

Grader agreement, and sensitivity and specificity of digital photography in a community optometry-based diabetic eye screening program

Luckni Sellahewa^{1,2}
Craig Simpson²
Prema Maharajan²
John Duffy²
Iskandar Idris³

¹Diabetic Medicine Department, Nottingham University Hospitals,
²North Nottinghamshire Eye Screening Service, Sherwood Forest Hospitals Foundation Trust,
³Division of Medical Sciences and Graduate Entry Medicine, School of Medicine, University of Nottingham, Nottingham, UK

Background: Digital retinal photography with mydriasis is the preferred modality for diabetes eye screening. The purpose of this study was to evaluate agreement in grading levels between primary and secondary graders and to calculate their sensitivity and specificity for identifying sight-threatening disease in an optometry-based retinopathy screening program.

Methods: This was a retrospective study using data from 8,977 patients registered in the North Nottinghamshire retinal screening program. In all cases, the ophthalmology diagnosis was used as the arbitrator and considered to be the gold standard. Kappa statistics were used to evaluate the level of agreement between graders.

Results: Agreement between primary and secondary graders was 51.4% and 79.7% for detecting no retinopathy (R0) and background retinopathy (R1), respectively. For preproliferative (R2) and proliferative retinopathy (R3) at primary grading, agreement between the primary and secondary grader was 100%. Where there was disagreement between the primary and secondary grader for R1, only 2.6% (n=41) were upgraded by an ophthalmologist. The sensitivity and specificity for detecting R3 was 78.2% and 98.1%, respectively. None of the patients upgraded from any level of retinopathy to R3 required photocoagulation therapy. The observed kappa between the primary and secondary grader was 0.3223 (95% confidence interval 0.2937–0.3509), ie, fair agreement, and between the primary grader and ophthalmology for R3 was 0.5667 (95% confidence interval 0.4557–0.6123), ie, moderate agreement.

Conclusion: These data provide information on the safety of a community optometry-based retinal screening program for screening as a primary and as a secondary grader. The level of agreement between the primary and secondary grader at a higher level of retinopathy (R2 and R3) was 100%. Sensitivity and specificity for R3 were 78.2% and 98.1%, respectively. None of the false-negative results required photocoagulation therapy.

Keywords: retinopathy, screening, public health, community, optometry, diabetes

Introduction

Diabetic retinopathy is a highly specific microvascular complication of diabetes and the leading cause of blindness in people under the age of 60 years in industrialized countries.^{1–4} Data from the Early Treatment of Diabetic Retinopathy Study showed that early laser treatment would be more than 90% effective in preventing blindness,⁴ and as such, early detection of sight-threatening disease is crucial in preventing blindness in this group of patients. To this end, previous studies have shown the effectiveness of diabetes eye screening programs to prevent blindness in patients with diabetes.^{2–9} The United Kingdom National Screening Committee therefore recommended a systematic population screening program¹⁰ which was implemented in 2003. As a result, the current National Health Service (NHS) Diabetic Eye Screening Programme is in place.¹¹

Correspondence: Iskandar Idris
Division of Medical Sciences and Graduate Entry Medicine, School of Medicine, University of Nottingham, Royal Derby Hospital, Uttoxeter Road, DE22 3NE, UK
Tel +44 1332 724 668
Fax +44 1332 724 697
Email iskandar.idris@nottingham.ac.uk

Digital retinal photography with mydriasis is the preferred modality for diabetic eye screening based on its reported values for sensitivity and specificity,^{12–15} and its ability to quality assure screening standards.^{16,17} This modality of retinopathy screening fulfils the Exeter minimum standard for sensitivity and specificity of 80% and 95%, respectively, for robust and safe diabetic retinopathy screening.^{18,19} Conventionally, this utilizes technicians to perform the primary grading, with secondary grading performed by more experienced screeners or clinicians, and arbitration grading performed by an ophthalmologist or a diabetologist with expertise in diabetic retinopathy screening. However, in selected screening programs, primary and secondary gradings are performed by trained opticians. Whilst data are available on the effectiveness of individual screening modalities,^{10–13,17–19} there is currently only one study that has looked at the interobserver agreement between primary graders and an expert grader.²⁰ Information on the safety, effectiveness, and agreement between primary and secondary graders for images of patients undergoing routine diabetic eye screening in a community optometry-based retinopathy screening program has not yet been reported.

