Accuracy of diabetic retinopathy screening by trained non-physician graders using non-mydriatic fundus camera

Mayuri Bhargava1,2, MD, Carol Yim-Lui Cheung1, PhD, Charumathi Sabanayagam1, MD, Ryo Kawasaki2, MD, PhD, C Alex Harper3, MD, Ecosse L Lamoureux4,2, MSc, PhD, Wai Leng Chow3, MBBS, Adrian Ee4, MBBS, Haslina Hamzah1, BSc, Maisie Ho1, BSc, Wanling Wong1, MBiostat, Tien Yin Wong1,2,5, PhD, FRCSE

INTRODUCTION
We compared the agreement of diabetic retinopathy (DR) assessment between trained non-physician graders (NPGs) and family physicians (FPs) in a primary healthcare setting.

METHODS
This was a cross-sectional study conducted retrospectively over a period of one month. The participants were diabetic patients from two primary healthcare clinics (polyclinics) in Singapore. Single-field digital retinal images were obtained using a non-mydriatic 45-degree fundus camera. Retinal images were graded for the presence or absence of DR by FPs at the polyclinics and by NPGs at a central ocular grading centre. The FPs’ and NPGs’ assessments of DR were compared with readings by a single retinal specialist (reference standard).

RESULTS
A total of 367 diabetic patients (706 eyes) were included in the study. The mean age of the patients was 63 years, and the majority were Chinese (83.8%). For DR assessment, the agreement between NPGs and the retinal specialist was substantial (κ = 0.66), while the agreement between FPs and the retinal specialist was only fair (κ = 0.40). NPGs’ assessment showed higher sensitivity (70% vs. 45%) and comparable specificity (94% vs. 92%) as compared to FPs’ assessment. The area under the receiver operating characteristic curve of NPGs’ assessment of DR was greater than that of the FPs’ (0.82 vs. 0.69, p < 0.001).

CONCLUSION
This study has demonstrated that trained NPGs are able to provide good detection of DR and maculopathy from fundus photographs. Our findings suggest that DR screening by trained NPGs may provide a cost-effective alternative to FPs.

Keywords: diabetic retinopathy, family physicians, screening, single-field digital retinal images, trained graders

An alternative could be the utilisation of trained NPGs for DR screening, which may require the evaluation of hundreds of photographs every day, most of which have no abnormalities. This would help to reduce FPs’ burden and provide faster service to the patients as well. To date, the performance of NPGs for DR detection from digital fundus images has never been directly compared with that of FPs, with only one study that reported on the diagnostic accuracy in screening DR by FPs. Thus, it has never been clearly shown whether DR screening by NPGs is an acceptable alternative to the current practice (i.e. DR screening by FPs). The purpose of the present study was to compare the agreement of DR assessment between trained NPGs and FPs, with a retinal specialist’s assessment as the reference standard.

METHODS
We performed a cross-sectional study of DR assessment from retinal images taken in two government primary healthcare clinics (polyclinics) in Singapore. Currently, diabetic eye screening in Singapore is conducted via the Diabetic Retinal Photography (DRP) programme across polyclinics, where single-field digital retinal images are read and decisions to refer to ophthalmologists are
Table I. Demographic characteristics of diabetic patients, stratified by diabetic retinopathy status.

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Diabetic retinopathy</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No (n = 334)</td>
<td>Yes (n = 33)</td>
</tr>
<tr>
<td>Male gender</td>
<td>149 (44.7)</td>
<td>20 (44.4)</td>
</tr>
<tr>
<td>Chinese race†</td>
<td>280 (83.8)</td>
<td>27 (81.8)</td>
</tr>
<tr>
<td>Age (yrs)</td>
<td>62.9 ± 11.0</td>
<td>65.5 ± 14.9</td>
</tr>
<tr>
<td>Duration of diabetes mellitus (yrs)</td>
<td>7.3 ± 5.2</td>
<td>9.9 ± 7.2</td>
</tr>
</tbody>
</table>

Note: Data is presented as number (%) or mean ± standard deviation.

*p-value for differences between diabetic retinopathy and no diabetic retinopathy, by t-test or chi-square test, as appropriate. p < 0.05 is statistically significant.

