought to contain examples of all possible outcomes. However, consensus was not reached by the expert graders, so grade R2 was not represented in the study set. It is difficult to assess experience and training as a predictor of performance. One would expect the more experienced and better trained to have superior performance, as we have crudely shown in Table 2. However, a better, more refined assessment of these factors will inform where educational efforts could best be focused. In general, our sample is a self-selected group who have expressed an interest; how these findings can be applied to the broader population performing such screening processes is unclear, although we have no reason to expect that our sample is not representative.

Conclusion

The process was well supported by participants and was able to demonstrate safety and areas of weakness that require training. The next SA EQA is scheduled to run in November 2014.

References

1. Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract 2010;87(1):4-14. [http://dx.doi.org/10.1016/j.diabres.2009.10.007]
2. Backlund L, Algvere P, Rosenqvist U. New blindness in diabetes reduced by more than one-third in Stockholm County. Diabet Med 1997;14(9):732-740. [http://dx.doi.org/10.1002/(SICI)1096-9136(199709)14:9<732::AID-DIA474>3.0.CO;2-J]
3. Ryder R. Screening for diabetic retinopathy in the 21st century. Diabet Med 1998;15(9):721-722. [http://dx.doi.org/10.1002/(SICI)1096-9136(199809)15:9<721::AID-DIA694>3.0.CO;2-B]
4. Goatman KA, Philip S, Fleming AD, et al. External quality assurance for image grading in the Scottish Diabetic Retinopathy Screening Programme. Diabet Med 2012;29(6):776-783. [http://dx.doi.org/10.1111/j.1464-5491.2011.03504.x]
5. Zamora J, Abraira V, Muriel A, et al. A software for meta-analysis of test accuracy data. BMC Med Res Methodol 2006;6:31. [http://dx.doi.org/10.1186/1471-2288-6-31]
6. Wong T, Klein R, Sharrett A, et al. Diabetic retinopathy and risk of ischemic stroke: The atherosclerosis risk in communities study. Invest Ophthalmol Vis Sci 2004;45(Suppl 2):U617-U617.
7. Kawasaki R, Tanaka S, Tanakas AS, et al.; Japan Diabetes Complications Study Group. Ophthalmology 2013;120(3):574-582. [http://dx.doi.org/10.1016/j.ophtha.2012.08.029.
8. Harding SP, Broadbent DM, Neoh C, White MC, Vora J. Sensitivity and specificity of photography and direct ophthalmoscopy in screening for sight threatening eye disease: The Liverpool Diabetic Eye Study. BMJ 1995;311(7013):1131-1135. [http://dx.doi.org/10.1136/bmj.311.7013.1131]
9. Early Treatment Diabetic Retinopathy Study Research Group. Grading diabetic retinopathy from stereoscopic color fundus photographs – an extension of the modified Airlie House classification: ETDRS Report Number 10. Ophthalmology 1991;98(5):786-806. [http://dx.doi.org/10.1016/S0161-6420(13)38012-9]