Materials and methods

The North Nottinghamshire diabetic retinopathy screening service has utilized an optometry-based model since April 2006 and involves 36 optometrists across 21 sites. Screening is undertaken by local optometrists, and two-field digital images of the retina are recorded in the database and graded. All models and makes of the retinal cameras in use, as well as their age, are approved based on criteria set by the NHS Diabetic Eye Screening Programme. Tropicamide 1% is used to dilate the pupils to an acceptable size for screening, which is performed according to a standard national screening protocol. Primary and secondary grading is carried out by optometrists on the digital retinal images, and a web-based referral to an ophthalmologist is required if there is disagreement between primary and secondary graders or if sight-threatening retinopathy is observed.

For this study, data were collected retrospectively between January 2011 and December 2011 from a cohort of 8,977 patients registered in an optometry-based retinal screening program database currently in place in North Nottinghamshire. These patients were reviewed by optometrists who carried out digital retinal photography. Images were stored in a web-based database and graded according to the national screening standard.¹¹ Grading levels were as follows: no retinopathy (R0), background retinopathy (R1), preproliferative retinopathy (R2), proliferative retinopathy (R3), and maculopathy (M1). Any retinopathy detected by

a primary grader (R1, R2, M1) and 10% of images with no evidence of retinopathy (R0) was sent for secondary grading performed by another optometrist. If there was any disagreement between the primary and secondary grader, the images were sent to arbitration, which was performed by an ophthalmologist. The presence of proliferative retinopathy (R3) would require an urgent referral to ophthalmology. However, during 2011, due to an internal quality audit that was being undertaken, all patients with R1 were referred to the ophthalmologist for screening. Retinal images that were not gradable by the primary grader for reasons such as previous surgery or cataracts were referred directly to ophthalmology. Patients under ophthalmology follow-up were kept under ophthalmology review with follow-up appointments until their retinopathy was stable. The screening program also has in place a fail-safe mechanism (monitored by a fail-safe officer) whereby images of patients subsequently found to have R3 or have undergone photocoagulation therapy are traced back to see whether this was missed during screening on an ongoing basis. No R3 was being missed at screening during the period of this audit. Once the patients had stable retinopathy with no immediate intervention required, they were referred back into the local retinal screening recall process.

We calculated the agreement between the primary and secondary grader as well as between individual graders and ophthalmologists by means of Kappa statistics.²¹ We also looked at the proportion of disagreement leading to an upgrading of the retinopathy level. Assessment of sensitivity and specificity values in this study was limited to images graded as R3, since all R3 are referred to an ophthalmologist for arbitration or a final grading. R3 grading from the primary grader was compared against the “gold standard” ophthalmological diagnosis. Sensitivity is calculated as the (number of true positives/true positives + false negatives) while specificity is calculated as the (number of true negatives/true negatives + false positives). This work is labeled as service evaluation. The audit work and data derived from this work are part of the program’s ongoing clinical governance exercise to maintain standards of retinopathy screening within the service. The statistical analysis was performed using SPSS version 14 software (SPSS Inc., Chicago, IL, USA).

Results

Of 8,977 patients (15,583 images), 734 patients were graded as R0 by the primary grader. Of these, 377 were graded as R0 by the secondary grader. This resulted in 51.4% agreement between the primary and secondary grader for patients graded as R0 at primary grading. The other 357 patients had no agreement between the primary and secondary grader.

From these, 4.8% (n=17) were downgraded and 3.6% (n=13) were upgraded by ophthalmology (Table 1).

Background retinopathy grading (R1) was given to 7,784 patients by the primary grader and 1,448 of these were graded by ophthalmology. The level of agreement between primary and secondary graders in this group was 79.7% (n=6,204). Among these patients, 15.5% (n=207) of agreement was reported between the primary grader and ophthalmology, while the agreement between the secondary grader and ophthalmology was 10.7% (n=835). For the proportion in which there was disagreement between the primary and secondary grader, 2.6% (n=41) were upgraded, of which 1% (n=16) were upgraded to R3 (Table 1). For the proportion in which there was disagreement between the primary and secondary grader, 0.8% (n=13) were downgraded to a different grade by ophthalmology (Table 1). Where patients were graded R2 (n=210) at primary grading, agreement between the primary and secondary grader was 100% (Table 1); 207 of the 210 that were graded as R2 by the primary grader were graded by the secondary grader as well as ophthalmology. This was due to an internal quality assurance audit that was taking place in 2011.

Proliferative retinopathy (R3) was detected in 249 patients by the primary grader, but only 31.7% (79) of these were subsequently confirmed as R3 by ophthalmology. Of the total population screened (n=8,977), 8,728 were found not to have R3 by the primary grader, while 1,777 patients were confirmed by ophthalmology not to have R3. From these data, the sensitivity and specificity for R3 in our cohort is 78.2% and 98.1% (Table 1); 3.6% of normal (R0) and 2.6% of background retinopathy (R1) had a disagreement in grading, leading to an upgrading of retinopathy level by ophthalmology. Ten percent of images graded as R0 went

through to ophthalmology for arbitration. Of these, there was no agreement between the primary and secondary grader, but there was 56.6% agreement between the primary grader and ophthalmology, and 36.6% agreement between the secondary grader and ophthalmology.