†Other races in the study include Malay: n = 12 (3.27%); Indian: n = 40 (10.89%); Others (Filipino, Eurasian, Sikh, Burmese) n = 8 (2.18%).

made by trained FPs. The participants were diabetic outpatients from two polyclinics who were screened under this programme in March 2009.

At the polyclinics, nurses were trained on how to operate the fundus camera and take photographs. The pupils were dilated only if they were too small for adequate imaging and after consent was obtained from patients. Single-field retinal photographs centred on the macula and encompassing the optic disc were obtained from both eyes of each participant using a non-mydriatic 45-degree digital retinal camera (Canon CR-DGi with a 10-D SLR back; Canon, Tokyo, Japan). The photographs were digitally stored and copies of the grading report were kept in the patients’ case notes.

The digital images were randomly graded by five FPs at the polyclinics. On average, 470 patients per day were seen by the five FPs in a single polyclinic. Due to their busy schedules, FPs are currently trained by a retinal specialist for a two-hour period and are accredited every two years. After grading the fundus images, the FPs filled out a standard form derived from grading outcomes based on the Ministry of Health Singapore Clinical Practice Guidelines on Diabetes Mellitus, for each patient. The same photographs were also read by trained NPGs at the ocular grading centre of the Singapore Eye Research Institute, in a masked fashion using the same form. The centralised grading centre comprises five trained NPGs who have been validated by the University of Melbourne and have a high reliability index (κ = 0.8). NPGs undergo one year of rigorous training and regular yearly auditing. Finally, the images were graded separately in a masked fashion by a retinal specialist at the Centre for Eye Research Australia, University of Melbourne, Australia. Neither the FPs nor the NPGs were aware that their assessments were being compared to those of the retinal specialist.

The primary grading outcome used in the study was that of the ‘presence/absence of disease’, based on standard definitions and clinical guidelines. Each digital fundus image was graded according to the lesions present and then categorised as per the severity level of DR. ‘No/absent DR’ was defined as consisting of isolated microaneurysms (MAs), including questionable MAs, MAs, retinal haemorrhages, hard exudates, cotton wool spots and neovascularisation constituted the ‘presence of DR’. Nonproliferative diabetic retinopathy (NPDR) was defined by retinal haemorrhages, cotton wool spots/venous beading and intraretinal microvascular abnormalities, which were categorised as mild, moderate or severe, depending on the extent of the retinal lesions. A diagnosis of proliferative diabetic retinopathy (PDR) was made if one or more of the following was present: (a) any neovascularisation elsewhere (NVE); (b) neovascularisation of the disc less than one-third of the disc’s diameter; and (c) vitreous or preretinal haemorrhage with NVE less than half of the disc’s area. Maculopathy was defined as retinal oedema, thickening or hard exudates within 500 microns of the macular centre, and retinal oedema or thickening one disc diameter or larger in any size, with any part within one disc diameter of the centre of the macula. In either retinopathy or maculopathy grading, the raters were allowed to label images as ungradable, based on their judgement. All the patients in the study were referred based on four categories: (a) referral within one week for PDR and any maculopathy; (b) referral in less than one month for moderate to severe NPDR; (c) referral between 1–3 months for mild NPDR; and (d) referral within one year for annual photograph in cases with no DR. The FPs did not categorise DR into nonproliferative and proliferative types.

All statistical analyses were performed using Stata version 10.0 (StataCorp, College Station, TX, USA). The kappa statistic (linear weighted kappa was used in case of more than two categories) was used to assess inter-rater agreement. A kappa (κ) value between 0.0 and 0.2 indicates slight agreement, between 0.21 and 0.40 indicates fair, between 0.41 and 0.60 indicates moderate, between 0.61 and 0.80 indicates substantial, and between 0.81 and 1.00 indicates almost perfect agreement. Sensitivity, specificity, accuracy and the area under receiver operating characteristic curves (AUCs) were performed to compare the diagnostic performances of the NPGs and FPs against the assessments of the retinal specialist (reference standard).