We used Kappa statistics to evaluate the level of agreement between primary and secondary graders and between primary and arbitration graders for R0–R2. There was an observed kappa of 0.3223 (95% confidence interval 0.2937–0.3509) and 0.269 (95% confidence interval 0.216–0.321), respectively (Tables 2 and 3). The level of agreement between the primary grader and ophthalmology for R3 using Kappa statistics gives an observed kappa of 0.5667 (95% confidence interval 0.4557–0.6123).

Discussion

For a systematic screening program to be effective, it needs a database that is robust and well maintained. The system currently in place in North Nottinghamshire uses a central call/recall center with ongoing quality assurance taking place at all stages of the process. In addition to their professional qualification registered by the General Optical Council which regulates dispensing opticians and optometrists, all screeners/graders would have undertaken a certificate for diabetic retinopathy screening by City and Guilds, as well as undergoing a test training set mandated by the NHS Diabetic Eye Screening Programme. During the period of the audit, one test training set was performed by the opticians. However, data for the intergrader agreement based on this exercise were not available. Although the national program recommended only 10% of R0 to be secondarily screened, we performed an internal audit for the year 2009–2010, where all R0 underwent secondary grading as a result of a quality

Table 1 Percentage of agreement, disagreement, upgrading, and downgrading of images in the North Nottingham screening program

	R0 (n=734) n (%)	R1 (n=7,784) n (%)	R2 (n=210) n (%)	R3 (n=249) n (%)
Agreement between primary and secondary grader	377 (51.4%)	6,204 (79.7%)	210 (100%)	249 (100%)
Agreement between primary grader and ophthalmology	Not evaluated	1,207 (15.5%)	78 (37.1%)	79 (31.7%)
Agreement between secondary grader and ophthalmology	Not evaluated	835 (10.7%)	78 (37.1%)	Not evaluated
Disagreement leading to downgrading by ophthalmologist	17 (4.8%)	Not evaluated	Not evaluated	113 (45.4%)
Disagreement leading to upgrading by ophthalmologist	13 (3.6%)	41 (2.6%)	Not evaluated	Not evaluated
Disagreement leading to upgrading to R3 by ophthalmologist	Not evaluated	13 (0.8%)	Not evaluated	Not evaluated

Notes: Using Kappa statistics to evaluate agreement between primary grader and ophthalmology for R3, the observed κ is 0.57 (95% confidence interval 0.46–0.61), ie, moderate agreement. Sensitivity and specificity for detecting R3 are 78.2% and 98.1%, respectively.

Abbreviations: R0, no retinopathy; R1, background retinopathy; R2, preproliferative retinopathy; R3, proliferative retinopathy.

Table 2 Agreement and disagreement for primary grader (horizontal axis) and secondary grader (vertical axis)

	R0	R1	R2
R0	17	185	6
R1	12	1,207	122
R2	0	36	78

Notes: Using Kappa statistics to evaluate overall level of agreement between primary and secondary graders for R0–R2, the observed κ is 0.3223 (95% confidence interval 0.2937–0.3509).

Abbreviations: R0, no retinopathy; R1, background retinopathy; R2, preproliferative retinopathy.

assurance exercise recommended by the NHS Retinopathy Screening Programme. No sight-threatening retinopathy (R2 or higher) was identified.

The above study provides novel information on the safety and effectiveness of a community-based retinal screening program that uses optometrists at both the primary and secondary grader level compared with other optometry or nonoptometry-based programs that use senior graders, diabetologists, or ophthalmologists as secondary graders.

Evidence for the effectiveness of screening is based on evidence of treatment efficacy especially after early detection and on cost-effectiveness. Comparing this screening program with the Exeter standards,^{18,19} ours achieved a specificity level above the expected 95% but the sensitivity level was marginally short of the recommended 80% threshold. Of note, the sensitivity data here refer to data analysis specific to R3 rather than data from the whole program. Moreover, it is conceivable that the slightly higher level of false-positives observed here reflects a slightly overcautious approach by optometrists to grading in patients with a higher likelihood of abnormalities in their eyes. In addition, image arbitration was performed by an ophthalmologist who may decide on the final “grade” based on clinical need for photocoagulation therapy rather than actual reporting of the images. Nevertheless, the importance of appropriate sensitivity and specificity for any screening modality has become more important in view of some recent evidence which may advocate for a different frequency of retinopathy screening for different individuals depending on the risk of retinopathy progression, based on

Table 3 Agreement and disagreement for primary grader (horizontal axis) and arbitration grader (vertical axis)

	R0	R1	R2
R0	377	1,107	0
R1	354	6,204	0
R2	3	261	210

Notes: Using kappa statistics to evaluate overall level of agreement, between primary and secondary graders for R0–R2, the observed κ is 0.269 (95% confidence interval 0.216–0.321).