RESULTS

Among the digital retinal images obtained, 794 images from 397 patients were graded by the FPs and the retinal specialist. Due to missing data, the NPGs only analysed 760 images (380 subjects), which were then case-matched, leaving 706 digital retinal images of 367 participants for the final analysis. All the images were gradable. Table 1 shows the demographics of the study population. The mean age ± standard deviation of the study population was 63.18 ± 11.37 years. 54.00% of the participants (n = 198) were female. A majority of the patients were Chinese (n = 307, 83.65%).
Patients with DR (n = 33, 8.99%) were older (mean age 65.5 years) and had a longer duration of diabetes mellitus (mean duration 9.9 years), although these variables were not found to be statistically significant (p > 0.05).

The DR and maculopathy assessments of the 706 images, as well as referrals for the 367 patients by the three grading groups, are shown in Table II. The NPGs identified 17.0% of the images with DR, which was consistent with the assessment of the retinal specialist, who detected 17.8% of eyes with DR (κ = 0.66). However, the FPs recorded the presence of DR in only 12.3% of eyes (κ = 0.40). The NPG group detected maculopathy in 2.83% of the images, with a higher reliability (κ = 0.43) in contrast to the FPs, who assessed 2.69% of the images to have maculopathy (κ = 0.29). The ‘< one week’ referrals of cases by the NPGs (4.9%; κ = 0.35) in contrast to the FPs (2.2%; κ = 0.33) were comparable. These results suggested that NPGs’ assessments of DR and maculopathy were reliable and in agreement with those of the retinal specialist.

Table III shows the diagnostic performance among the graders. The sensitivity and specificity of NPGs’ assessment in detecting the presence of DR were higher (70% vs. 45%) and comparable (94% vs. 92%) to those of the FP’s assessments, respectively. The AUC of NPGs’ grading was greater than that of the FP’s (0.642 vs. 0.615; p < 0.001). Sensitivity in detecting maculopathy was also found to be higher among the NPGs when compared with the FPs (58% vs. 42%), with a greater AUC for the NPGs (p = 0.314).

**DISCUSSION**

In this study, we demonstrated that the diagnostic performance and agreement of the NPGs were comparable to, and in some instances, better than that of the FPs in the assessment of DR from fundus photographic images, when compared against a gold standard reference (the retinal specialist). This implies that the use of NPGs for DR screening, in addition to being accurate and economical, could ultimately ease FPs’ workload.

Previous studies have reported that the assessment of DR by NPGs is highly sensitive. Scotland et al. and Philip et al. have shown that NPGs’ assessments have sensitivity values ranging from 87% to 99% in DR detection. DR assessment of digital fundus photographs by NPGs has been shown to have 100% sensitivity and 71% specificity when compared with ophthalmoscopy in a study by Lin et al. Similar results are mirrored in our study, which shows a good consensus between retinal specialists’ and trained graders’ interpretations of digital images.

Various studies have been done on DR screening where direct ophthalmoscopy was performed by FPs. However, there is limited data in the literature with regard to FPs’ assessment of DR detection from retinal images, which is still commonly practised in many countries, including Singapore. Shanit et al. in their pilot study, explored the possibility of grading retinal images by FPs for the evaluation of various retinal diseases. Owens et al. showed that FPs have 79% sensitivity against the centralised grading centre for DR assessment, when compared by ophthalmoscopy and grading mydriatic 35 mm digital retinal images. Farley et al. using 1,040 single-field non-mydriatic digital images, found FPs to have a decreased sensitivity of 85% for DR detection when compared to retinal specialists’ evaluation. Lim et al. have shown that in Singapore, only 38% of diabetic patients referred by FPs to tertiary ophthalmic centres in 2002 were found to have DR. The lack of participation of FPs in DR screening was indicated as a limitation in another study. As the methods for fundus photography and image quality can influence the grading outcome, we thought it

---

### Table II. Inter-grader agreement of diabetic retinopathy (DR) assessment and referral for non-physician graders (NPGs), family physicians (FPs) and the retinal specialist (RS).

<table>
<thead>
<tr>
<th>Presence of DR (n = 706)</th>
<th>No. (%)</th>
<th>Kappa ± standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPG</td>
<td>120 (17.0)</td>
<td>0.656 ± 0.038</td>
</tr>
<tr>
<td>FP</td>
<td>87 (12.3)</td>
<td>0.400 ± 0.036</td>
</tr>
<tr>
<td>RS</td>
<td>126 (17.8)</td>
<td>0.425 ± 0.036</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Presence of maculopathy (n = 706)</th>
<th>No. (%)</th>
<th>Kappa ± standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPG</td>
<td>20 (2.83)</td>
<td>0.573 ± 0.028</td>
</tr>
<tr>
<td>FP</td>
<td>19 (2.69)</td>
<td>0.356 ± 0.028</td>
</tr>
<tr>
<td>RS</td>
<td>12 (1.69)</td>
<td>0.314</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Referral type (n = 367)</th>
<th>No. (%)</th>
<th>Kappa ± standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 1 wk</td>
<td>18 (4.9)</td>
<td>0.66 ± 0.038</td>
</tr>
<tr>
<td>&lt; 1 mth</td>
<td>22 (6.0)</td>
<td>0.68 ± 0.042</td>
</tr>
<tr>
<td>1-3 mths</td>
<td>49 (13.4)</td>
<td>0.57 ± 0.028</td>
</tr>
<tr>
<td>Within 1 year</td>
<td>278 (75.7)</td>
<td>0.314</td>
</tr>
</tbody>
</table>

### Table III. Diagnostic analysis of diabetic retinopathy (DR) assessment for non-physician graders (NPGs) and family physicians (FPs) vs. the retinal specialist as the gold standard.

<table>
<thead>
<tr>
<th>DR (n=126)</th>
<th>Sensitivity (%) (95% CI)</th>
<th>Specificity (%) (95% CI)</th>
<th>Accuracy (%) (95% CI)</th>
<th>AUC (%) (95% CI)</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPG</td>
<td>69.8 (61.3, 77.2)</td>
<td>94.4 (92.3, 96.1)</td>
<td>90.1 (87.7, 92.1)</td>
<td>0.822 (0.78, 0.863)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>FP</td>
<td>44.7 (36.5, 53.2)</td>
<td>92.4 (90.1, 94.2)</td>
<td>84.3 (81.6, 86.7)</td>
<td>0.686 (0.642, 0.729)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maculopathy (n=12)</th>
<th>Sensitivity (%) (95% CI)</th>
<th>Specificity (%) (95% CI)</th>
<th>Accuracy (%) (95% CI)</th>
<th>AUC (%) (95% CI)</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPG</td>
<td>58.3 (32, 80.7)</td>
<td>98.1 (96.8, 98.9)</td>
<td>97.5 (96, 98.4)</td>
<td>0.782 (0.608, 0.956)</td>
<td>0.314</td>
</tr>
<tr>
<td>FP</td>
<td>41.7 (19.3, 68)</td>
<td>98.0 (96.6, 98.8)</td>
<td>97.1 (95.5, 98.1)</td>
<td>0.698 (0.512, 0.884)</td>
<td></td>
</tr>
</tbody>
</table>

*Pairwise comparison of AUCs between the grading of NPG and FP. AUC: area under receiver operating curve; CI: confidence interval.
was important to directly compare the grading results between NPGs and FPs on the same fundus images. Our study is the first to directly compare the diagnostic performance of NPGs and FPs against that of a retinal specialist (reference standard), with the results showing NPGs to have the sensitivity and diagnostic accuracy comparable to that of a retinal specialist.

Theoretically, there are several advantages of employing NPGs at a centralised grading centre to assess DR as compared to the employment of FPs for the same purpose. First, this would enable FPs to have more time to focus on treating and monitoring their patients if DR grading were to be carried out by trained graders. Second, Lin et al has demonstrated a reduction in the underdiagnosis of DR, from 24% by ophthalmoscopy to 8% by digital photography, by a centralised grading centre when compared with the reference standard of seven-field standard mydriatic stereoscopic colour photographs, which may eventually prove to be cost-effective for DR screening.(27) Third, it is suggested that the reliability of the DR screening programme can be enhanced via internal auditing by a centralised grading centre in order to institute baseline standards, against which future interobserver agreements can be measured for quality assurance.(28) Finally, a centralised grading centre will provide information about the accuracy and adequacy of single-field fundus photography implemented by the current DRP programme. Using trained graders in a centralised reading centre, inherent with ease of archiving, retrieval and remote data transmission of digital images(29) with retinal specialists as backup, has been described as the future strategy in diagnostic eye care.(27)