Abbreviations: R0, no retinopathy; R1, background retinopathy; R2, preproliferative retinopathy.

baseline and/or previous screening results.²⁴ Despite a high false-negative rate, none of the false negatives required urgent photocoagulation therapy, which reflects a subsequent “clinical” diagnosis by the ophthalmologist rather than a misdiagnosis by the optometrist. This has been confirmed by regular audit of our data based on the governance structure currently in place in our screening program. It was also reassuring to note that the levels of agreement between primary and secondary graders for higher levels of retinopathy (R2 and R3) were both 100%. For lower levels of retinopathy, ie, R0 and R1, agreement between primary and secondary graders were lower at 51.4% and 79.7%, respectively. Of these, 3.6% of normal (R0) and 2.6% of background (R1) retinopathy showed a disagreement in grading, leading to an upgrading of retinopathy level by ophthalmology, but none required photocoagulation therapy.

Some limitations to this study needs to be highlighted. To calculate sensitivity and specificity, we analyzed data specific to R3 only. This was because only 10% of R0 and some of R1 and R2 were referred to ophthalmology, whereas all R3 were referred to an independent ophthalmologist. Because of this, we were unable to look at the sensitivity and specificity for the whole cohort, which affects the results reported in our study. We used the ophthalmologist grade as the gold standard, so it would be important to have all retinopathy graded as R2 by the primary grader reviewed by ophthalmology to ensure that none of these would need to be upgraded to R3, which would mean they will need ophthalmology follow-up and potential treatment. The study was carried out by retrospective data collection, which would also be considered as a limitation, due to the presence of confounding biases. We were also not able to reliably determine results for maculopathy within our program. Further, we were not able to accurately adjust results for ungradable images, due to poor patient compliance with the screening protocol, poor mydriasis, or other factors. Interpretation of the results is limited to this program and cannot necessarily be generalized to other programs. Lastly, although Kappa statistics is a recognized method for assessment of agreement, the magnitude of kappa reflecting adequate agreement is unclear. However, arbitrary guidelines are available to indicate level of agreement, although these are not evidence-based. Generally, however, it is accepted that a kappa score >80% would suggest very good agreement.^{25,26} Despite this, due to methodological limitations of other research in this area, and due to a lack of data and evidence of optometrists as primary and secondary graders in detecting R3 in a retinopathy screening program, we believe data from this study would enhance available knowledge concerning the

safety and effectiveness of an optometry community-based retinopathy screening program.

There is no clear evidence suggesting who has the best sensitivity and specificity for detecting sight-threatening retinopathy, ie, whether it is independent graders, optometrists, diabetologists, general practitioners, or ophthalmologists. A single study showed that retinal photographs assessed by optometrists could achieve >91% sensitivity in detecting R3 or sight-threatening retinopathy.²⁰ Data on the effectiveness of individual screening modalities are widely available.^{13,17,19,23} However, our study provides unique data on the safety, effectiveness, and agreement between primary and secondary graders for images of patients undergoing routine diabetes eye screening in a community optometry-based retinopathy screening program.

Author contributions

LS contributed to the data acquisition and analysis, and interpretation of the data, and wrote the first draft of the manuscript. CS supported the acquisition and analysis of the data. JD and PM contributed to analysis or interpretation of data. II conceptualized the study and contributed to the design, analysis, and interpretation of the data. II is the guarantor for this study. All authors contributed to the writing of the manuscript and agreed on the final draft.

Disclosure

The authors report no conflicts of interest in this work.

References

- Owens DR, Gibbins RL, Kohner E, et al. Diabetic retinopathy screening. *Diabet Med.* 2000;17(7):493–393.
- Stefánsson E, Bek T, Porta M, et al. Screening and prevention of diabetic blindness. *Acta Ophthalmol Scand.* 2000;78(4):374–385.
- Garvican L, Clowes J, Gillow T. Preservation of sight in diabetes: developing a national risk reduction programme. *Diabet Med.* 2000;17(9):627–634.
- Scanlon P, Aldington S, Wilkinson C. Early Treatment Diabetic Retinopathy Study Research Group. Early photocoagulation for diabetic retinopathy, ETDRS report number 9. *Ophthalmology.* 1991;98(5):766–785.
- James M, Turner D, Broadbent D, et al. Cost effectiveness analysis of screening for sight threatening diabetic eye disease. *BMJ.* 2000;320(7250):1627–1631.
- Buxton M, Sculpher M, Ferguson B, et al. Screening for treatable diabetic retinopathy: a comparison of different methods. *Diabet Med.* 1991;8(4):371–377.
- Sculpher M, Buxton M, Ferguson B, et al. A relative cost-effectiveness analysis of different methods of screening for diabetic retinopathy. *Diabet Med.* 1991;8(7):644–650.
- Bachmann MO, Nelson S. Impact of diabetic retinopathy screening on a British district population: case detection and blindness prevention in an evidence based model. *J Epidemiol Community Health.* 1998;52(1):45–52.
- Davies R, Roderick P, Canning C, et al. The evaluation of screening policies for diabetic retinopathy using simulation. *Diabet Med.* 2002;19(9):762–770.
- UK National Screening Committee. Available from: <http://www.screening.nhs.uk>. Accessed May 31, 2013.
- NHS Diabetic Eye Screening Programme. Available from: <http://diabeticeye.screening.nhs.uk>. Accessed May 31, 2013.
- Ferguson BA, Humphreys JE, Altman JFB, et al. Screening for treatable diabetic retinopathy: a comparison of different methods. *Diabet Med.* 1991;8(4):371–377.
- Hutchinson A, McIntosh A, Peters J, et al. Effectiveness of screening and monitoring tests for diabetic retinopathy – systematic review. *Diabet Med.* 2000;17(7):495–506.
- Scanlon PH, Wilkinson CP, Aldington J, et al. Screening for diabetic retinopathy. In: Scanlon PH, Wilkinson CP, Aldington SJ, Matthews DR, editors. *A Practical Manual of Diabetic Retinopathy Management.* Oxford, UK: Wiley-Blackwell; 2009.
- Taylor D, Fisher J, Jacob J, et al. The use of digital cameras in a mobile retinal screening environment. *Diabet Med.* 1999;16(8):680–686.
- 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.
- Sallam A, Scanlon PH, Stratton IM, et al. Agreement and reasons for disagreement between photographic and hospital biomicroscopy grading of diabetic retinopathy. *Diabet Med.* 2011;28(6):741–746.
- Harding SP, Broadbent DM, Neoh C, et al. Sensitivity and specificity of photography and direct ophthalmoscopy in screening for sight threatening eye diseases: the Liverpool Eye Study. *BMJ.* 1995;311(7013):1131–1135.
- Harding S, Greenwood R, Aldington S, et al. Grading and disease management in national screening for diabetic retinopathy in England and Wales. *Diabet Med.* 2003;20(12):965–971.
- Patra S, Gomm EM, Macipe M, et al. Interobserver agreement between primary graders and an expert grader in the Bristol and Weston diabetic retinopathy screening programme: a quality assurance audit. *Diabet Med.* 2009;26(8):820–823.
- Donner A, Shoukri M, Klar N, et al. Testing the equality of two dependent Kappa statistics. *Stat Med.* 2000;19(3):373–387.
- Gibbins RL, Owens DR, Allen JC, et al. Practical application of the European field guide in screening for diabetic retinopathy by using ophthalmoscopy and 35 mm retinal slides. *Diabetologia.* 1998;41(1):59–64.
- Olson J, Strachan F, Hipwell J, et al. A comparative evaluation of digital imaging, retinal photography and optometrist examination in screening for diabetic retinopathy. *Diabet Med.* 2003;20(7):528–534.
- Stratton IM, Aldington SJ, Taylor DJ, Adler AI, Scanlon PH. A simple risk stratification for time to development of sight threatening diabetic retinopathy. *Diabetes Care.* 2013;36:580–585.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–174.
- Fleiss JL. *Statistical Methods for Rates and Proportions.* 2nd ed. New York, NY, USA: John Wiley; 1981.

Clinical Ophthalmology

Publish your work in this journal

Clinical Ophthalmology is an international, peer-reviewed journal covering all subspecialties within ophthalmology. Key topics include: Optometry; Visual science; Pharmacology and drug therapy in eye diseases; Basic Sciences; Primary and Secondary eye care; Patient Safety and Quality of Care Improvements. This journal is indexed on

Submit your manuscript here: <http://www.dovepress.com/clinical-ophthalmology-journal>

Dovepress

PubMed Central and CAS, and is the official journal of The Society of Clinical Ophthalmology (SCO). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.