Compared with other population-based studies (e.g. Singapore Malay Eye Study [SiMES]), the prevalence of DR (17.8%) and clinically significant macular oedema (CSME) (1.69%) in this cohort was lower.(1,15,26) This could be attributed to a number of reasons. First, more than half of the patients (50.9%) in our study were observed to have diabetes mellitus of less than five years' duration, when DR incidence is known to be low.(27) Whereas in SiMES, only 39% of the patients had diabetes mellitus for less than five years. Also, the current sample had digital single-field images, which may limit the detection of DR and CSME when compared with the two-field photographs used in SiMES(20) or the multiple-field stereoscopic photographs used in other studies.(6)

Non-mydriatic retinal photography was used in this study, as most studies have reported no significant difference in sensitivity/specificity between mydriatic and non-mydriatic methods in detecting the presence of DR.(10) Patients are also known to report discomfort and functional limitations from pharmacological mydriasis.(24) Due to the lack of stereoscopic images, we had used the grading protocol of ‘presence/absence of disease’ derived from the National Health Medical Research Council, which is based on the International Clinical Diabetic Retinopathy Disease Severity Scale and Diabetic Macular Oedema Severity Scale.(29) Lin et al reported that the sensitivity of DR detection was 78% using single-field images,(27) which may be improved by establishing a new standardised DR grading protocol for single-field retinal photography. Even though there is level I evidence(5,27) that single-field fundus photography can act as a screening tool, the United Kingdom currently uses two-field fundus photography in their DRP screening programme.(30) This addition may increase accuracy and reduce unwanted referrals in future. Furthermore, computer-assisted programmes with automated algorithms for DR detection(15,10,19) have recently been developed and shown to achieve performance comparable to a single retinal expert reader, thus allowing for cost-effective early detection of DR, and may prove to be an important DR screening tool in future.

This study is not without its limitations. First, the FPs did not grade the severity of DR due to its retrospective nature. This parameter could not be compared among the three categories of graders. Second, the comparison between rigorously trained NPGs and FPs with busy schedules and shorter training duration may not be entirely valid. However, in the present study, we were only analysing the current practice and screening scenario, and the possibility of NPGs substituting FPs in the screening so as to unburden FPs and implement a faster and equally accurate grading process. Finally, we did not evaluate cost-effectiveness in our study, as we had assumed that the implementation of trained graders for DR screening would prove to be more cost-effective than FPs’ assessment when unnecessary referrals to an ophthalmologist are reduced.(11) Prospective studies are needed to corroborate the cost-effectiveness of DR assessment in a centralised grading centre.

In summary, the use and interpretation of digital single-field retinal images for DR screening by trained NPGs is comparable to that by retinal specialists. Hence, our findings suggest that trained graders may play a useful and possibly more cost-effective role in DR screening. These findings have broad policy implications for countries such as Singapore, where physicians are burdened with increasing clinical workload due to an ageing population, and where the prevalence of diabetes mellitus and other chronic diseases is increasing. Our suggestion is that FPs do not need to screen for DR when this can be done better and in a more cost-effective manner by NPGs.

REFERENCES

7. Lin DY, Blumenkranz MS, Brothers RJ, Grosvenor DM. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy

SMJ Best Paper Award 2012
The Singapore Medical Association will be presenting awards for the Best Research Paper published in the SMJ in 2012. All original research papers that are published in the SMJ during the one year period from January 1, 2012 to December 31, 2012 will be considered for this award. The following are the judging criteria:

1. The paper with the most potential impact on healthcare
2. Most rigorous study design/research methodologies
3. Comprehensive data analysis and balanced discussion
4. Data interpretation

Distinguished members of the medical profession will be invited to serve on our panel of judges for selecting the winning papers. The authors of the winning papers selected by our panel of judges will receive cash prizes for the first, second and third places. Prize winners will also receive a commemorative trophy and certificate.

Thank you for your support of the SMJ. The quality of our journal depends on the quality of your submissions.